CARNEGIE MELLON UNIVERSITY

School of Architecture

College of Fine Arts

Thesis

Submitted in Partial Fulfillment of the requirements for the degree of

MS Computational Design

TITLE:

Reality Integrated Design: Harnessing Real-World Data for Next-Generation Virtual Environments

AUTHOR:

David Lee Troetschel

ACCEPTED BY ADVISORY COMMITTEE:

Daniel Cardoso Llach		May 8, 2024
Professor NAME	Principal Advisor	DATE

Abstract

This thesis explores the integration of real-world data for the design and development of next-generation virtual environments, shifting away from traditional asset modeling towards a reality-based design approach. By using advanced reality capture techniques and both intentional and incidental media, this research aims to streamline the design process, enhance our understanding of existing resources, and facilitate the creation of innovative virtual realities.

Surveying reality capture techniques, processing, and segmentation techniques, the research reduces the workload associated with traditional design methods while improving accuracy and utility in virtual environment creation. This approach not only facilitates the reuse of real-world elements in multiple design contexts but also expands the potential for contemporary visualization and manipulation of design elements.

Studies utilized consumer-grade equipment to capture real-world data, ensuring the broad applicability and accessibility of these techniques. Moreover, the research anticipates future technological advancements by integrating current practices with emerging tools and methods, promising enhanced immersion and interaction within virtual spaces. This thesis sets the groundwork for a novel paradigm in design and virtual environment creation, emphasizing the practical application of integrated real-world data in computational design.

Table of Contents

Title **1**

Abstract **2**
- Photogrammetry 6
- Digital Camera(s) 6
  - Chemical/traditional photography 7
- Stereo Cameras / Camera Arrays. 7
- Light Field Systems 7
- Video Game Engines 7
- Virtual Reality 8
  - Wearable Computing / Augmented Reality (AR) 8
  - Teleoperation 8

Hypothesis **10**

Research Questions **11**
- Applications 12
- Need for media in other dimensions and immersion. 14
- What are the shortcomings of current modeling art approaches? 15
- Deep parametric systems 15
- Photogrammetry in Immersive Environments 20
- Virtual Production 20
- Heritage Site Material Management 20
- Single Images to 3D Reconstruction 21
- Prior video to 3D; Shining360 by Clair Hentshker 21
- Six Degree of Freedom Video 22
- Plenoptic systems 22
- Segmentation 24

…

Method Statement **25**
Digital/DSLR Camera Systems 26
Drone Paradigms and Video Systems: 27
Hardware Compute Benchmark: 28
Tufa Bridge, Panther Hollow 29
Second Tufa 32
Heinz Chapel 33
St Nicholas 36
St Nicholas - Unity 38
From Video 39
Penguin Walk - Kamakura 39
Noheji Japan 39
ThePOVchannel - Unity 43
Light L16 Testing 46
Philadelphia Vietnam Veterans Memorial 46
Frick Battlement Fort 47
Mars to Terrain Maps 48
AI-Based Experiments 49

Results/Findings **51**
Visual ID, manual and AI classification 53
Experiential system considerations 53
Conclusion 54

Contributions **55**

Limitations and Next Steps **56**
AI continuity 56
Further considerations: 58

Bibliography **59**
Additional sources by category: 63

Introduction and Arguments

I propose a shift from traditional asset modeling, advocating instead for utilizing real-world data as the foundation for next-generation virtual environments. Rather than creating assets from scratch or relying solely on parametric designs, we should enable designs by harnessing the knowledge and resources available in the world around us; not from abstract spaces but from tangible concepts and analogies, combining disparate elements into innovative wholes.

Most of our design endeavors are influenced by existing creations. By utilizing reality as our primary source of inspiration, we gain access to a vast array of design morphologies, both historical and contemporary. By capturing reality through established techniques and applied across diverse scenarios, we can streamline the design process, cultivate knowledge, and facilitate the creation of novel design realities.

By employing techniques such as reality capture and scene segmentation, we can significantly reduce the workload associated with design while enhancing our understanding of existing resources. Additionally, we can isolate and segment various elements from captured scenes, enabling their reuse across multiple contexts.

While standardized practices and equipment (Lachambre, 2017) for producing high-quality assets are still state of the art, such as those utilized in video game and cinema production, they are often underutilized or inaccessible to a wider audience. (Unity, 2024)

Simultaneously, there is a growing trend towards AI-based (Luma Labs, 2024) processing services, albeit with certain trade-offs, emphasizing ease of visualization over absolute fidelity. In contrast, this research primarily focuses on natural capture conditions using consumer-grade equipment and photogrammetric methods. These conditions are closer to typical common practices in media production, ensuring broad utility.

Our ability to convert large amounts of sequential photo and video footage into 3D representations opens up new opportunities for contemporary visualization and design manipulation. Advancements in related research, such as the translation of 3D models and spatio-temporal information enhance the potential of this approach.

This methodology considers the current technological landscape and anticipates future developments in interfaces and immersion systems. While the capture of reality is considered moderately challenging, the processing, display, understanding, manipulation, and replication of captured reality pose varying degrees of difficulty, with breakthroughs needed in certain areas as outlined in my conclusions.

Definitions

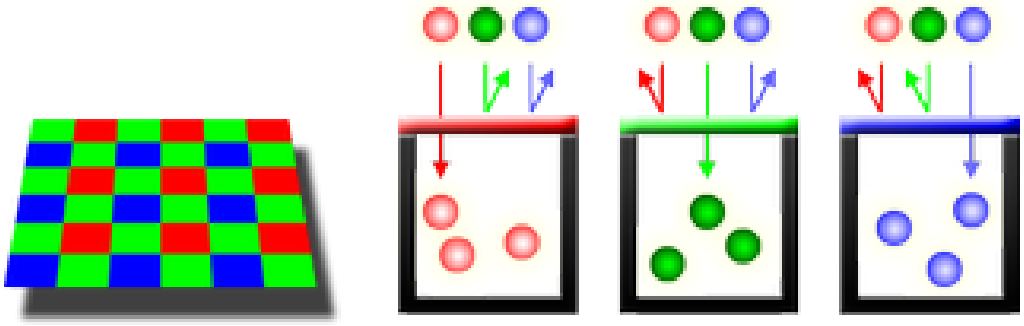
Photogrammetry

Photogrammetry is a technique that extracts precise geometric information about objects and environments from photographic images. This method involves analyzing the position, orientation, and perspective of multiple photographs to accurately reconstruct three-dimensional structures with a high level of detail. It enables the creation of detailed and accurate digital models from a series of two-dimensional images, making it invaluable in fields ranging from archaeology to engineering.

In this research, Agisoft Metashape serves as the primary processing software. This proprietary tool has proven reliable and effective, providing robust features for photogrammetric processing. The investment in its license fee is justified by its comprehensive utility and the consistent accuracy of the results it produces.

Digital Camera(s)

Digital cameras capture images using a sensor composed of numerous tiny "buckets," each equipped with color filters arranged in a grid. When light passes through the camera's lens, it hits these filters, allowing photons of specific wavelengths to fill the corresponding buckets. After the shutter closes, the sensor discharges the collected photons during the exposure time, effectively emptying these buckets. The number of photons in each bucket is then measured, converted into a digital signal, and rendered as a digital image.



(RED Digital Cinema. "Color Monochrome Camera Sensors." 2024)

Chemical/traditional photography

Traditional photography uses film coated with chemically reactive grains, which serve a similar function to the buckets in digital sensors. These grains vary in density and change composition based on their exposure to light, capturing the image as a result of this chemical reaction. For photogrammetry, traditional photographs can be utilized by scanning or digitizing multiple images of the same subject from different angles. This process allows for the extraction of depth information, potentially yielding actual 3D data from historical photographs, and offering unique insights into past perspectives and dimensions.

Stereo Cameras / Camera Arrays.

Stereo cameras, also known as camera arrays, employ two or more synchronized cameras to simultaneously capture still images or videos from different perspectives. This synchronized capture is essential for generating accurate three-dimensional depth information. The processed images allow for the creation of immersive visual experiences by providing depth perception that enhances the realism of virtual environments.

Light Field Systems

Such as those used in Lytro and Raytrix cameras, capture not only the intensity and color of light but also its direction. By recording the direction of light rays at each point in space, light field systems enable the creation of immersive, interactive visualizations with adjustable focus and perspective.

Structured light projectors, scanners and SolidLight technologies employ calibrated emitters to project patterns of light onto surfaces. By analyzing the deformation of these patterns, these systems can derive detailed three-dimensional information about the objects or environments being scanned. Structured light systems are often used in conjunction with other imaging technologies to enhance accuracy and resolution.

Video Game Engines

Video game engines are software platforms used to create and develop interactive digital experiences, primarily video games. They provide all the tools for designing, rendering, and simulating virtual environments, characters, and gameplay mechanics. Unity and Unreal Engine are two of the most widely used in the industry. They offer comprehensive features for game development, including rendering, physics simulation, animation, audio processing, and scripting.

Virtual cameras enable users to simulate and control the perspective and movement of the in-game camera. This allows for dynamic camera angles, cinematic sequences, and player-controlled viewpoints within the virtual environment.

These game engines provide tools for implementing player controllers, which govern player movement, interaction, and input within the game world. They also support interactive elements, such as objects, characters, and user interfaces of immersive and engaging gameplay experiences.

One of the key advantages of using video game engines is their support for easy distribution across multiple platforms and devices. Developers can deploy their games to various platforms, including consoles, PCs, mobile devices, and virtual reality headsets, with relative ease.

Virtual Reality

VR technologies create fully immersive environments using projected images of recorded, transmitted, or generated content that has no direct correlation to the real world. Headsets and sophisticated applications aim to achieve photorealism through advanced rendering pipelines. Interaction in VR environments is facilitated by various controllers, ranging from single to dual custom fixtures, enhancing user engagement.

Special equipment like the Infinadeck, an omnidirectional treadmill seen in the movie "Ready Player One" and Ekto VR shoes from Pittsburgh, allow users to move naturally in virtual spaces, although they face challenges like noise, sizing, and weight. Disney's experimental Holotiles offer a modular floor system that enhances movement in VR and assisted movement by using powered tilt-angled rotating discs for dynamic user interaction.

Wearable Computing / Augmented Reality (AR)

Modern wearable computing and augmented reality systems, such as Realwear, Hololens, Apple Vision, and Magic Leap, are designed to seamlessly integrate virtual objects into real-world environments. These technologies support hands-free operation, making them highly suitable for enterprise applications where efficiency and multitasking are crucial. The overlay of digital information onto the physical world enhances situational awareness and decision-making.

Augmented reality systems are gaining traction not only in business settings but also among private users and prosumers, driven by their ability to enrich interactions with immersive experiences. As these technologies continue to evolve, their adoption is expected to expand, revolutionizing how users engage with both digital content and the environment around them.

Teleoperation

Teleoperation technology facilitates remote control capabilities, which are essential for operating drones, remotely operated vehicles (ROVs), and automated systems in environments where direct human control is either impractical, unsafe, or in situations that demand enhanced precision or involve hazardous conditions.

The applications of teleoperation are diverse, spanning sectors such as cinema, surgery, construction, and space exploration. In the film industry, teleoperated drones are commonly used for camera control, providing dynamic and complex shooting angles without the need for human camera operators directly on the scene.

Similarly, in surgical applications, teleoperation enables precise manipulations in minimally invasive procedures. In construction, teleoperated machinery can perform tasks in unsafe or inaccessible areas, improving safety and productivity. Looking to the future, teleoperation will play a pivotal role in space missions, where the remote operation of equipment on other planets or celestial bodies is necessary due to the impossibility of direct human involvement.

Additionally, in the realm of photogrammetry, the use of teleoperated drones equipped with cameras allows for the collection of high-quality source material from unique vantage points, greatly enhancing the detail and accuracy of 3D models created from this data.

Hypothesis

Efficient capture of real-world information on the scene level, the segmentation of scenes into smaller assets, and the manipulation of these assets to facilitate rapid design iterations build upon existing research in the field of 3D scanning, photogrammetry, and computational design. By leveraging procedural techniques for reality capture, this research seeks to streamline the design process by reducing the overall need for manual modeling.

I investigated a variety of capture techniques including high-resolution photogrammetry, structured light scanning, and light-field imaging to ensure comprehensive and precise data collection. These methods allowed for the detailed modeling of complex environments which are critical for accurate asset creation. Additionally, I studied segmentation methods for the decomposition of large scenes into usable 3D models, significantly enhancing workflow efficiency for designers, artists, and content creators across disciplines.

Designers, artists, journalists, content creators, and documentarians can use 3D models to visualize captured data and create interactive stories, documentaries, and historical events firsthand. This will revolutionize storytelling by offering an immersive, exploratory component to understanding complex issues and events.

Linking processed historical footage and photos with current videos to create time-lapse models will allow users to virtually navigate through time in a given location. This "virtual time travel" provides insights into urban development, environmental changes, and the impact of significant events on physical spaces. The creation of personalized VR experiences from individual videos will also allow users to revisit their memories in a more immersive format. Imagine converting footage from family events, vacations, or personal milestones into 3D spaces that can be explored virtually, enhancing the way we remember and relive personal history.

The expected trajectory of related research is to truly encompass temporality, that is, movement, space and time, natural motion, six degrees of freedom, etc. We will have the holodecks of Star Trek one day via the current trajectory of emitter/sensor arrays of sufficient resolution and power.

Preserving the now, even in mediums of unknown or questionable qulity and brink utility *is* of practical importance. Future data repair, analysis, and augmentation could unlock much of the history of produced media in many interpretations. Data already exists; we need to process it differently.

Artists will use this technology to create new forms of digital art, combining real-world footage with virtual elements to produce experiences that blend reality with imagination, opening up new avenues for artistic expression and engagement as well as blurring the lines between the digital and physical worlds.

Research Questions

- *How beneficial will the creation of high-quality, readily usable reality-based assets be for the future of design and visualization?*

- *Can this approach provide designers with accurate and vivid foundations, enhancing immersion in virtual environments?*

- *What methodologies and tools are necessary to integrate scanned data with parametric modeling pipelines for the development of high-quality assets across various applications?*

- *In what areas can research further enhance processing pipelines, including interpolation, extrapolation, and camera alignment?*

- *How can we assess the strengths and weaknesses of equipment in terms of convenience, affordability, quality, and utility?*

- *What are the most effective tools and strategies for capturing and transforming real-world information into usable assets for virtual environments?*

- *Will the integration of hybrid geometric designs enhance the immersion and informativeness of virtual experiences?*

- *How can the application of high-quality assets derived from real-world data benefit sectors such as design, architecture, and education?*

- *Can future video-to-3D processing technologies effectively resolve incoherent data to accurately estimate or visualize limited representations? Will old video processed into 3D appear correct from various angles, or can the data be compounded or estimated for coherence in future analyses?*

What is at stake?

The stakes in modern design center on our ability to capture and interpret the historical context, nuances, and subtle details of forms that traditional manual methods and even generative design technologies struggle to preserve. As Kayvon Fatahalian articulates, the ambition is to "capture everything visible, enhance human communication, and empower machines to fully understand and process visual information" (Fatahalian, 2016). This drive pushes the design community beyond traditional creation from scratch, shifting focus towards realism and the utilization of a vast repository of existing assets. Today's advanced tools and technologies allow intricate capture of reality across different scales, from expansive scenes to individual objects, greatly influencing the accuracy of the final products.

The ability to transform these captured realities into practical, usable forms is essential. It involves more than mere image capture; it is about encapsulating the very essence of moments and environments, making them accessible for applications as varied as virtual reality and historical preservation.

Applications

Personalized Virtual Reality (VR) Experiences
Personalized VR experiences involve the conversion of individual videos or footage into immersive virtual environments that users can explore and interact with. By leveraging VR technology, these experiences offer a new dimension of engagement and immersion, allowing users to revisit memories or events in a more interactive format.

Once converted into VR environments, users can navigate and interact with these virtual spaces using VR headsets and controllers. They can walk through virtual environments, interact with objects, and explore scenes from different perspectives, enhancing their sense of presence and immersion.

