A Trade-offs Characterization of Accessible 3D Foot Scanning with Model Reconstruction from Photogrammetry

Joshua Chiu
University of British Columbia
joshchiu@student.ubc.ca

ABSTRACT

The rapid evolution of 3D Photogrammetry has revolutionized the process of capturing and modeling physical objects as it extracts three-dimensional information from multiple two-dimensional images. Its efficiency with a straightforward user experience and a low equipment requirements allow many industries to revolutionize and 3D foot reconstruction is no different. The implications spans across the medical and consumer industry from diagnosing foot posture and morphology, to custom orthotics to consumer goods personalization. With smartphones, apps can perform a 3D reconstruction of the foot with tolerances of less than a centimeter in less than ideal conditions, completely on-device. The different camera systems on smartphones varies with the number of cameras and any auxiliary sensors such as LiDAR or dot projectors, can affect the reconstructed model. Therefore, it is needed to find a new balance in the trade-off between the precision and accuracy of the reconstructed model and the ubiquity and accessibility.

KEYWORDS

photogrammtery, model reconstruction

ACM Reference Format:
Joshua Chiu. 2024. A Trade-offs Characterization of Accessible 3D Foot Scanning with Model Reconstruction from Photogrammetry. In *Trade-offs in Designing Computer Systems, 2024, Vancouver, BC, Canada.* ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

The field of 3D Photogrammetry has and continues to evolve quickly transforming the process for capturing and modelling physical environments. From the early days of photography, the concept of extracting three- dimensional information from multiple two-dimensional images laid the foundation for modern structure reconstruction techniques. The proliferation in the use of digital cameras enables the analysis of a series of photos captured from different perspectives. Combined with modern computing techniques, this allows the recreation of a 3D model in a process known as photogrammetry [9].

The term photogrammetry itself refers to obtaining or making measurements from photographs, and 3D Photogrammetry specifically involves the reconstruction of three-dimensional structures from a series of images from different perspectives known as Multi-View Stereo (MVS) [15]. The advent of computer vision and available mobile computational power, and onboard accelerators has allowed greater capabilities in our handheld mobile phones. One key innovation is the application of Structure from Motion (SfM). This technique reconstructs 3D structures by analyzing the movement of the camera between images. The MVS and SfM pipeline has enabled the creation of detailed and accurate models from a collection of photographs independent of camera calibration and parameters [13]. There is no longer a need to carefully calibrate the lens and obtain a ground truth value prior to a scan. This facilitates the ease of use and portable nature of scanning in a non-controlled environment, greatly increasing the possible applications of this technology [12].

2 BACKGROUND

In recent years, the proliferation of consumer-level smartphones equipped with increasingly sophisticated camera systems has made the usage of photogrammetric techniques much more accessible to the average consumer [12]. Further, powered by the virtual and augmented reality advances, the technology incorporated into each mobile device is gaining more features related to 3D processing. One such example, monocular and stereo cameras that are integrated into these devices have expanded the scope of computer vision applications, allowing individuals to engage in 3D modeling with relative ease [2]. This accessibility has the potential to broaden the field's impact, from professional surveying to commodity customization to augmented reality experiences [13].

While 3D Photogrammetry has improved significantly and continues to develop, it exists within a landscape of more establish reconstruction technologies [10]. Laser scanning with point cloud methodologies, considered as industry standards, offer high accuracy and precision but often come with

CPEN 534, 2024, Vancouver, BC, Canada
2024.

higher costs ranging from 6,000 to over 20,000 USD in initial equipment costs in addition to operational complexities requiring trained personnel for function [14]. The ongoing research and development in 3D Photogrammetry continuously address these challenges, seeking to strike a balance between flexibility, accessibility, and accuracy [2].

The problem of 3D foot reconstruction presents novel and unique constraints to the traditional photogrammetric techniques [12]. The human foot is a complex and dynamic non-rigid body with curves, creases, and varying contours leading to shadows. Further, the skin surface may may have reflectivity differences on a range of factors (such as skin shade, sweat, surface texture, etc.) which reflect light differently, leading to inconsistencies in the photographs. All these factors can potentially form a challenging environment to recreate an accurate and cohesive 3D model [2].