The immersive nature of personalized VR experiences can enhance memory recall and emotional engagement, allowing users to relive past events or experiences in a more vivid and immersive way. This can be particularly impactful for preserving and sharing personal memories and stories. (Nebel et al, 2020)

Interactive Documentary and Journalism
Interactive documentaries and journalism leverage 3D models and virtual environments to create immersive and interactive storytelling experiences. By combining real-world footage, historical archives, and interactive elements, these projects offer a dynamic and exploratory way to engage with complex issues and events.

Textures continue to be a problem if settings are not optimized before importing models. Perspective and quality are off, potentially due to an initial player and camera scale compared to the environment. Resolving these problems and several others in progress or completed will help solve prior problems, going back all the way to capture.

Digital Heritage

Objects of significance need to be preserved, it is possible to use the tools at hand for nondestructive capture. It should be argued that any capture is better than none, and that crowdsourced and archival images could be used to attempt reconstruction. Ideally, this is done before an artifact is lost to allow progressive levels of quality.

HBIM (Historic Building Information Modeling)

Historic Building Information Modeling (HBIM) involves the digital representation and management of historical buildings and structures using Building Information Modeling (BIM) techniques. HBIM enables the comprehensive documentation, analysis, and preservation of historical architecture, facilitating informed decision-making and conservation efforts.

HBIM allows for systematically archival historical documents, including architectural drawings, photographs, and structural data, in one model. Virtual tours, augmented reality applications, and immersive experiences allow users to explore and interact with historical buildings in new and engaging ways. This also enables the simulation and visualization of historical buildings in various states, including past, present, and future conditions. (Banfi et al, 2017)

Virtual Time Travel

Linking historical footage and photos with current videos to create time-lapse models will allow users to virtually navigate through time in a given location. This "virtual time travel" will provide insights into urban development, environmental changes, and the impact of significant events on physical spaces.

Need for media in other dimensions and immersion

The expected ongoing trajectory of research is to truly encompass temporality (that is, movement/space and the sense of time), natural motion, 6DOF, etc. Preserving the now, even in mediums of unknown or questionable quality and brink utility is of practical importance. Future data repair and augmentation may unlock the total history of produced media in many interpretations. Current near-future possibilities using Looking Glass, Light Field Labs, etc. are allowing for easier external visualization of dimensional media. More commonly we are seeing the accessibility of VR and AR devices increase, and with it a desire for diverse new content.

Albert Parra notes “True immersion requires fully spherical capture that does not compromise one part of the spherical environment over another.” … “The ability to reliably capture and reconstruct dynamic real-world environments opens doors for automating dataset generation for machine learning techniques, e.g., depth reconstruction, view synthesis, and potentially more sophisticated forms of dynamic scene understanding.” (Albert Parra et al, 2019)

Virtual cameras allow novel opportunities in capture and timing as well as adaptable post view for visual analysis. In cinema, they allow new dynamic shots and compositing opportunities enhancing all aspects of studio workflow as reference, geometry, or art. 1:1 displays allow actors to stand within scenes with the point of view adjusted along with camera systems to re-render the background scene for near-perfect capture as if on set. Added wind, dust, etc and rotating platforms enhance this effect. Actors can also interact with virtual objects through feedback systems for more realism.

What are the shortcomings of current modeling art approaches?

Movies, advertising, and virtual production have different requirements, as do virtual museums and architectural visualizations. Most of my research has focused on larger environments with the goal of sub-feature extraction. For example, we scan a bar but only want a chair or a clock. Context is useful when it expedites and anticipates unstaged capture protocols.

Much current computational design has leveraged generative systems for artifact generation, which despite their advantages in modularity and logic, often fall short in creating outputs that reflect actual objects in complexity and specificity such as transcribing video to 3D (YouTube, etc.). Digital reality, and frankly all technology, should be understood within its proper context. Discussions around accuracy and "truth" - a hard-to-evaluate metric currently are important to have. For example, we can scan a statue, but to "check" the scan would require a much more laborious manual, error-prone measurement technique to verify, and even then would be dubious in accuracy. Should we just trust scanners after calibration? What about in the case of photogrammetry? This doesn't even address the fact that the digital reconstruction is just that, a copy.

These questions stem from the recognition that existing design systems often yield indistinguishable or extreme outlier results when explored from an atomic or hyper-parametric standpoint. Similarly, AI-driven systems output results in unconventional or unusable file types, adding complexity to their practical utility.

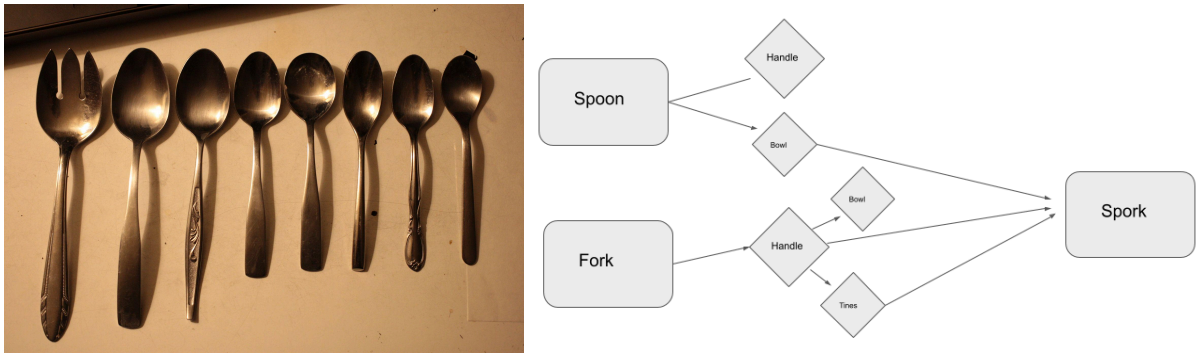
Deep parametric systems

Spronksys:

I was able to generate over thirty-two quadrillion discrete feasible forks, spoons, and sporks in Grasshopper for Rhino 3D. This approach demonstrates the breadth of possible further exploration of form segmentation-based relationships and manufacturing constraints from master patterns.

Surprisingly, I was not satisfied with most results. It becomes clear that novel forms lie at unusual parameter descriptions - it may be possible to isolate groups on "taste" and then explore further subsets and hybrids much like training a model. However, given the breadth of possible domains, this requires random sampling and appraisal. Having run through many such sets, picking the most appealing models, and then 3D printing, many of the models feel off in a nuanced way with balance and form. While a group of monkeys with typewriters like analogy may apply in this case to find acceptable forms in the dataset, the raw computation time using common hardware is too excessive.

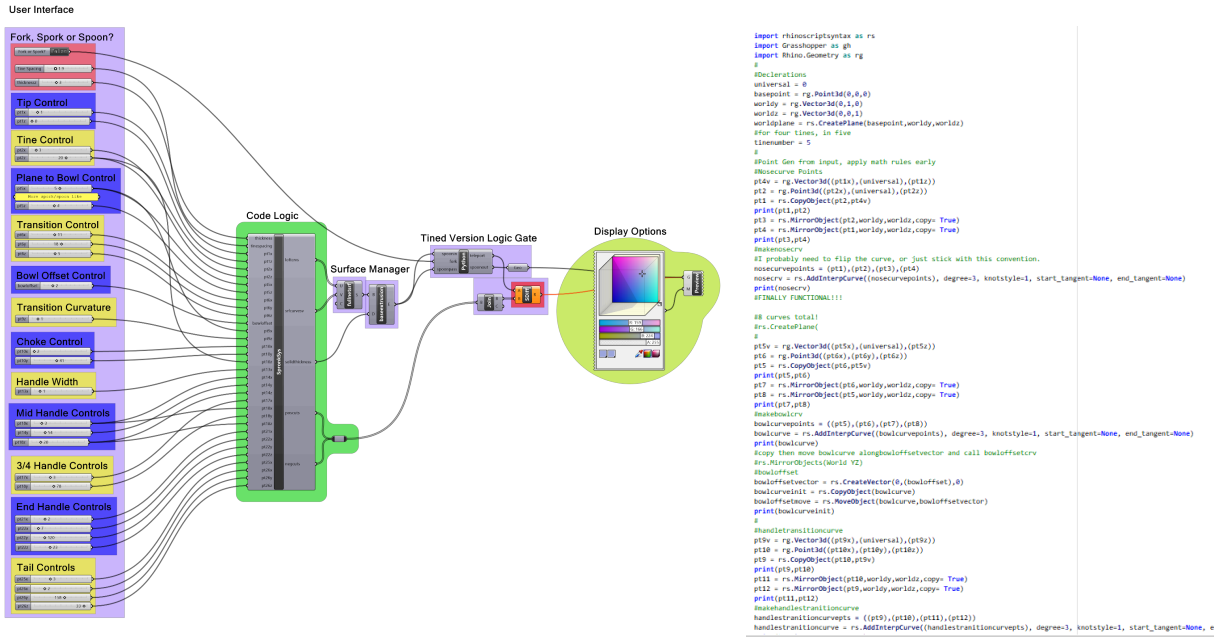
To create my generative model I sampled several forks and spoons. On the same grid, I developed outline and elevation plans, which were manually measured. These became the bounds of my system. Of note, this experience incentivized my desire to 3D scan the referenced utensils. Additionally, I looked at ratio tables for tine length to bowl and handle length features to create a domain and tunable relationship allowing fine control. This allows fine manual control to achieve a close approximation to an intended result with permutations for a designer to then apply non-linear feature variables such as filets and tailored modifications.



(Measured utensils. Logic outline. Photo and screenshot by author, 2023)

Still, while these assets would pass in any game or movie prop setting, they require careful selection for real-life desirability at a high resource and computation cost. It may have been preferable to 3D scan the samples and develop a morphing type attractor solution, but a sensor of the proper resolution for reflective cutlery proved difficult to obtain. Being able to generate so many delineations in the form of something as a pronk system has value, if nothing else than to show us the complexity in seemingly simple forms.






```
import rhinoscriptsyntax as rs
import Grasshopper as gh
import Rhino.Geometry as rg
#
#Declarations
universal = 0
basepoint = rg.Point3d(0,0,0)
worldy = rg.Vector3d(0,1,0)
worldz = rg.Vector3d(0,0,1)
worldplane = rs.CreatePlane(basepoint,worldy,worldz)
#for four tines, in five
tinenumber = 5
#
#Point Gen from input, apply math rules early
#Nosecurve Points
pt4v = rg.Vector3d((pt1x),(universal),(pt1z))
pt2 = rg.Point3d((pt2x),(universal),(pt2z))
pt1 = rs.CopyObject(pt2,pt4v)
print(pt1,pt2)
pt3 = rs.MirrorObject(pt2,worldy,worldz,copy= True)
pt4 = rs.MirrorObject(pt1,worldy,worldz,copy= True)
print(pt3,pt4)
#makenosecrv
#I probably need to flip the curve, or just stick with this convention.
nosecurvepoints = (pt1),(pt2),(pt3),(pt4)
nosecrv = rs.AddInterpCurve((nosecurvepoints), degree=3, knotstyle=1, start_tangent=None, end_tangent=None)
print(nosecrv)
#FINALLY FUNCTIONAL!!!

## curves total!
#rs.CreatePlane(
#
pt5v = rg.Vector3d((pt5x),(universal),(pt5z))
pt6 = rg.Point3d((pt6x),(pt6y),(pt6z))
pt5 = rs.CopyObject(pt6,pt5v)
print(pt5,pt6)
pt7 = rs.MirrorObject(pt6,worldy,worldz,copy= True)
pt8 = rs.MirrorObject(pt5,worldy,worldz,copy= True)
print(pt7,pt8)
#makebowlcrv
bowlcurvepoints = ((pt5),(pt6),(pt7),(pt8))
bowlcurve = rs.AddInterpCurve((bowlcurvepoints), degree=3, knotstyle=1, start_tangent=None, end_tangent=None)
print(bowlcurve)
#copy then move bowlcurve alongbowloffsetvector and call bowloffsetcrv
#rs.MirrorObjects(World YZ)
#bowloffset
bowloffsetvector = rs.CreateVector(0,(bowloffset),0)
bowlcurveinit = rs.CopyObject(bowlcurve)
bowloffsetmove = rs.MoveObject(bowlcurve,bowloffsetvector)
print(bowlcurveinit)
#
#handletransitioncurve
pt9v = rg.Vector3d((pt9x),(universal),(pt9z))
pt10 = rg.Point3d((pt10x),(pt10y),(pt10z))
pt9 = rs.CopyObject(pt10,pt9v)
print(pt9,pt10)
pt11 = rs.MirrorObject(pt10,worldy,worldz,copy= True)
pt12 = rs.MirrorObject(pt9,worldy,worldz,copy= True)
print(pt11,pt12)
#makehandlestransitioncurve
handlestransitioncurvepts = ((pt9),(pt10),(pt11),(pt12))
handlestransitioncurve = rs.AddInterpCurve((handlestransitioncurvepts), degree=3, knotstyle=1, start_tangent=None, e
```

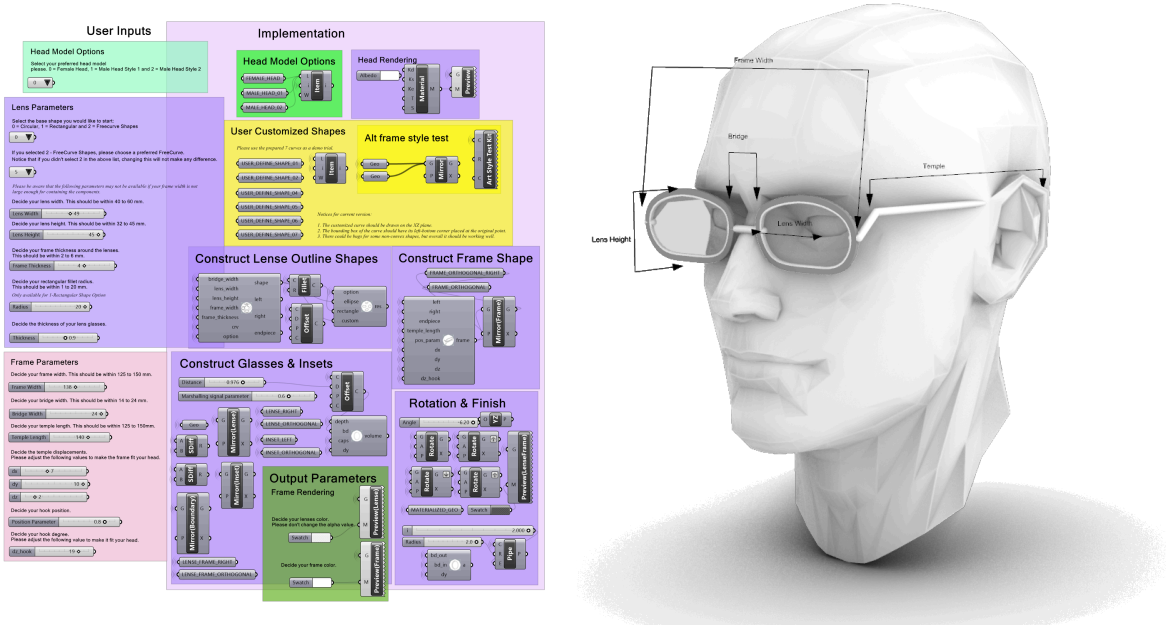
(3D printed models beside reference fork. UI and script. Photo and screenshots by author, 2023)

A compressed and refined GhPython script runs much more efficiently than the standard definition. However, we are left with a highly expansive but under-constrained design space with an absurd number of nearly indistinguishable results. Moreover, without random sampling techniques or manual manipulation of variables, the exploration of possible configurations is cumbersome.

Lingheng Tao and I also advanced a parametric glasses generation system, ”Glades2” (Tao, Troetschel. 2023) enhancing its processing efficiency and usability features. The interface is designed with an initial setup that flows from left to right and top to bottom, facilitating a natural workflow. The controls are organized into groups based on shared features, which simplifies navigation and adjustment.

We recommend using a dual monitor setup to fully view the control array and monitor live changes effectively. A detailed written guide accompanies the system, providing users with step-by-step instructions on navigating and utilizing the various options.





(3D printed frame, User interface, Design visualizer. Photo and screenshots by author, 2023)

In the context of parametric systems designed for niche applications like sporks, bicycles, and glasses, there is a valuable opportunity to integrate real-world data to refine the control logic and functionality. This approach involves employing universal modules of ontological hierarchy, which can be enhanced through segmentation labeling and the analysis of relationships. By segmenting and understanding the relationships and interactions within the data, we can create more intuitive and effective design tools that respond dynamically to user inputs and environmental contexts.

Background

Photogrammetry in Immersive Environments

"Photogrammetry for Game Environments 2014-2019: What Happened Since The Vanishing of Ethan Carter" provides an extensive overview of the adoption and impact of photogrammetry in video game development from 2014 to 2019. This period marked when photogrammetry moved from a novel technique to a standard expectation in high-quality game productions. Their study highlights the incorporation of photogrammetry in 26 notable games, including major franchises like Battlefront, Far Cry, and Resident Evil. The research underlines the technological advancements, the challenges of capturing and processing digital environments, and the legal and ethical considerations of using real-world locations. The study also examines the media's portrayal of photogrammetry as a transformative, almost magical tool, which oversimplifies the complexities involved while changing audience expectations. (Statham et al., 2020)

"History Is Our Playground": Action and Authenticity in Assassin's Creed: Odyssey," also explores the blend of historical narrative and gameplay within Ubisoft's Assassin's Creed series, specifically focusing on the title set in ancient Greece. They discuss the game's role as both a source of entertainment and a potential educational tool through its historically enriched settings and narratives while assessing the game's impact on players' understanding of history through its reconstruction of ancient Greece. (Politopoulos et al., 2019)

Virtual Production

Virtual Production (VP) technologies within the audiovisual industry offer new opportunities, especially the needs of small and medium independent productions. The authors of “Using Immersive Technologies to Facilitate Location Scouting in Audiovisual Media Production: A User Requirements Study and Proposed Framework” explore the potential of integrating Virtual Reality (VR) and Augmented Reality (AR) with real-time graphic engines, highlighting how such technologies can enhance creativity, collaboration, and efficiency in media production. A significant focus is placed on the process of location scouting, a critical pre-production activity that can be revolutionized by virtual technologies. (Bodini et al., 2023)(Shapiro, 2024)

Heritage Site Material Management

Monika Stobiecka's paper "Archaeological Heritage in the Age of Digital Colonialism" looks into the ethical and political implications of digital archaeological practices, particularly focusing on the reconstruction of destroyed heritage sites. Stobiecka critiques the digital reproduction of the Syrian Arch of Triumph, which was destroyed in 2015 and later replicated using digital technologies. The replica was exhibited in various Western cities but never returned to Syria,

raising questions about the colonial nature of such digital interventions in heritage. (Stobiecka, 2020)

Single Images to 3D Reconstruction

The study "SDF-SRN: Learning Signed Distance 3D Object Reconstruction from Static Images" by Chen-Hsuan Lin, Chaoyang Wang, and Simon Lucey at Carnegie Mellon University explores the innovative potential of reconstructing 3D objects from single images without direct supervision. The researchers introduce SDF-SRN, a method that uses signed distance functions (SDF) derived from 2D silhouettes to create 3D shapes. This technique simplifies the training process by requiring only single-view images rather than multi-view annotations. This method shows considerable promise in improving machines' ability to interpret complex 3D environments from minimal inputs, which could substantially streamline the creation and modification of digital models. (Lin et al, 2020)

Panoptic Systems

At Carnegie Mellon University, the "Panoptic Studio" project (Hanbyul Joo et al., 2015) developed 3D capture stages for social observation and later acquired by Meta (formerly Facebook) using a dome array setup for dynamic motion capture, processed on dedicated servers. This setup exceeds the current work discussed in this thesis, which instead focuses on static scenes and imagery and their application in existing display and navigation systems. This process is also extremely memory-intensive, now requiring the use of portable, container-based memory servers that are transported for archiving and processing at significant cost. Their primary goal is to streamline avatar creation, using just a few images from a cell phone and extrapolating from a high-quality dataset to enable 3D social interactions. An excellent example of the system's application is Lex Fridman’s 2023 interview with Mark Zuckerberg in the Metaverse. (Fridman, 2023.)

Prior video to 3D; Shining360 by Clair Hentshker

Clair Hentshker's Shining360 project represents a novel approach to visualizing camera perspective and geometry derived from entire film sequences. This work illustrates how specific camera movements can affect the quality of photogrammetric results.

Beyond this, Hentshker’s portfolio explores broader themes, using time and space as foundational elements for experiencing and visualizing environments. Her projects such as "Memories of Home" capture evocative snapshots in time, providing a vivid 3D portrayal of personal and historical spaces. Similarly, "Ghost Coaster" revisits a demolished cultural icon, reimagining it through 3D technology. These works highlight the potent combination of narrative

and technical prowess, offering new ways to perceive and interact with our past and present surroundings. (Hentschker, 2016)

Six Degree of Freedom Video

Free-viewpoint video, also known as six degrees of freedom (6DOF), is a type of immersive video that allows users to select their viewing angles in a 3D space, offer a more immersive and comfortable viewing experience but is limited by the need for complex and costly equipment such as camera arrays and light-field cameras, processing, and content delivery compared to traditional video pipelines. (Shaohui Jiao et al., 2023)

Jingwei Huang and their colleagues highlight the limitations of 360-degree videos through virtual reality (VR) headsets in "6-DOF VR Videos with a Single 360-Camera." One notable issue is that 360 cameras capturing monoscopic videos render identical visuals for each eye, failing to replicate the sense of depth found in 3D environments. These videos only cater to rotational movements, while modern VR headsets are equipped to handle six degrees of freedom. This flaw tends to make the VR experience less interesting and can cause motion sickness since visual information does not align with the user's physical movements. (Huang, 2017)

As far as 3D content delivery and systems, the paper "A 6DoF VR Dataset of 3D Virtual World for Privacy-Preserving Approach and Utility-Privacy Tradeoff" by Yu-Szu Wei et al. examines various datasets used for 360° video streaming, an early application of VR technology. These datasets serve different research purposes, including the design and evaluation of viewport/gaze prediction algorithms using data on users' heads and eye movements. Specific datasets focus solely on eye movements to classify these actions while others aim to understand the relationship between simulator sickness and head rotations or to explore relations between biometrics, head/eye movements, and user emotions. However, these datasets do not contain personal attributes like age, gender, and height, limiting their use in studies focusing on privacy-preserving methods for personal attributes. (Yu-Szu Wei et al., 2023)

6DOF video is generally focused on the subject/content within a scene, often with little to no scene. Labs such as Canon, etc. have developed Livestream systems for free viewpoint rendering. These are often used for dynamic capture of an actor. The complexity and expense of these systems made personal experimentation impractical, but an important part of next-generation media is an outline of the technologies and their connections to other systems, cost, complexity, and research into cheaper alternatives.

Plenoptic systems

Plenoptic systems utilize light field technology which captures light from multiple angles to record different viewpoints of a scene. This allows for capturing a wealth of visual information that traditional cameras cannot, such as binocular vision simulation, motion parallax, and effects such as highlights and reflections. This level of detail makes plenoptic systems ideal for applications like virtual reality, computer games, teleconferencing, film production, and broadcasting. (Grinspun et al, 2016)

Despite their advantages, plenoptic cameras such as the first-generation Lytro, have limitations. They can only provide detailed depth maps at close distances and have difficulties in modeling curved or transparent objects. This limitation restricts their use in certain cases, such as complex scenes or materials. Additionally, these systems are typically very cost-prohibitive.

Unlike traditional methods such as laser scanning, plenoptic imaging captures only the light that exists in the scene without emitting any additional light. As Murgia points out, this makes it particularly useful in sensitive environments where additional light may damage photosensitive materials. Plenoptic cameras provide a non-invasive alternative to other 3D reconstruction technologies similar to photogrammetry. (Murgia, 2023)

Advancements such as the Fraunhofer IIS' light-field plug-in suite demonstrate the post-production possibilities of these systems enabling the synthesis of various camera views post-capture, enabling virtual camera movements, changes in depth of field, or shifts in zoom. This flexibility is ideal for professional settings like film production and virtual set design. (Fraunhofer IIS, 2015)

The method of capturing a sequence of images at different distances from the subject, as described by Shreedhar Rangappa, is pivotal in generating accurate depth information. This process involves carefully measuring changes in camera position relative to the object to ensure the depth data is not just relative but absolute. This approach is fundamental in achieving the precision required for detailed 3D modeling and applications in scientific and industrial fields.(Rangappa et al., 2018)

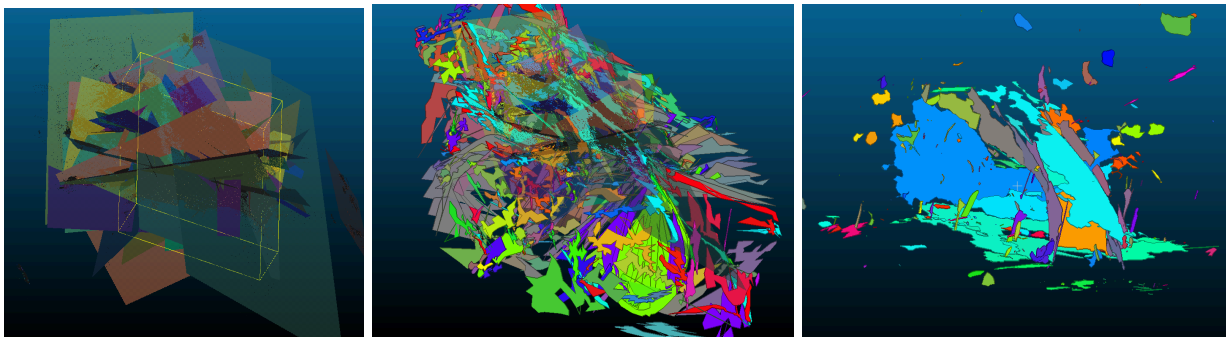
Segmentation

Segmentation processing involves breaking down complex data into manageable segments to better understand the structure and features of the objects being analyzed. The segmentation process can be performed through various methodologies, including region-growing, where segments are formed by grouping neighboring pixels based on predefined criteria, and scalar-based segmentation, which utilizes RGB values to distinguish different parts of a dataset (Anuar et al., 2016)

For instance, projects like Microsoft COCO involve segmenting images into instances for precise object localization, demonstrating the practical application of segmentation in enhancing the utility of large datasets for machine learning and other applications. However, limitations remain, such as the inability to infer finer details like the internal mechanics or assembly features of objects solely from their external imagery (Lin et al., 2015)

Recent research emphasizes the importance of integrating 3D representations with conditional generative modeling to overcome challenges in image synthesis from multiple viewpoints. This integration allows for high-resolution, 3D-aware rendering that is responsive to user inputs, ensuring that each 3D point is assigned accurate semantic labels for color and density, facilitating precise image and label map rendering (Deng et al., 2023)

I captured both Tufa bridges in Schenley Park, Pittsburgh, Pennsylvania multiple times shown later in this document and utilized CloudCompare's built-in segmentation functions to see which elements could be isolated. I tested four methods and found that for these organic forms, Fastmarch provided the most usable results.



(Ransac, Kdtree, and Fastmarch segmentation using CloudCompare. Image by the Author.)

Method Statement

This research utilized image sequences and edited videos to enhance the quality and utility of photogrammetric 3D model processing. Multiple field photography projects were performed focusing on large-scale scene reconstruction of both natural and manmade structures. Whether photos with consistent metadata from varying perspectives or videos segmented into scenes devoid of occlusions like unexpected hands or selfies, the subsequent generation of 3D models form the backbone of this study.

Kdenlive, an open-source video editor built on Qt and KDE Frameworks libraries, was employed extensively throughout the research. It facilitated detailed video editing and rendering to image frames, which were then processed further using photogrammetric techniques. Comparisons of outputs were made in Metashape to assess strengths and weaknesses. This comparison helped in understanding the impact of various frame rates and settings on resulting geometry and textures. The research suggests opting for sub-framerate sampling to avoid redundancy in frames, which can lead to excessive use of processing resources without improving the model quality.

Digital/DSLR Camera Systems

Cameras manufactured after 2010 generally produce high-quality media, making them suitable for this research. Secondary cameras are recommended to serve as backups or for reference. Smartphone cameras can also be utilized, but users should be aware of their inherent image stabilization and pre-processing features. For lenses, fixed focal length options with minimal distortion are preferred. Avoid using the zoom function and lock the lens at its widest field of view to ensure consistency across shoots. Photos should always be taken in landscape orientation as this simplifies the overlapping of images; portrait mode can complicate image stitching processes. Avoid taking photos while walking or in motion unless using a stable, vibration-free platform.

It is advisable to use a strong ring flash in conjunction with a polarized filter to manage natural lighting conditions and eliminate reflections and highlights. While not essential and not used in this research, these lighting adjustments can lead to more reliable results. Ensure images are saved in RAW or JPG format to retain the integrity of the data for post-processing. Always carry extra batteries or a power source to avoid disruptions during shoots. The Canon 1D Mark III, with its ASP-C sensor and 40mm lens that effectively provides a 27mm focal length, was the primary camera used in this research, chosen for its wide viewing angle and low distortion properties.



(Canon 1D Mark III Camera. Photo by author, 2024)

Drone Paradigms and Video Systems:

Drones equipped with RTK and GPS are invaluable for capturing aerial footage in remote or hazardous locations, functioning effectively as flying cameras that enhance the efficiency of data capture.

When setting up video systems, use prime lenses to reduce distortion and set the shutter speed high enough to eliminate blur while maintaining even lighting. Use mechanical or gimbal stabilizers over internal stabilization to avoid cropping issues. Maintain the camera's landscape orientation consistently, and pause shooting if an obstruction enters the frame, resuming once the path is clear. Utilizing self-announced audio cues can be helpful for post-production editing to manage these interruptions by looking for level spikes provided the video is not intended to also serve as supplementary material.

The Insta360 OneR used in this research is a modular action camera particularly effective when settings are adjusted to correct for the linearity of images from the wide-angle lens. The internal setting avoids distortion for this application. The camera's long battery life makes it well-suited for prolonged use in the field, although gyro stabilization should not be used in post-processing to prevent image cropping errors.



(Insta360 1R Camera. Photo by author, 2024)

Experimental Computational Photography:

Light L16: The Light L16 camera, designed with 16 lenses, aimed to replace traditional DSLRs in a large cellphone form factor running Android with Bluetooth and WiFi. Each lens is individually focused, allowing a composite 52mp image to be perfectly sharp, and adaptable in post-processing for different focal lengths. Despite its initial high cost, slow focus, exposure speeds, and the failure of its manufacturing company, the L16 remains available at a much-reduced price through resellers such as Ebay, offering an enticing option for experimental photography platforms. The dense sensor spacing of the L16 allows it to capture multiple stereo images simultaneously. However, its proprietary software does not currently support exporting each raw image individually. Fortunately, open-source efforts are ongoing to reverse-engineer and unlock more of its capabilities for 3D imaging and other advanced applications.



(Light L16 Camera. Photo by author, 2024)

Hardware Compute Benchmark:

Desktop system: Microsoft Windows 10 Home, Physical Memory (RAM) 32.0 GB, Total Virtual Memory 36.7 GB, AMD Ryzen 5 3600 6-Core Processor 3.60 GHz.