The field of 3D Photogrammetry faces critical challenges as it strives to deliver a balance of the demand for accurate but accessible three-dimensional modeling techniques [4]. While advancements in mobile computing accelerated SFM and the utilization of smartphone cameras have expanded accessibility, the lack of a review of the trade-offs in equipment specifications, reconstructed model accuracy, and user accessibility hampers informed decision-making. Furthermore, the absence of a comparison with industry-standard technologies and professional methodologies impede the identification of optimal solutions. This paper seeks to address these gaps by conducting a simplified initial investigative review, aiming to quantify and analyze the trade-offs inherent in various 3D Photogrammetry techniques, particularly in the context of static foot tracking and modeling applications. This will provide potentially valuable insights for practitioners, researchers, and industries seeking an efficient and cost-effective three-dimensional reconstruction, tracking and modelling solution.

3 DEMAND FOR PERSONALISATION

In the evolving landscape of spatial data acquisition and its large impact on various sectors, coupled with a rise for consumer demand for personalization in goods, it is only natural that tailored fitting for clothing becomes popular [11]. One particular example is that online shopping has introduced various business models, but for items like clothing and shoes, which need to align with personal physical characteristics, finding suitable products remains a persistent challenge. Innovative solutions have emerged to address this issue. One such example is WANNABY, which employs mobile augmented reality (AR) applications to enable users to virtually try on shoes, jewelry, and nail polish. These applications create a visual representation of how the commodities would look when worn by the user, bridging the gap between the virtual and physical worlds. In 2017, Invertex developed an in-store 3D foot scanner where customers' feet are scanned, and a 3D foot model is generated and sent to their mobile phones. A system then guides consumers to discover the best-fit shoes using a matching engine. Nike recognized the potential of this technology and acquired Invertex in 2018. Building on this progress, Nike launched the Nike Fit mobile application in May 2019, allowing users to scan their feet at home and accurately determine their ideal shoe size and fit. These advancements in accessible and accurate 3D Photogrammetry highlight the effect of enhancing the online shopping experience and addressing the challenges of fit testing especially for a shoe [2].

As industries and research domains increasingly demand accurate and efficient methods for three-dimensional modeling, understanding the trade-offs and optimizing the use of available techniques becomes paramount. This is especially important as this technology is expected to used by the end user who may not fully understand the technology and expects resiliency in detection.

For fitting purposes, at present, most companies still use traditional measurement methods which are manual and requires contact. There are two categories which are common in use today. The first is taking manual measurement using a cloth ruler, steel ruler or other tools. This leads to errors in the measurement position or technique and is inconsistent between users. The second is the foot scanner available in stores, which often uses laser scanning where laser lines are emitted and a series of cameras from a known position are used to collect multiple frames of images from different angles to obtain an accurate foot model. This equipment, however, is expensive, bulky, has poor portability and high maintenance cost should service be required [10, 12]. With image data being more available and with higher quality, many efforts are spent on study how to effectively and efficiently recreate this model in ideal lab conditions [7, 9]. However, they did not consider the deployment of this technology to the end user.

4 UTILIZING FUNCTIONALITY IN OUR POCKETS: THE SMARTPHONE

The widespread availability of consumer-level smartphones equipped with advanced cameras presents an unique opportunity to popularize 3D Photogrammetry. Understanding the capabilities and limitations of such devices for this purpose is crucial, as it can open doors for individuals and organizations with varying levels of resources to engage in 3D modeling projects. This aligns with the broader trend of making sophisticated technologies accessible to a wider audience [11]. However, mobile phones vary greatly in their abilities and auxiliary sensors. Flagship devices are often equipped with a multi camera system with other depth sensing sensors, whereas other devices may only have a single low resolution camera [6].

Alongside, mobile computing is becoming more and more capable and powerful, allowing for on-demand, always available computing power. The performance of camera systems on mobile devices has been the focus of development over recent years, with a sharp increase in camera sensor quality (increasing pixel density, pixel count, low light performance, etc.), camera count, and augmented data capture with efficient image processing pipelines. Vision technologies are thus becoming more integrated in mobile devices [4, 11].

With a rise in consumer awareness of security and privacy, the implementation of data laws and general heightened senses of data management, it is increasingly important to evaluate this technology with on device processing as opposed to uploading all captured images for cloud processing. This bounds our reconstruction in a timely manner to the available resources onboard.

With the variety of smartphones available in the market, a choice was made to evaluate using the Apple iPhone lineup due to the available existing implementation built for Apple's custom SoCs, their large market share of the in terms of device configuration, and their consistency across the entire lineup.