Methods and Processing

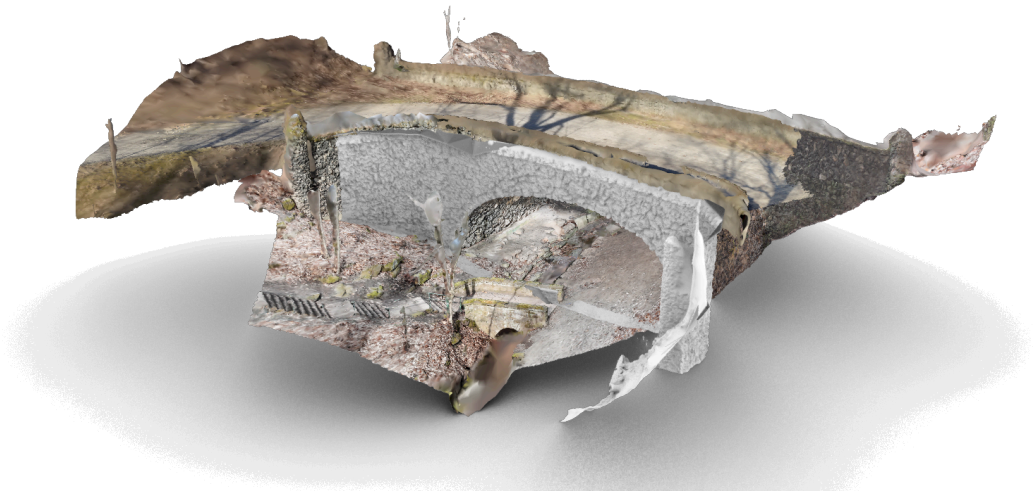
Tufa Bridge, Panther Hollow

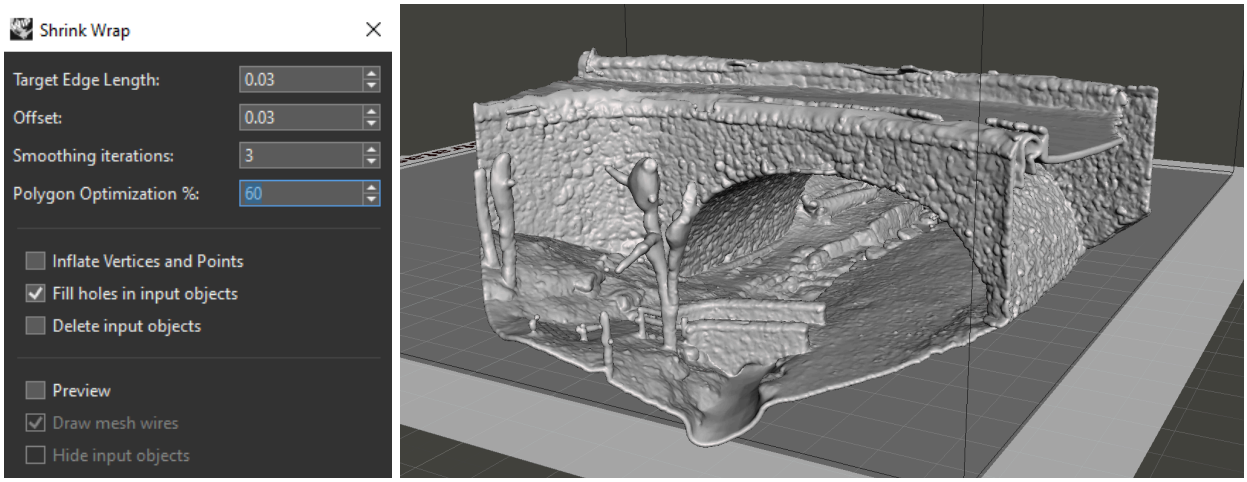
The Tufa Bridges, a pair of closed-spandrel arch structures on the Schenley Park Bridle Path in Pittsburgh, Allegheny County, Pennsylvania, served as the subject of this photogrammetry study.

For this project, I used various camera systems. Initially, I conducted a preliminary site survey using a Sonim XP7 phone and an Insta360 OneR, capturing 56 images to create an initial low-resolution model. This early model helped evaluate the site's lighting conditions during the day and structural challenges, particularly in how the bridge's edges transitioned from the sides to the top. These images were compiled into a mosaic using Microsoft Image Composite Editor to summarize the dataset visually.

I returned to the Panther Hollow site equipped with my Canon 1D Mark III and captured an extensive dataset of 1800 images from multiple perspectives. Initially, I attempted to process this dataset as a single batch, but computational limitations forced a reevaluation. After several unsuccessful attempts, I segmented the data into five distinct chunks, each focused on key features of the bridge. This segmentation allowed for successful individual processing of each chunk over a week using my computing resources at the time.

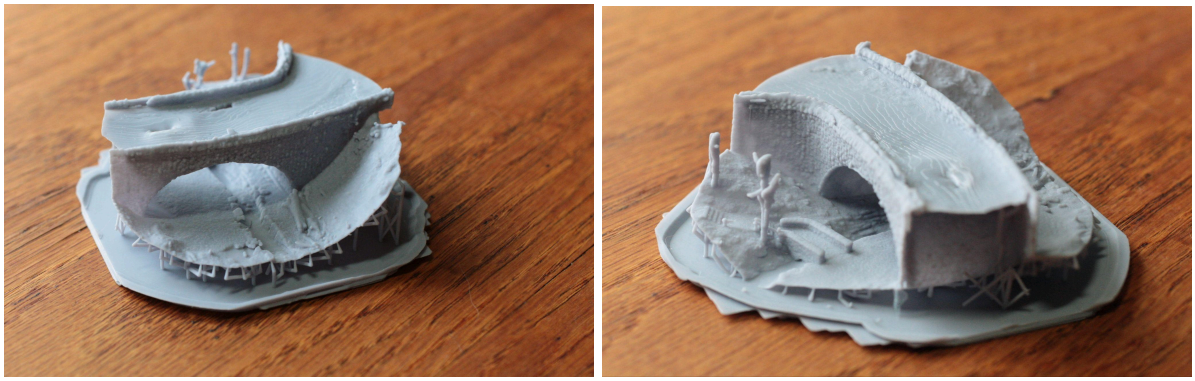
Combining the processed chunks into a unified model proved problematic. The alignment issues and the sheer data volume overwhelmed my available hardware capabilities at the time, a Dell M4700 laptop. I explored alternative software solutions like MeshLab, Meshmixer, and Zeiss to aid in merging the datasets, but none proved effective given the noisy datasets. Eventually, I utilized Rhino to adjust and align the meshes manually, then used the shrink-wrap function. This approach required down-scaling to lower-resolution meshes, which, while not ideal, facilitated a more manageable processing workflow.





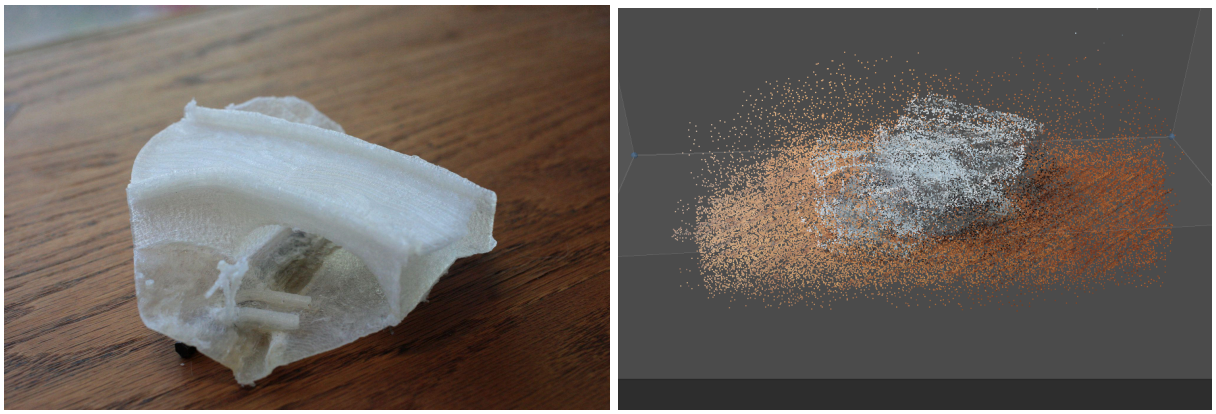
(Rhino renderings of stages of mesh processing. Captures by author, 2023)

With the meshes aligned, I 3D printed models of the Tufa Bridge. The initial 3D prints, created with the assistance of Cody Soska at Carnegie Mellon University, utilized resin printing to capture detailed features of the bridge and its textures. The results were promising, showing detailed stonework and structural integrity, though the top surfaces presented some minor issues.



(Photos of 3D prints of digital models. Photos by author, 2023)

Further attempts to utilize FDM printing highlighted the limitations of this method for detailed architectural features. The larger-scale prints lacked the finer details and were not representative of the original bridge's intricate architecture. Still, these combined prints served as a valuable proof of concept for the potential of 3D printing in architectural preservation, visualization, and presentation models.



(Photos of 3D prints of digital models. Photos by author, 2023)

I photographed the resulting 3D prints under varied lighting conditions to attempt to create a new set of photogrammetry data. Unfortunately, this data was too noisy to be useful, demonstrating the challenges of re-replicating precise architectural details through photogrammetry from 3D printing. Recreations and novel designs in mediums like stone now have robust foundations. (Monumental Labs, 2024)

Second Tufa

Given the prior Tufa Bride, capture and processing I formulated my approach to the second Tufa Bridge site, using a figure eight scanning pattern from top of the bridge, left inside of the curve to the other side, valley, pass under, return to valley origin. The imagery was clean and comprehensive and the resulting 3D models were found to exhibit better visual qualities in extrapolation given this context.



(Photos of the Schenley Hill Tufa Bridge. Photos by author, 2023)

Second Tufa - Unity



(Screenshots of the 3D model environment in Rhino3D and Unity. Captures by author, 2023)

Heinz Chapel

I scheduled testing before St. Nicholas Chapel using a larger but similar environment in order to develop a long-term strategy. I performed two outside counterclockwise loops at a distance and for close-up details with fifteen percent frame overlaps. Due to the proximity of trees, some obstructions caused tracking errors in processing for approximately one-third of the exterior. For the interior I did one clockwise lap from the front to rear and then a zig-zag through the column sections focusing on the altar.

In preparation for the St. Nicholas Chapel project, I conducted preliminary testing at Heinz Chapel to refine our capture strategy and processing techniques. This involved performing two counterclockwise loops around the chapel's exterior to capture the structure from a distance and gather close-up details, ensuring fifteen percent frame overlaps. Unfortunately, due to the proximity of trees, approximately one-third of the exterior images faced tracking errors during processing.

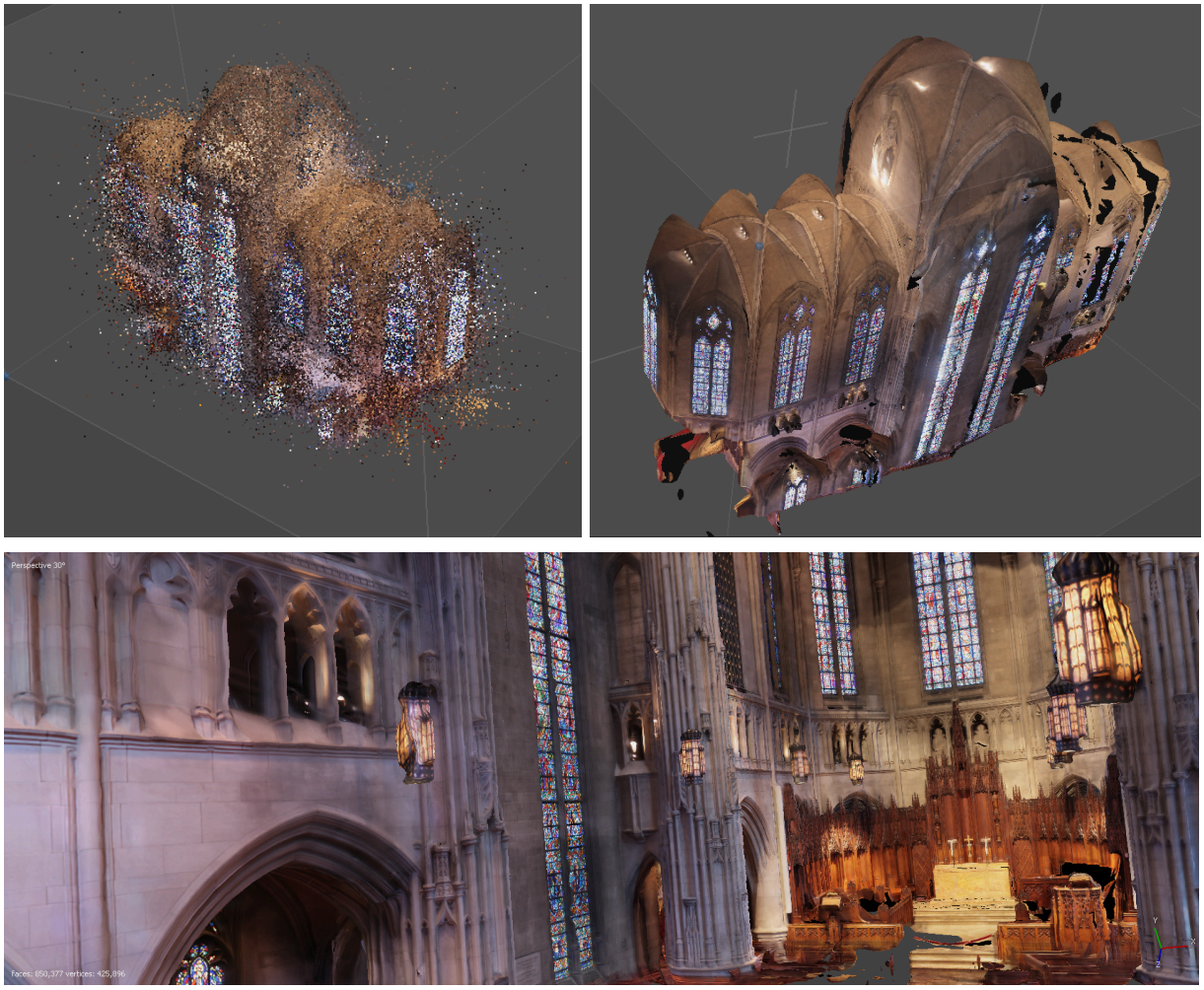
For the interior, I adopted a different approach, performing a clockwise lap from the front to the rear of the chapel and then executing a zig-zag pattern through the columns, with a particular focus on the altar and ceiling spans.



(Exterior photos of the Heinz Chapel. Photos by author, 2023)



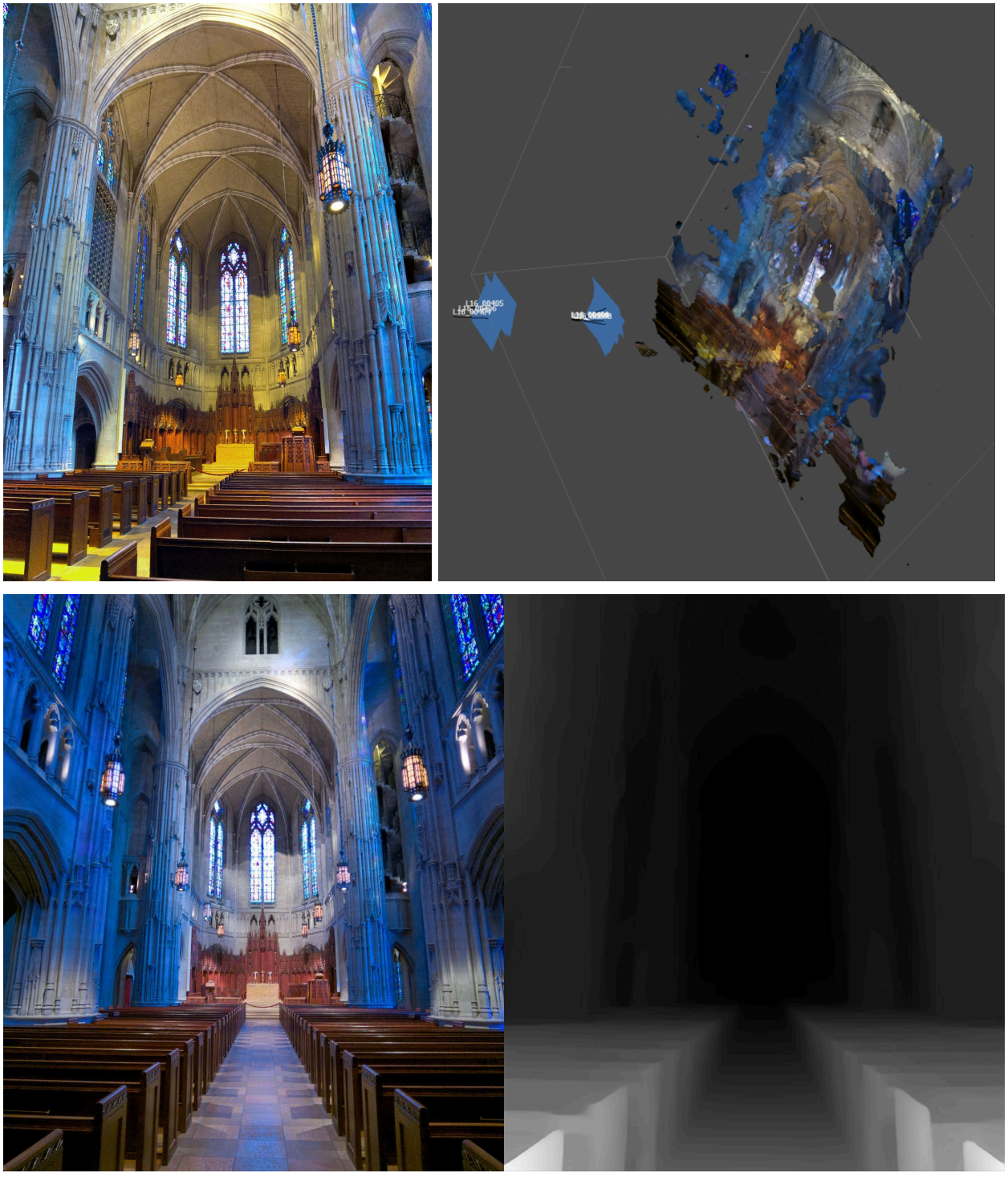
(Interior photos of the Heinz Chapel. Photos by author, 2023)



(Agisoft Metashape steps and renders of the Heinz Chapel. Captures by author, 2023)

The interior scans yielded excellent results, capturing most architectural details vividly. However, there were notable exceptions including the floor, benches, and some areas of the ceiling that did not meet quality expectations. These areas were specifically noted for attention in the subsequent sessions at St. Nicholas Chapel.

Despite the challenges posed by the chapel's exterior size and the visual obstructions from surrounding foliage, the data collected provided valuable insights into potential improvements for future scans. Utilizing a drone for exterior scans could potentially mitigate issues related to sight angles and obstructions, providing clearer and more consistent coverage. A limited dataset trial using the Light L16 camera produced outstanding results, generating composites of over 100 megapixels that were successfully converted into 2.5D and proved compatible with light field displays. This test underscores the potential of using advanced camera systems to enhance the quality and utility of photogrammetric data in complex architectural environments.



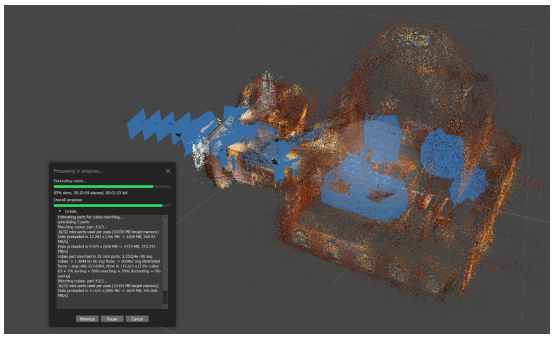
(HDR Light L16 photo composite, Agisoft Metashape processing, and Lightfield Labs depth map processing for the holographic display of the Heinz Chapel. Photos by author, 2023)

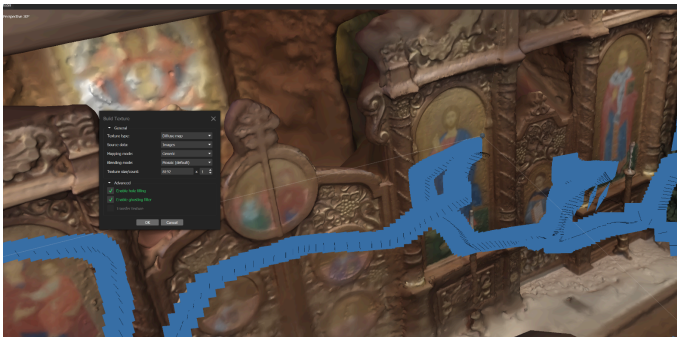
St Nicholas

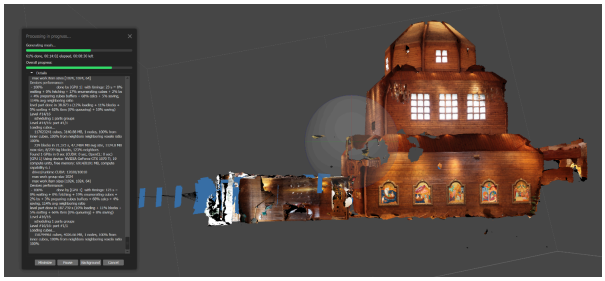
As an independent study under Daniel Cardoso Llach's coordination, our team conducted an extensive documentation project at St. Nicholas Chapel in Beaver, Pennsylvania, utilizing a diverse range of photographic and video methods. I captured approximately 2700 images, which were then used to generate detailed 3D models via photogrammetry in Agisoft Metashape.

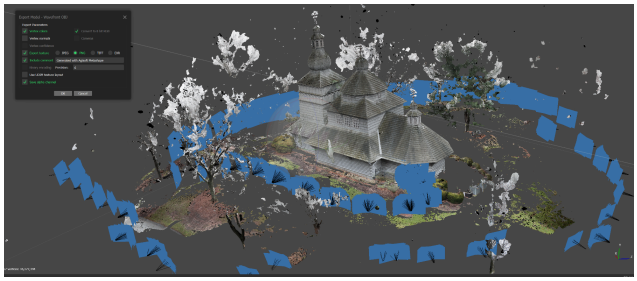
In a collaborative effort, additional datasets and sub-models provided by colleagues Isaac Martinotti, Michael Hasey, and Felicia Luo, who contributed with Lidar and further photogrammetry data, were integrated into another file version.

Given our plan to scan St. Nicholas Chapel, it was a natural progression to integrate these models into a game environment using the Unity game engine, aiming to create an immersive educational tool. Our meticulous data collection included nearly three thousand photos, six videos, and various other captures. Felicia and Isaac utilized a Matterport Lidar scanner, achieving highly accurate results. This blend of Lidar and photogrammetry proved crucial, especially in areas my data did not cover.











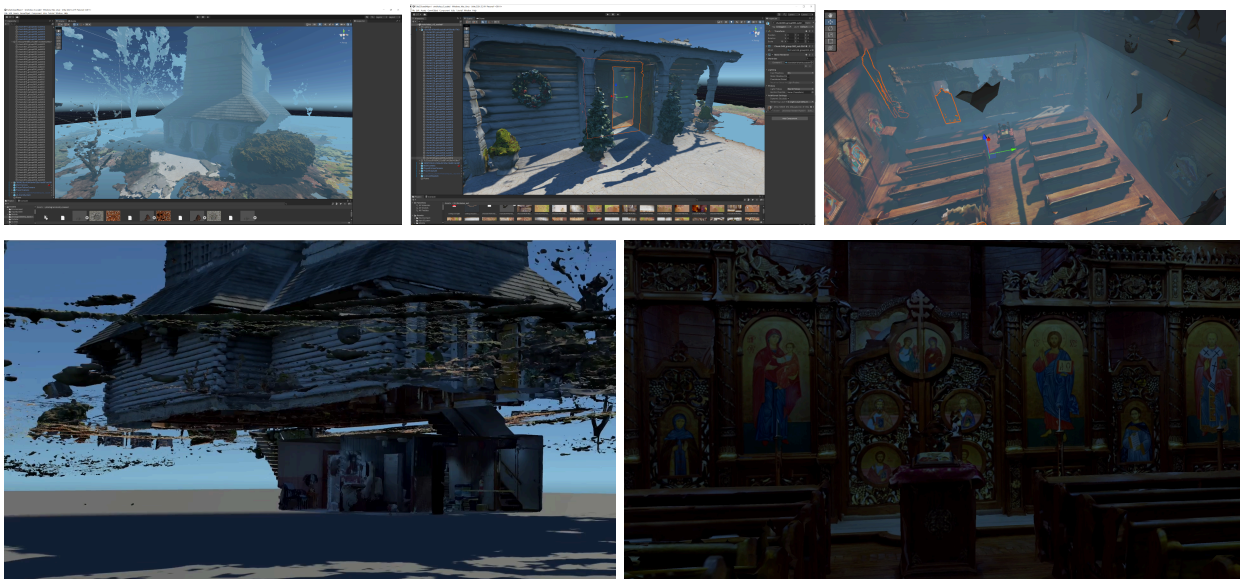
(Agisoft Metashape steps and renders of the St. Nicholas Chapel. Captures by author, 2023)

St Nicholas - Unity

The photorealistic textures and models from St. Nicholas Chapel were imported into the Unity game engine, allowing users to navigate the scanned space either in first-person or third-person views or even fly through the scene at will. This adaptability extends to potential multi-player, web, or mobile deployments, facilitating a broad distribution through a single executable installation.

I manually aligned and scaled the scan data in Unity to ensure accurate rendering, with player character dimensions adjusted for easy navigation through doors and tighter areas. Despite challenges in capturing the floor, roof, and finer details like benches, the interactive controls (using WASD for movement, Spacebar to jump, and Shift to run, along with mouse controls for camera movement) allow for an engaging exploration of the environment.

As of the latest update, the independent installer for the virtual environment, V7, functions smoothly. It requires about 500 MB for installation and using approximately 1.3 GB of memory when fully installed. This version is tailored for Windows systems and is capable of running on contemporary hardware configurations. Access to V7 is currently available through direct contact with the author.



(Unity renders of St. Nicholas Chapel. Captures by author, 2023)

From Video

Penguin Walk - Kamakura

This project utilized video footage from the “Penguin Walk” channel, titled "[4K Japan Walking] Strolling in the back alleys where the Enoden runs in Kamakura," available on YouTube. The selected segment, running from 00:27 to 15:42, was processed into 4,518 source images using KdenLive, with a naming sequence starting from Trimmed_00002 to Trimmed_27104, using every six keyframes (e.g., 98, 92, 86, and 80). (Penguin Walk - 4K Japan Travel, 2023)

Two distinct 3D photogrammetry models were constructed using Agisoft Metashape. The first model, which was extrapolated, required three days of computing time, while the second, interpolated model was completed in approximately two days using a Dell M4700. Both models encountered issues with tracking at certain transitions, resulting in a fractal-like scaling within the larger scene. These anomalies, however, provided unique high-resolution secondary scene models that were isolated during post-processing. This could be due to excessive frames with micro-variations.

The resulting models were archived and then rendered at twenty-five percent of their original quality to optimize rendering speed for viewing in real-time using Rhino 3D’s viewport render combined with OBS video screen capture.

Results were mixed and initial outputs were somewhat disappointing. Elements of each processing method contributed valuable data. But they would need to be segmented for a comprehensive scene, again highlighting the need for continued research in segmentation for this approach.

Noheji Japan

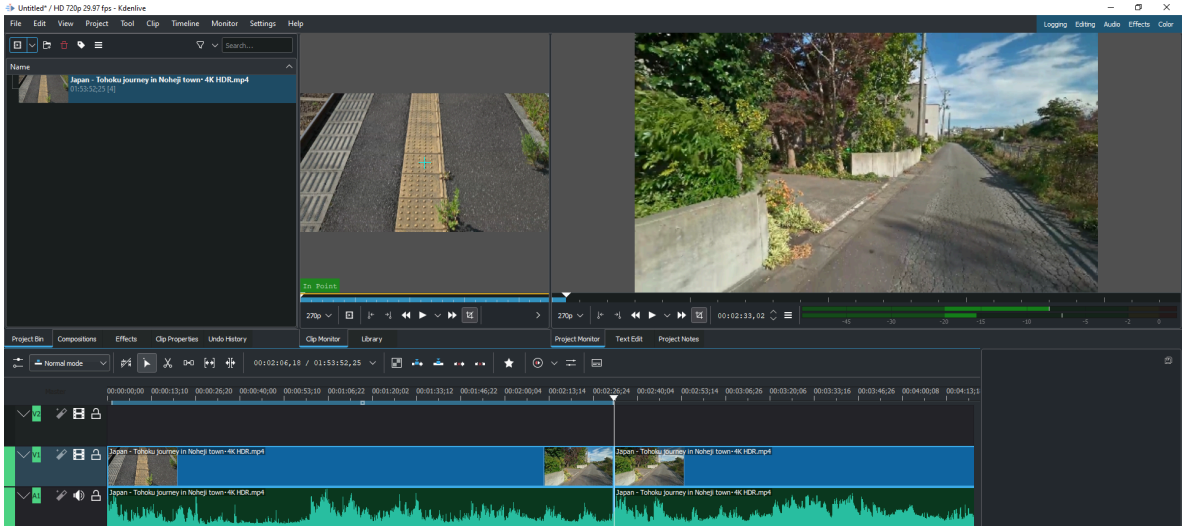
"Noheji Japan" continued my investigation into 3D tourism through the detailed rendering of locations to provide immersive and detailed experiences, highlighting the importance of texture quality and volumetric preservation in creating realistic visual representations.

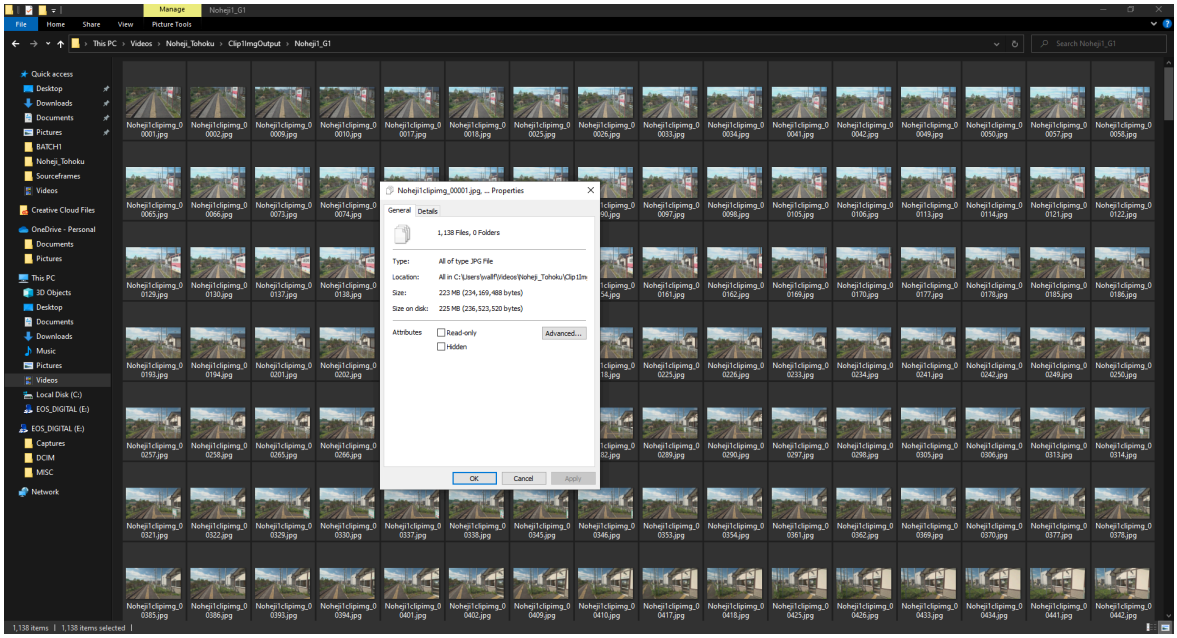
A significant challenge was the noisy movement nature of the input data, which complicates texture fidelity. To address this, transitioning from two minutes of video to over one thousand images, I then converted varying sample groups from the core dataset. This iterative process of trial and refinement helped me overcome the limitations posed by the initial data quality.