5 SETTING BOUNDS FOR THE EXPERIMENT

The primary methodology here is to implement and evaluate the photogrammetric techniques under different environmental conditions. To accurate represent something that can be easily performed by a user at home without the sophisticated knowledge or an advanced understanding on what is optimal for a photogrammetry system, the choice was made to use existing guided capture algorithms where it returns feedback and guides the user to capture. Further, as the scope have been limited the ecosystem of Apple Mobile Devices, testing will occur on an iPhone SE, iPhone 12 and iPhone 14 Pro, each representing a different combination of camera technology. These devices will be evaluated with the following implementations on the resiliency and model accuracy.

- **Object Capture** Presented in Apple WWDC '23, this technology generates a 3D model from a series of images captured from different angles fused with the LiDAR sensor on Apple's own Object Capture API as part of RealityKit, and thus will be the best case given the fully integration optimization in hardware and the Apple refined guided capture. Available for both macOS and iOS, only iOS will be evaluated (to conform to our on device processing restriction) for model accuracy and resiliency in different lighting conditions [6].
- **Scaniverse 2.0** Developed under Niantic (developer of Pokémon Go and a leader in augmented reality), Scaniverse is the arguably the industry standard in 3D Scanning App with comprehensive device support across different devices on varying camera systems and resolutions including fusion with a LiDAR or TrueDepth (Dot Projection) technology employed by Apple. Across devices, we can evaluate the final reconstructed geometry again in the various conditions on the specific MVS SfM pipeline [3]. It should be noted that Scaniverse is a closed source implementation, but the cross platform support and optimization are the important factors.

Each technique will be tested with the direct guidance from the app prompts for the best possible reconstruction of the model. This approach gives the maximum likelihood for detail and consistency in the reconstruction process, as the goal is to achieve the most accurate and high-fidelity model representation.

The plastic anatomical foot model being evaluated has a known 3D ground truth. This will be the target object being reconstructed in all scenarios. This model has been directed to by the Sensorimotor Physiology Laboratory Team led by Dr. Jean-Sébastien Blouin. This is the best alternative as testing on real humans require the approval of the ethics board. As this is a static rigid body model, this will not achieve a comprehensive representation of the skin factors and the under arch representation, but this should serve as an adequate initial understanding of these technologies.

In essence, the techniques and implementation evaluated and their respective representation are

- **Object Capture on iPhone 14 Pro** Stereoscopic Photogrammetry with LiDAR Sensor and Gravity Data, Best Available Technology
- **Scaniverse on iPhone SE** Industry Standard Monoscopic Photogrammetry
- **Scaniverse on iPhone 12** Industry Standard Stereoscopic Photogrammetry
- **Scaniverse on iPhone 14 Pro** Industry Standard Stereoscopic Photogrammetry with LiDAR Sensor

Each of the above setup will be tested in the two environments representative of an ideal scenario, room in bright light (colour temperature daylight 6500k) with no shadows cast on itself and a room in warm white (colour temperature 2500k) with some shadows cast on itself and unevening lighting.

6 EVALUATION CRITERIA

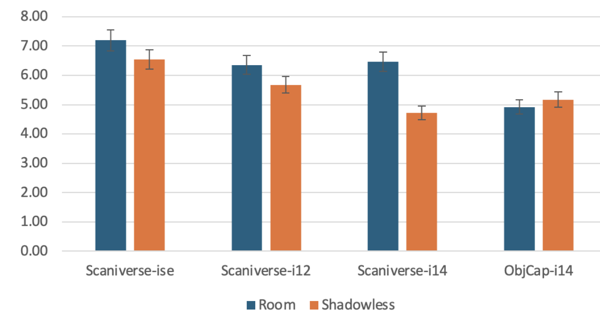
The criteria for evaluating each reconstructed model has the focus on accuracy compared to the ground truth and observation of how and the severity of each varying factor affects the accuracy. Since each tradeoff is difficult to independently evalute, assumptions will be made between two system that closely matches and conclusions being drawn for each category of LiDAR sensor with stereoscopic vision, pure stereoscopic vision and monocular vision and the effects of the lighting environment. The element of cost will not be considered since it is not directly representative of the camera system and can not be easily decoupled from all the features of a phone.