Each processing step was characterized by distinct strengths and weaknesses. Some methods excelled in noise reduction but compromised on detail, while others preserved detail at the expense of leaving residual artifacts. The final model preserves detail close to the original camera path but also extends the viewing range of side streets and architectural details not focused on in the initial video.

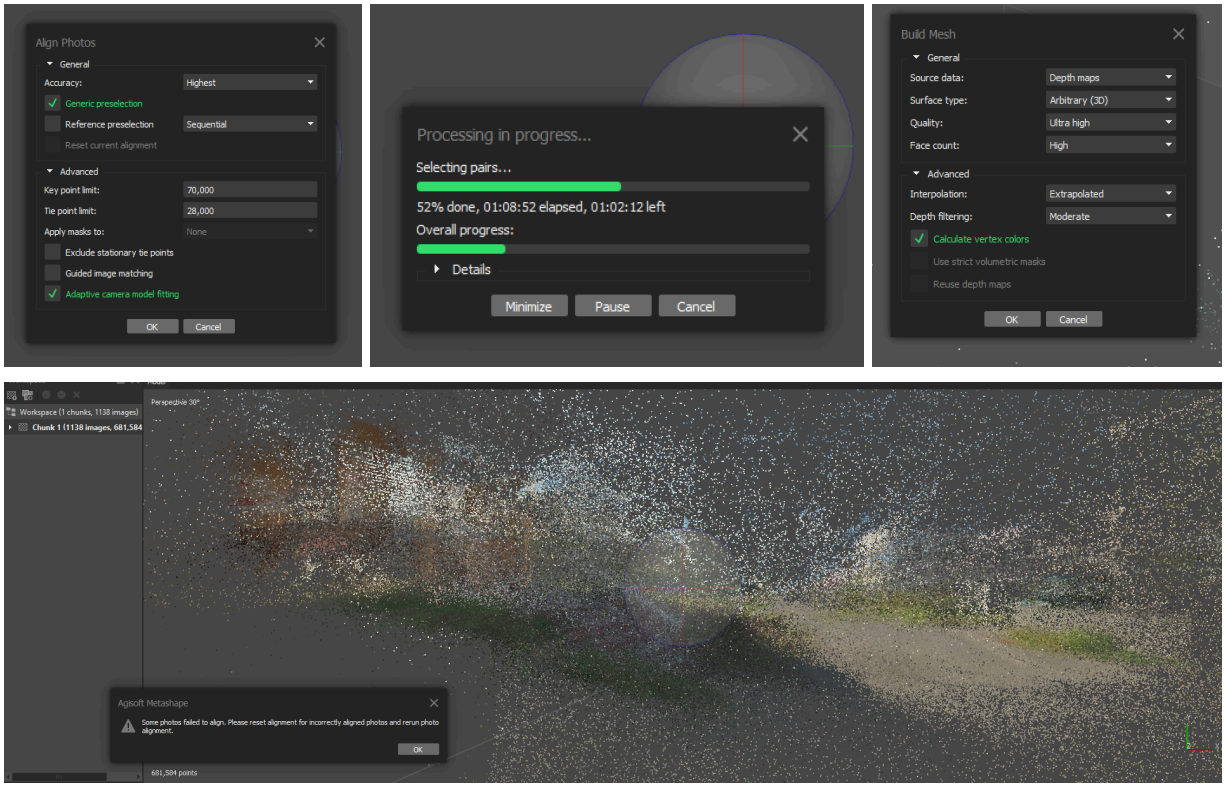


(Image frames from video. Captures by author, 2023; Source imagery by Rambalac, 2023.)





(KdenLive editing and resulting frames. Captures by author, 2023; Source imagery by Rambalac, 2023.)



(Agisoft Metashape processing settings and point-cloud. Captures by author, 2023)



(Second dataset round, processing, and extrapolated mesh. Captures by author, 2023)

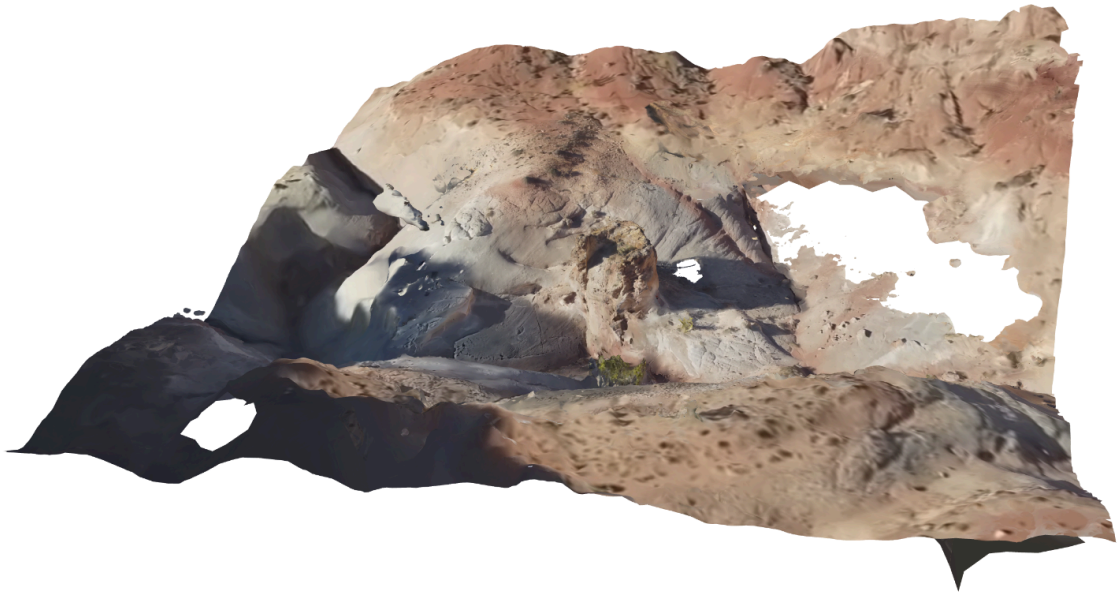
ThePOVchannel - Unity

ThePOVchannel, operated by a YouTuber named Nolan Fisher, specializes in exploring and documenting geographical formations and potential archaeological sites using satellite imagery. Fisher's work includes capturing drone and handheld footage suitable for photogrammetry, particularly highlighting unique natural structures. A notable video featuring a central pillar-like formation in a depression caught my attention for its potential in 3D modeling.



(Lightfield Labs depthmap. Image from video by Nolan Fisher, author processed, 2023)

These 3D models were developed from YouTube footage obtained with the author's permission. The collaboration yielded excellent feedback, particularly highlighting the immersive, free-walk experience that the 3D render provided. This interaction allows users to appreciate the true scale of the formation, sparking a desire for an unbounded exploratory experience as one approaches the model's edges.



(3D mesh rendering in Rhino. Capture by author, 2024)

Nolan expressed enthusiasm about the accuracy of the models' scale and the fidelity of the captured shapes, although he noted that the textures need improvement. Encouraged by these results, he has produced a new video centered on Lidar data (Fisher, 2024) and is planning a photogrammetry-specific collaboration set for the summer of 2024.

In his feedback, Nolan mentioned: “Hey David! The world scale looks pretty spot on, assuming the camera level is at the player's eye level. I can't believe how well you pinned down the shape of the object. Yeah, the texture is off, but other than that it's a pretty good render of the area for scale. I'm curious as to what you think could be the widespread application of this type of thing. I'm definitely down to catch up on a call and brainstorm some ideas. Also, not sure if you saw the newest video — I have a large volume of raw drone files of some of these geometric-shaped rocks. The object in the next video might be more interesting for you. I’ll also shoot you a text. Cheers!” - (Nolan Fisher of “thePOVchannel.”, email to author, February 13, 2024)



(Virtual camera renderings in Unity. Captures by author, 2024)

Additionally, the same input data was utilized to test Gaussian Splatting. While this method was adequate for creating a fast visual impression, it was not suitable for detailed applications due to numerous visual abnormalities.



(Gaussian Splat render of corresponding photogrammetry data. Capture by author, 2024)

Light L16 Testing

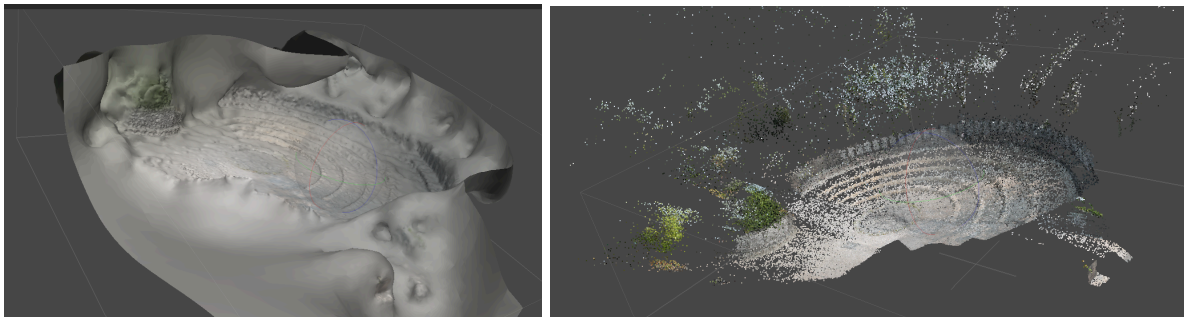
Philadelphia Vietnam Veterans Memorial

Input images were made using a right-to-left panning motion, eleven frames.



(L16 HDR input images. Captures by author, 2024)

Processing in Agisoft Metashape; interpolated, extrapolated. The extrapolated model was seamless but with lower texture quality, the interpolated processing looked better both in appearance and surface quality. Merging the models would be ideal. I also processed a Microsoft Image Composite Editor photomosaic for a total overview giving early insight into abnormalities in the 3D model, of which I detected no significant anomalies.



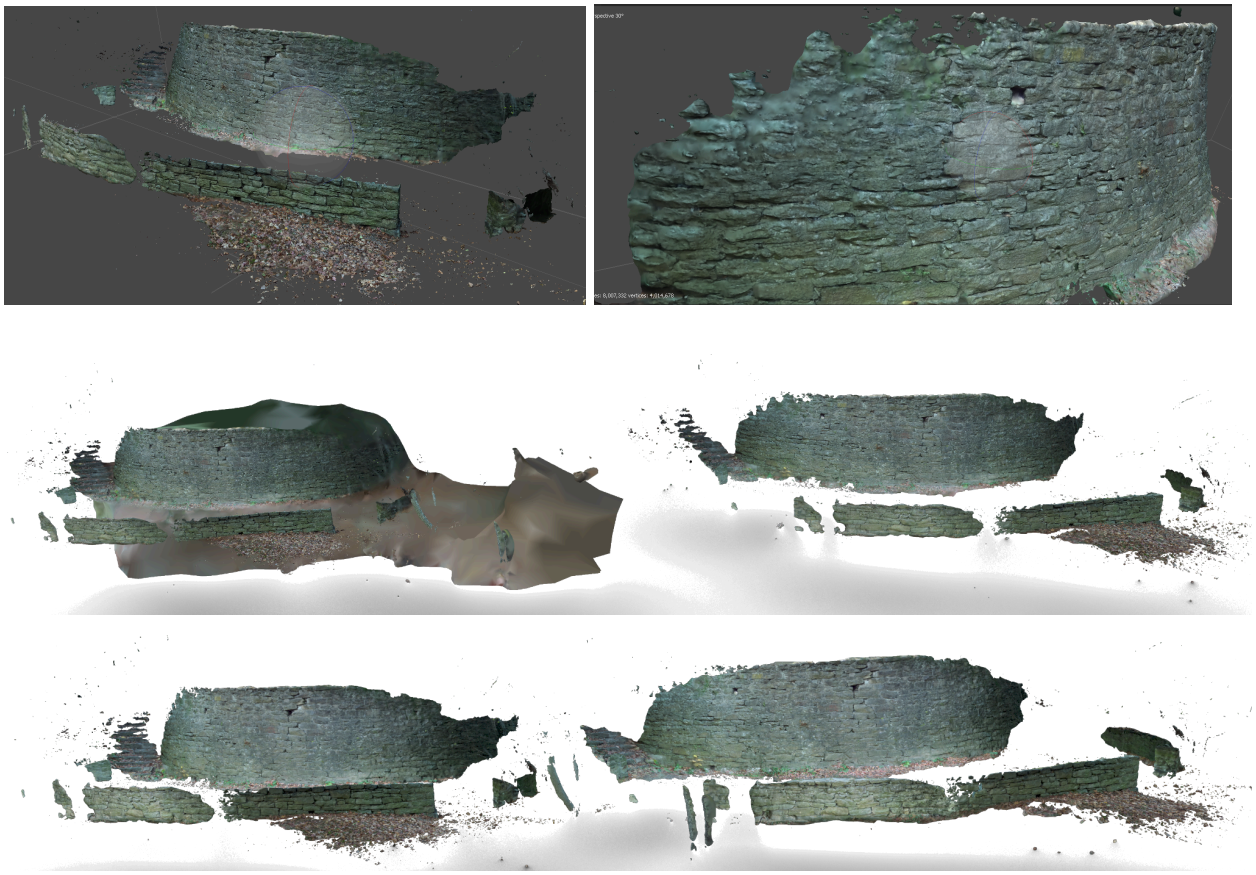
(Extrapolated and interpolated processing in Agisfot Metashape. Captures by author, 2024)



(Whole scene photo composite by author, 2024)

Frick Battlement Fort

Using the L16 I captured thirty-five input images from the front of the fort structure on the Homewood Trail near the Frick Museum in Pittsburgh, Pennsylyvania. 3D models were processed via four methods. Again, much of the ground was missing. This may be due to insufficient aggressive depth filtering of the images in mesh construction. Wall features were accurate and detailed with accurate spacing between the primary elements.

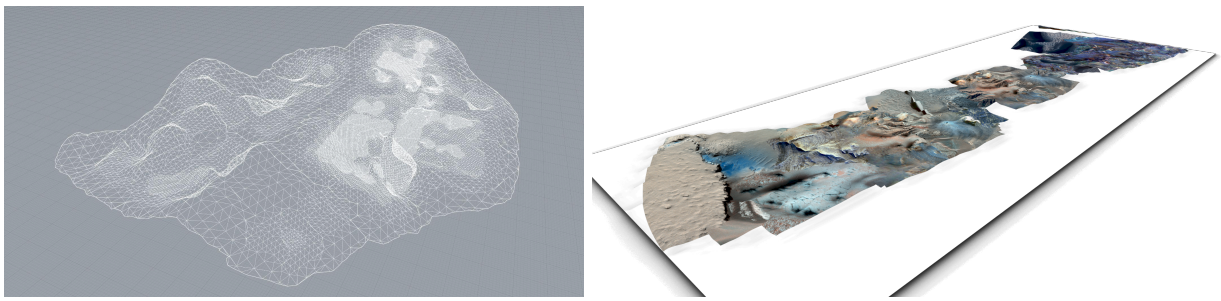


(Agisoft Metashape initial models and full series rendered in Rhino3D. Captures by author, 2024)

Mars to Terrain Maps

Using imagery from Mars in the public domain and Microsoft Image Composite Editor (ICE) I combined, in various alignments, a range of mosaics. I then processed height maps in Rhino 3D and 2.5D processing in Agisoft Metashape to develop terrain and textures for a video game map concept. Greater access to stereo camera data would allow more accurate reconstructions on future missions.

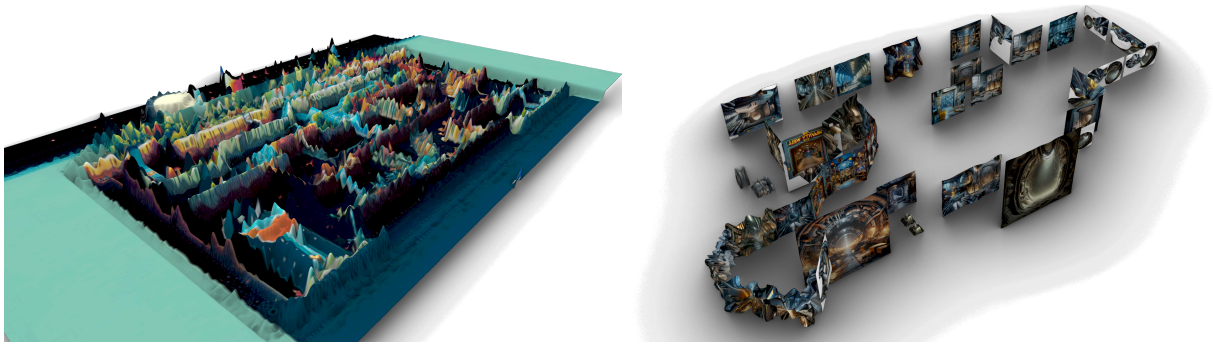
Note that 2.5d is not geometrically accurate. ICE can be used to merge random image data in various configurations for creative purposes to create terrain. ICE can also be used to photo stack images for higher resolution although this seems to rarely be of benefit. This method needs further investigation, similar to the later discussed L16.



(Martian orbiter imagery to Rhino height-field processing. Captures by author, 2024)

AI-Based Experiments

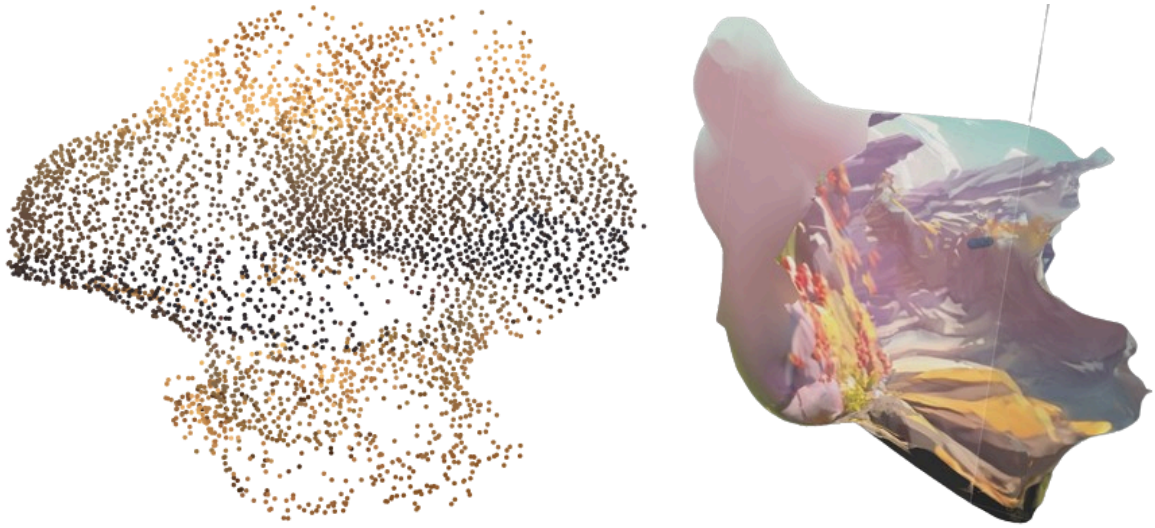
In exploring the applications of AI to generate 3D models from imagery, experiments with various AI tools such as Point-E, MidJourney, and Skybox by Blockade Labs yielded suboptimal results. Notably, Point-E, despite its potential, produced point clouds that were too abstract for practical use in photogrammetry, and the software did not permit downloading the points, only providing rendered images. Similarly, attempts to employ MidJourney and Skybox for generating usable textures in photogrammetric applications consistently resulted in 2.5D projections when processed via Agisoft Metashape.



(Dall-E AI-generated images to height fields. Captures by author, 2024)

These challenges show the current limitations of AI-generated imagery, primarily suited for creating basic surface maps or static backdrops rather than dynamic, detailed 3D models. The utilization of Runway ML to simulate movement through a scene from a single image also did not yield successful outcomes.

Human Esmaeili and Harold Thwaites's study on virtual photogrammetry investigated the feasibility of using rendered images and virtual cameras to create 3D models similar to contemporary AI imagery. Their research, conducted at the Centre for Research-Creation in Digital Media at Sunway University, Malaysia, illustrates the potential of non-physical cameras in reconstructing detailed virtual environments and objects, suggesting a future where virtual and AI-assisted photogrammetry could overcome current technological constraints. (Esmaeili and Thwaites, "Virtual Photogrammetry" 2016).



(AI-generated point cloud and matching pose photogrammetric mesh. Screenshots by author 2024)

Results/Findings

When transcribing found video footage from platforms like YouTube to 3D, we gain access to diverse perspectives and angles that differ from the original footage. This variability allows us to highlight different aspects of the scene, some of which may be clearer than others based on the initial camera path. By creatively using the same footage, we can produce much longer and more intriguing videos and enable free exploration.

Using intentionally captured, static site-based imagery allows for the recreation of scenes with geometric and visual accuracy. These generated scenes integrate seamlessly into video game engines and online platforms, facilitating remote viewing of locations. For designers, these scenes serve as precise references or a foundation for further creative exploration.

Lidar scans offered an unexpected advantage: the scans often consist of many segmented pieces. While it can be cumbersome to apply colliders to each object, this allows us to selectively omit colliders from certain elements like closed doors. This capability lets players walk through walls without the need for animating door openings—a crude yet effective solution that reduces the need for mesh manipulation, essentially acting as pre-segmentation.

Hybrid geometric design emerges as a powerful byproduct of these techniques, offering the potential to intuitively modify or enhance based on existing models. This design approach is underscored by the need for careful delineation. For example, if you wanted to add cat ears to a bus, only a few logical placement options exist. Similarly, elongating a fork involves extending the handle but not the tines, unless specifically intended, ensuring that modifications are logical and coherent.

Robust methods of analyzing sets of source imagery

The creation of a sophisticated video processing pipeline tailored to handle motion and abrupt scene changes would represent a significant advancement in transforming sequential or cohesive datasets into 3D assets. This pipeline, currently in its early stages of implementation, needs to be designed to split video clips or image sequences into batches for optimal processing through photogrammetry.

One of the core challenges in this process is managing interruptions during video capture, such as a rock climber's hand obscuring the camera. These disruptions can severely affect the processing techniques, leading to errors such as multiple fractal environments or visual distortions—manifesting as fogs in Neural Radiance Fields (NeRFs) or issues in Gaussian Splats. Correcting these errors requires significant cleanup, which can be time-consuming and resource-intensive.

To enhance the efficacy and accuracy of the pipeline, video motion-based subdivision and pixel shift methodologies should be employed. This strategy involves fine-tuning a model to achieve the best output of clips by establishing specific criteria related to camera motion and style, identified through experience and prior research. This setup would not only filter out undesirable video segments or images but also helps in categorizing the footage into sub-folders based on the camera's motion and interruptions.

There are explorations into the application of 3D object classifiers (Bradski, G G. "The OpenCV Library." *Dr. Dobb's Journal of Software Tools* (2000.)) that utilize multi-perspective or stereo views to refine the segmentation and classification process. While my current lack of experience with sophisticated tools like OpenCV poses a challenge, it also presents an opportunity. By setting custom parameters and defining clear criteria along with appropriate training data, OpenCV could significantly improve this process, making it more robust and versatile.

Dividing meshes

Mesh subdivision significantly impacts the workflow of 3D asset creation. Manual subdivision, while precise, is labor-intensive and often impractical for large datasets. Conversely, automatic subdivision, a more novel approach, enables the identification of primitive shapes, specific features, and planes, facilitating streamlined processing.

To achieve broader adoption of large-scale 3D models to individual artifacts requires effective segmentation; it's crucial to focus on selectively extracting desired elements from complex scenes. For instance, extracting only a bench from an extensive scene necessitates advanced labeling and retrieval systems to ensure ease of use. Contemporary labeling systems, primarily reliant on fixed or two-dimensional imagery, are inadequate for accurate 3D environment segmentation. Successful 3D segmentations require the extrusion of label boxes across multiple views and the meticulous splitting of resulting meshes, with ongoing label maintenance and adjustment. This often involves a regressive approach, segmenting from the overall scene down to individual sub-assets, especially when multiple objects intersect within a single scene.

Visual ID, manual, or AI classification to cropping windows based on multiple perspectives

Classification of 3D objects incorporates visual identification and manual, or AI-driven methods, to define cropping windows based on multiple perspectives. Classifiers vary widely in approach and accuracy are frequently biased towards specific training sets and lack generalization capabilities. Utilizing tools like the Google Images labeler on various perspectives of the same object may require multiple labeling passes to accurately segment the object from the parent mesh. This process could be carried out by locking viewport scales in Rhino, zooming to a focal subject, rendering output views, and processing these images through image classifiers to align labels and overlay model views, trace or extract label boundaries, and ultimately extrude and slice the mesh according to these results.

The potential of combining multiple data sources with these advanced segmentation techniques to produce high-quality assets suitable for virtual environments is significant. While optimizing geometry and textures to reduce file sizes and enhance runtime efficiency remains challenging, the development of extensive asset libraries demonstrates considerable promise for their application across various contexts.

Asset output libraries

To further enable this approach, robust asset management systems are essential. ID tags, based on sample input numbers, resolution, techniques, or custom tags, should be part of the file naming convention to facilitate easier removal of inappropriate references later. Such systems will not only aid in design foundation and reference generation but are also vital for immediate evaluative needs like discovery, archiving, forensics, and providing attributes and context for 3D knowledge databases. Advanced formats such as light-field and holographic media storage, as well as streaming capabilities for moving 3D assets, are also being explored to further enhance the functionality and accessibility of 3D asset libraries.

Experiential system considerations

The use of the Unity game engine has been critical in my implementation of experiential systems for virtual environments. The engine facilitates a range of interactive features, including first-person controls typical of video games (such as jumping, WASD, and mouse view control), multiplayer capabilities, and online play. This also allows us to embed extra content, responsive environments, and non-linear storytelling. These environments can be packaged into single executables or applications compatible with platforms like Steam, providing broad accessibility without the need for additional support at the base visualization or interaction level. (Von Stackelberg & McDowell, 2015)

However, there are challenges associated with these high-definition environments, primarily related to memory and computing limitations that may hinder users' ability to download and explore these spaces. These issues can be mitigated by optimizing the environments, such as by reducing the mesh polygon count, which helps to balance visual quality with performance.

Explorations into applying these technologies beyond gaming have included a range of potential sites for virtual recreation. Initial tests with the Unity-based "Tufa" experiments prompted investigations of other possible locations, such as museums and historic sites in Pittsburgh. Interesting opportunities like Trundle Manor, resisted digital archiving due to concerns about dissuading physical visits, and the Allegheny Observatory was unsuitable due to ongoing construction. More complex projects, like recreating the Pittsburgh Pennsylvania Kelly-Strayhorn Theater in high fidelity for replayable performances, were found to be too resource-intensive due to the need for multiple depth cameras and extensive data handling as well as logistical issues in recording performances at this time.

Conclusion

This research into reality-based asset creation and manipulation has uncovered several key findings and implications for design, visualization, and storytelling. By leveraging real-world data and advanced technologies such as photogrammetry and 3D modeling, designers and content creators can access a wealth of high-quality assets for use in virtual environments, interactive experiences, and historical preservation efforts. This approach not only streamlines the design process but also enhances immersion and engagement by providing users with realistic and contextually relevant content.

The potential impact of reality-based asset creation extends across various fields, from video game development and architectural visualization to interactive documentaries and historic preservation. By incorporating reality-based assets into their projects, designers, and storytellers can create more immersive and compelling experiences that resonate with audiences on a deeper level. Additionally, reality-based assets offer new opportunities for exploration, education, and cultural preservation, enriching our understanding of the past and present.

Contributions

- I established a robust comparison between photogrammetry, Lidar, and Neural Radiance Fields (NeRF), demonstrating the distinct advantages and application scopes of each method in capturing real-world data for virtual environment development.

- Investigations of the application of topological analysis tools were evaluated in the classification of 3D objects, enhancing the accuracy of digital models and facilitating complex architectural and environmental studies.

- I integrated real-world data into game engines for interactive and immersive user experiences. This included developing methods to refine real-time display technologies and enhance user interaction within these virtual spaces.

- Processes were developed and techniques refined for conventional and experimental image datasets, ensuring high-quality 3D models and paving the way for future innovations in computational photography and visualization.

- Testing with innovative camera systems was conducted with systems such as the Light L16 and I explored plenoptic and light field imaging technologies. This research contributed to understanding the limitations and potentials of these emerging technologies in the field.

- The conversion of synthetic data into 3D models was tested, providing insights into the challenges and potential of AI-generated art and photogrammetry in creating realistic virtual representations.

- I developed new methods for converting video footage into 3D models, significantly impacting how visual media is transformed into interactive and immersive virtual environments.

- I investigated the legal and ethical implications of converting global multimedia content into 3D assets, addressing concerns related to permissions, royalties, and fair use, and providing a framework for ethical content transformation.

Limitations and Next Steps

Looking ahead, there are several areas for further research and development in reality-based asset creation and manipulation. Continued innovation in data capture techniques, processing algorithms, and visualization tools will be essential for advancing the state-of-the-art in this field. Additionally, interdisciplinary collaboration between designers, technologists, historians, and storytellers can drive discoveries and applications of reality-based assets in diverse domains.

AI continuity

As of Spring 2024, the results from AI-driven sequential image generation, interpolation, and video models have been mixed, with generally poor outcomes. However, given the rapid advancements in this field, significant improvements are anticipated. A groundbreaking challenge is the need for enhanced multi-view synthesis that can consistently create at least 4-6 views of a generated object with uniform features and alignment crucial for achieving reliable results in photogrammetry.

Currently, artificially generated maps and assets tend to be limited to 2D or 2.5D representations. These often suffer from definition problems or issues with camera movement, which may stem from inadequate rendering instructions or a fundamental lack of knowledge about the hidden geometry from different angles. Although all attempts thus far have utilized photogrammetry, it may be that techniques such as Neural Radiance Fields (NeRFs)(Mildenhall et al., 2020) could offer better handling of synthetic data. Photogrammetry struggles with recognizing correlations in virtually generated scenes, which may be because the search algorithms used in Agisoft Metashape are unable to reconcile the scenes effectively. When viewed sequentially, it becomes apparent to the human eye that the generated imagery lacks many essential qualities needed for coherent scene expression.

Other contemporary research such as "3D-aware Conditional Image Synthesis" by Kangle Deng, Gengshan Yang, Deva Ramanan, and Jun-Yan Zhu at Carnegie Mellon University suggests promising directions for these technologies (Deng et al., 2023). This research underscores the potential of advancing how AI interprets and reconstructs complex visual data for more realistic and dynamically generated 3D models.

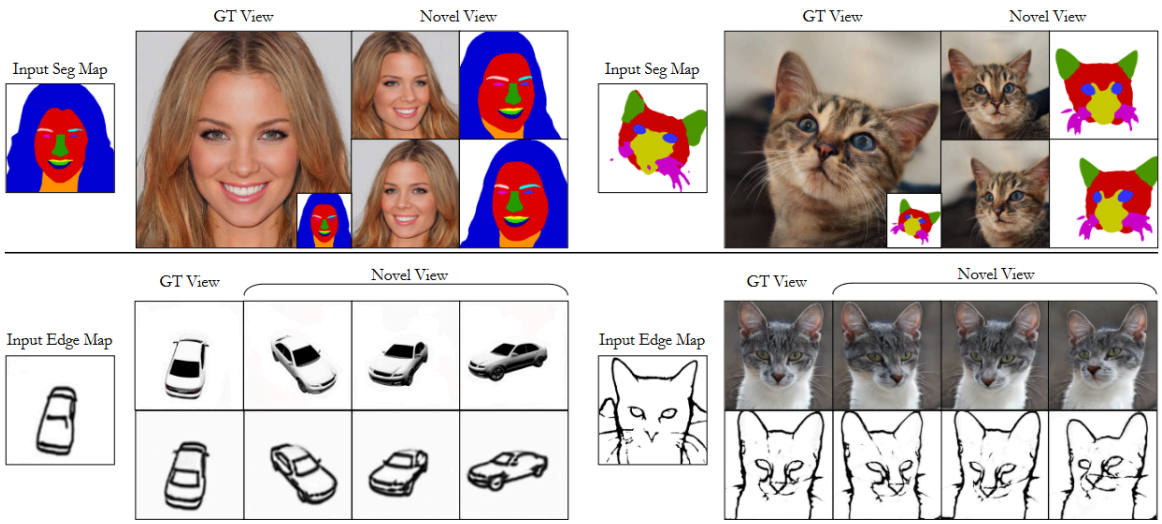


Figure 1. Given a 2D label map as input, such as a segmentation or edge map, our model learns to predict high-quality 3D labels, geometry, and appearance, which enables us to render both labels and RGB images from different viewpoints. The inferred 3D labels further allow interactive editing of label maps from any viewpoint, as shown in Figure 10.

(2D images to novel viewpoint transformations. Diagrams and caption Deng et al., 2023.)

The author additionally recommends continued research and investment into hardware systems and enabling technologies for advanced computing, extended reality, and perception that are worth intense focus such as sensors, emitters, FPGAs, Quantum Communications, and Memristors.

While the primary focus of this thesis is on static objects and minimizing interruptions, it contrasts sharply with approaches used in social media contexts. Arev et al. has developed a technique that utilizes footage from “social cameras”, which are focused on the activities being performed and their social weight being recorded. This method produces a coherent video that captures the personal and significant aspects of events as seen through the eyes of the participants, adding a layer of personalized narrative to the footage (Arev et al., 2014). An opposite system would be an interesting test for expedited 3D model generation from assorted input data.

Video Motion-based Subdivision uses pixel shift-based subdivision of video footage and would be able to isolate the most stable and clear frames for processing. By establishing specific criteria for camera motion and style, based on both demonstrated experience and prior research, we can effectively filter out unsuitable video segments or images. This fine-tuning would isolate only the most relevant clips used for further analysis, optimizing both the processing time and the quality of the resulting models.

Further considerations

When utilizing media from platforms such as YouTube or Vimeo, it is crucial to navigate the legal landscape concerning rights and royalties. As digital platforms evolve, there is a likelihood of updated user agreements or new monetization models that could mirror existing ad revenue systems but tailored for 3D models. Such changes would enable creators to derive financial benefits from 3D models based on their original video content while ensuring compliance with copyright laws and respect for artistic rights.

Capturing high-quality 3D data involves substantial costs. These include acquiring high-quality cameras and specialized scanning equipment, costs associated with accessing different locales, significant computing resources for processing and rendering 3D models, as well as investment in sophisticated software for data manipulation and processing.

Bibliography

(Sébastien Lachambre, Sébastien Lagarde, Cyril Jover, Developer - Rendering Research- Date: 2017–06–23) https://unity3d.com/files/solutions/photogrammetry/Unity-Photogrammetry-Workflow_2017-07_v2.pdf

Unity. "Photogrammetry." Accessed May 7, 2024. https://unity.com/solutions/photogrammetry.

Luma Labs. "Home page." Accessed May 7, 2024. https://lumalabs.ai/.

RED Digital Cinema. "Color Monochrome Camera Sensors." Accessed May 7, 2024. https://www.red.com/red-101/color-monochrome-camera-sensors.

Infinadeck. "Home page." Accessed May 7, 2024. https://www.infinadeck.com/.

Ektovr, "Home page," last modified 2024, accessed May 7, 2024, https://ektovr.com/.

The Walt Disney Company. "Lanny Smoot to Be the First Disney Imagineer Inducted into the National Inventors Hall of Fame." Last modified January 18, 2024. https://thewaltdisneycompany.com/lanny-smoot-to-be-the-first-disney-imagineer-inducted-into-the-national-inventors-hall-of-fame/.

Fatahalian, Kayvon. 2016. "100 Quadrillion Live Pixels: The Challenge of Continuously Interpreting, Organizing, and Generating the World's Visual Information." Lecture, Carnegie Mellon University.

Nebel, Steve, Maik Beege, Sascha Schneider, and Günter Daniel Rey. "A Review of Photogrammetry and Photorealistic 3D Models in Education From a Psychological Perspective." Frontiers in Education 5 (2020). https://doi.org/10.3389/feduc.2020.00144.

Barrile, Vincenzo & Fotia, Antonino. (2021). A proposal of a 3D segmentation tool for HBIM management. Applied Geomatics. 14. 10.1007/s12518-021-00373-4.

Pozo, Albert Parra, Michael Toksvig, Terry Filiba Schrager, Joyce Hsu, Uday Mathur, Alexander Sorkine-Hornung, Rick Szeliski, and Brian Cabral. "An Integrated 6DoF Video Camera and System Design." *ACM Transactions on Graphics* 38, no. 6 (November 2019): Article 216. https://doi.org/10.1145/3355089.3356555.

Statham, Nataska & Jacob, João & Fridenfalk, Mikael. (2020). Photogrammetry for Game Environments 2014-2019: What Happened Since The Vanishing of Ethan Carter. https://www.researchgate.net/publication/345995149_Photogrammetry_for_Game_Environments_2014-2019_What_Happened_Since_The_Vanishing_of_Ethan_Carter

Politopoulos, Aris, Angus A. A. Mol, Krijn H. J. Boom, and Csilla E. Ariese. 2019. "'History Is Our Playground': Action and Authenticity in Assassin's Creed: Odyssey." Advances in Archaeological Practice 7, no. 3 (August): 317-323. https://doi.org/10.1017/aap.2019.30.

Bodini, Aimone, Federico Colecchia, Arthi Manohar, David Harrison, and Vanja Garaj. "Using Immersive Technologies to Facilitate Location Scouting in Audiovisual Media Production: A User Requirements Study and Proposed Framework." Multimedia Tools and Applications 82 (2023): 12379-12400. https://doi.org/10.1007/s11042-022-13680-8.

Shapiro, Nicole. "Behind the Scenes: Product Spec Commercial with a Robot Arm." YouTube video, posted May 2, 2024. Accessed May 7, 2024. https://www.youtube.com/watch?v=Y7uBLGDEB-Y.

Stobiecka, Monika. "Archaeological Heritage in the Age of Digital Colonialism." Archaeological Dialogues 27 (2020): 113-125. https://doi.org/10.1017/S1380203820000239.

Lin, Chen, Chaoyang Wang, and Simon Lucey. "SDF-SRN: Learning Signed Distance 3D Object Reconstruction from Static Images." ArXiv, (2020). Accessed May 7, 2024. /abs/2010.10505.

Panoptic Studio: A Massively Multiview System for Social Motion Capture
Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, Yaser Sheikh In ICCV 2015.

Sep 28, 2023. Mark Zuckerberg: First Interview in the Metaverse | Lex Fridman Podcast #398 https://youtu.be/MVYrJJNdrEg?si=pSwwa90U0ER8J2f-

Hentschker, Claire. "Shining360°." YouTube video, 30:01. May 31, 2016. https://youtu.be/AupAFblRwgY?si=e6yG3wlaEj9m1R9a.

Jiao, Shaohui, Yuzhong Chen, Zhaoliang Liu, Danying Wang, Wen Zhou, Li Zhang, and Yue Wang. "Photo-Realistic Streamable Free-Viewpoint Video." In *ACM SIGGRAPH 2023 Posters*, Article 32. Los Angeles, CA, USA: Association for Computing Machinery, 2023. https://doi.org/10.1145/3588028.3603666.

Huang, Jingwei, Zhili Chen, Duygu Ceylan, and Hailin Jin. "6-DOF VR Videos with a Single 360-Camera." In *2017 IEEE Virtual Reality (VR)*, March 18-22, 2017, Los Angeles, CA, USA.

Wei, Yu-Szu, Xing Wei, Shin-Yi Zheng, Cheng-Hsin Hsu, and Chenyang Yang. "A 6DoF VR Dataset of 3D Virtual World for Privacy-Preserving Approach and Utility-Privacy Tradeoff." In *Proceedings of the 14th ACM Multimedia Systems Conference (MMSys '23)*, 444-450. Vancouver, BC, Canada: Association for Computing Machinery, 2023. https://doi.org/10.1145/3587819.3592557.

Dąbała, Ł., M. Ziegler, P. Didyk, F. Zilly, J. Keinert, K. Myszkowski, H.-P. Seidel, P. Rokita, and T. Ritschel. "Efficient Multi-image Correspondences for On-line Light Field Video Processing." In *EUROGRAPHICS 2016*, edited by E. Grinspun, B. Bickel, and Y. Dobashi, 35, no. 7. 2016.

Murgia, Francesca, and Daniele Giusto. "Low Cost Plenoptic Camera for 3D Modelling." In *Proceedings of the 24th Telecommunications Forum TELFOR 2016*, November 22-23, 2016, Belgrade, Serbia.

"Fraunhofer IIS Introduces Light-Field Plug-In Suite for The Foundry's Nuke." Erlangen/Amsterdam, NL, September 9, 2015, IBC, Hall 8, Booth B.80.

Rangappa, Shreedhar. "Absolute Depth Using Low-Cost Light Field Cameras." Loughborough University, January 2018.

Anuar, Syameer Faiz Khir, Amalin A'ishah Mohd Nasir, Suhaibah Azri, Uznir Ujang, Zulkepli Majid, Miguel González Cuétara, and Guillermo De Miguel Retortillo. "3D Geometric Extraction Using Segmentation for Asset Management." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIV-4/W3-2020 (2020): 61-69. https://doi.org/10.5194/isprs-archives-XLIV-4-W3-2020-61-2020.

Lin, Tsung-Yi, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. "Microsoft COCO: Common Objects in Context." *arXiv preprint arXiv:1405.0312* (2014). https://doi.org/10.48550/arXiv.1405.0312.

Deng, Kangle, Gengshan Yang, Deva Ramanan, and Jun-Yan Zhu. "3D-aware Conditional Image Synthesis." *arXiv preprint arXiv:2302.08509* (2023). https://doi.org/10.48550/arXiv.2302.08509.

Li, Lin, Fan Yang, Haihong Zhu, Dalin Li, You Li, and Lei Tang. "An Improved RANSAC for 3D Point Cloud Plane Segmentation Based on Normal Distribution Transformation Cells." Edited by Jixian Zhang, Xiangguo Lin, Guoqing Zhou, and Prasad S. Thenkabail. Wuhan

University, Wuhan, China. *Collaborative Innovation Centre of Geospatial Technology.* Published May 3, 2017. Received March 13, 2017; accepted April 30, 2017.

Ogawara, Koichi, Yasufumi Tanabe, Ryo Kurazume, and Tsutomu Hasegawa. "Detecting Repeated Motion Patterns via Dynamic Programming Using Motion Density." In *Proceedings of the IEEE International Conference on Robotics and Automation*, 1743-1749. 2009. https://doi.org/10.1109/ROBOT.2009.5152643.

Forcadel, Nicolas, Carole Guyader, and Christian Gout. "Generalized Fast Marching Method: Applications to Image Segmentation." *Numerical Algorithms* 48 (2008): 189-211. https://doi.org/10.1007/s11075-008-9183-x.

Monumental Labs. "Home page." Accessed May 7, 2024. https://www.monumentallabs.co/.

Penguin Walk - 4K Japan Travel. 2023. "[4K Japan Walking] Strolling in the back alleys where the Enoden runs in Kamakura," Video 00:27-14:42. https://youtu.be/UgYPErtoljI?si=XLwGLhUhdtNzP_0P

Rambalac. 2023. "4K・Japan - Tohoku journey in Noheji town・4K HDR" Video. https://www.youtube.com/watch?v=YXTkCrHdbpA

Nolan Fisher "thePOVchannel". 2024 "We made a 3D scan of a Ruin." Video, https://www.youtube.com/watch?v=IHlbbfTqxi4

NASA. "Mars Resources." Accessed May 8, 2024. https://science.nasa.gov/mars/resources/?search=curiosity+rover&types=images&content_list=true.

Esmaeili, Human, and Harold Thwaites. "Virtual Photogrammetry." In *Proceedings of the Academic Conference on Digital Media*, Sunway University, Malaysia, 2016. https://doi.org/10.1109/VSMM.2016.7863153.

Bradski, G. "The OpenCV Library." *Dr. Dobb's Journal of Software Tools* (2000).

von Stackelberg, Peter & McDowell, A.. (2015). What in the world? Storyworlds, science fiction, and futures studies. 20. 25-46. 10.6531/JFS.2015.20(2).A25.

Mildenhall, Ben, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis." In

Proceedings of the European Conference on Computer Vision (ECCV), 2020. https://arxiv.org/abs/2003.08934.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," ECCV 2020 (oral), https://doi.org/10.48550/arXiv.2003.08934.

Deng, Kangle, Gengshan Yang, Deva Ramanan, and Jun-Yan Zhu. "3D-aware Conditional Image Synthesis." arXiv, 2023. https://arxiv.org/abs/2302.08509.

Arev, Ido, Hyun Soo Park, Yaser Sheikh, Jessica Hodgins, and Ariel Shamir. "Automatic Editing of Footage from Multiple Social Cameras." ACM Transactions on Graphics 33, no. 4 (2014). https://dl.acm.org/doi/10.1145/2601097.2601198

Additional sources by category:

1. Computer Vision, Imaging, and 3D Modeling:

Aitkenhead, Matt, Clare Cameron, Graham Gaskin, Bastien Choisy, Malcolm Coull, and Helaina Black. "Digital RGB Photography and Visible-Range Spectroscopy for Soil Composition Analysis." Geoderma 313 (March 2018): 265–75. https://doi.org/10.1016/j.geoderma.2017.11.020

Armeni, Iro, Zhi-Yang He, Amir Zamir, Junyoung Gwak, Jitendra Malik, Martin Fischer, and Silvio Savarese. "3D Scene Graph: A Structure for Unified Semantics, 3D Space, and Camera." In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 5663–72. Seoul, Korea (South): IEEE, 2019. https://doi.org/10.1109/ICCV.2019.00576

Bolan, Jeffrey, Elise Hall, Chris Clifford, and Brian Thurow. "Light-Field Imaging Toolkit." SoftwareX 5 (2016): 101–6. https://doi.org/10.1016/j.softx.2016.05.004

Keselman, Leonid, and Martial Hebert. "Approximate Differentiable Rendering with Algebraic Surfaces." In Computer Vision – ECCV 2022, edited by Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, 13692:596–614. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022. https://doi.org/10.1007/978-3-031-19824-3_35

2. Design, Engineering, and Construction:

Demaid, Adrian, and John Zucker. "Prototype-Oriented Representation of Engineering Design Knowledge." Artificial Intelligence in Engineering 7, no. 1 (January 1992): 47–61. https://doi.org/10.1016/0954-1810(92)80006-C

3. Energy and Robotics:

Boukoberine, Mohamed Nadir, Zhibin Zhou, and Mohamed Benbouzid. "A Critical Review on Unmanned Aerial Vehicles Power Supply and Energy Management: Solutions, Strategies, and Prospects." Applied Energy 255 (December 2019): 113823. https://doi.org/10.1016/j.apenergy.2019.113823

4. Philosophical and Cognitive Studies:

Mitali, Banerjee, and Paul L. Ingram. "Fame as an Illusion of Creativity: Evidence from the Pioneers of Abstract Art." SSRN Electronic Journal, 2018. https://doi.org/10.2139/ssrn.3258318

5. Data and Information Retrieval:

Kowalski, Gerald. Information Retrieval Architecture and Algorithms. Boston, MA: Springer US, 2011. https://doi.org/10.1007/978-1-4419-7716-8

Sakka, Amir, et al. "A Profile-Aware Methodological Framework for Collaborative Multidimensional Modeling." Data & Knowledge Engineering 131–132 (January 2021): 101875. https://doi.org/10.1016/j.datak.2021.101875.