There exists many methods for evaluating 3D scan accuracy and resolution in relation to a known ground truth, with each bringing a different insight. First, we must define accuracy, and that is how close the measurement is to its true or tolerable value. And thus, we will be evaluating single scan accuracy with hausdorff and mean millimeter deviation.

The generated models will be compared against the ground truth file with hausdorff distance and mean error millimeter deviation, referred to as mean error. The Hausdorff distance serves as a metric for quantifying the dissimilarity between two sets by identifying the maximum distance from any point in one set to its nearest point in the other set. Conceptually, it represents the maximum of distances between the elements of two sets, reflecting the maximum separation between them. Analogously, envisioning two geometric shapes, the Hausdorff distance corresponds to the longest stretch of a tether that must span between the shapes to establish a connection, thus representing the extent of their spatial disparity. Intuitively, it is to find one point on the reconstructed model that is farthest away from the groundtruth model.

Millimeter deviation represents the absolute maximum deviation from the ground truth. This serve as a useful standard unit of measurement for directly assessing the deviation of real values at this scale, thereby representing a method for quantifying absolute accuracy. This value simply represents the mean deviation for all point has from the corresponding ground truth point. An artifact of this is that if the scanned model is 2mm longer than the ground truth, the millimeter deviation could range between 1mm and 2mm as both the farthest forward and farthest backward point could each have an error of 1mm or just one of the ends has an error of 2mm, assuming that the 1mm error axis is a the plane of highest error. To counter this, we calculate and align the centroid to balance this mean error in all models.

Figure 1: Mean Absolute Error in Reconstructed Model



7 RESULTS AND DISCUSSION

Through the evaluation of each method of photogrammetry from above, we are able to see repeated trends across the systems in their accuracy and variance of error. In general, the reconstructed model accuracy improves with more cameras and more auxiliary data (LiDAR, IMU, etc.) as can be seen in Figure 1. Each parameter is discussed below

7.1 Lighting Conditions

Through the conduction statistical paired t-tests with an alpha of 0.05, it can confidently said (with 95% certainty) that the lighting condition as specified does not affect the reconstructed model despite the greater mean error in the room lighting reconstruction as we are unable to reject the null hypothesis. Further, we can see that with more auxiliary data, the lighting condition no longer affects the accuracy of the reconstructed model.

7.2 Stereoscopic Vision

Through the conduction statistical paired t-tests with an alpha of 0.05, it can confidently said (with 95% certainty) that the increased number of cameras from one to two does affect the reconstructed model as seen with the lower errors in the reconstruction and the null hypothesis can be rejected. A camera system with stereoscopic vision can accurately detect depth similar to human vision, this is intuitive and aligns with the expectations.

7.3 Auxiliary Sensors

Through the conduction statistical paired t-tests with an alpha of 0.05, it can confidently said (with 95% certainty) that the addition of auxiliary data of a LiDAR sensor and a positional IMU with gravity data does affect the reconstructed model as seen with the lower errors in the reconstruction and the null hypothesis can be rejected. This is again intuitive as the positioning of the camera can be detected and assist in the perception of depth data.

Table 1: Reconstruction Mean Hausdorff Distances

System tested	Hausdorff	Stddev
Monoscopic Vision (SL[a])	25.93mm	5.62mm
Stereoscopic Vision (SL)	24.40mm	3.53mm
Stereoscopic LiDAR (SL)	22.09mm	1.83mm
Stereoscopic LiDAR IMU (SL)	18.70mm	2.19
Monoscopic Vision (RM[b])	27.09mm	7.77mm
Stereoscopic Vision (RM)	20.56mm	1.42mm
Stereoscopic LiDAR (RM)	22.96mm	4.47mm
Stereoscopic LiDAR IMU (RM)	18.82mm	1.91mm

Note: A table was chosen instead of a figure to prevent the bias in visualizing tail error.

[a]Shadowless lighting conditions with same light temperature from all directions to minimize form and cast shadows

[b]Room Lighting where there are form and cast shadows

8 RELATED WORKS

Three-dimensional computer vision encompasses the perceiving, recognizing, and comprehending three-dimensional scenes within the physical environment through computational means. This field utilizes cameras to capture images or videos of objects from various viewpoints, and through analyzing the projection correspondence between objects and their respective images, as well as the inter-image matching relationships derived from multiple viewpoints, objects are reconstructed by amalgamating positional, geometric, and contextual information. Essential components of three-dimensional computer vision systems encompass camera calibration, feature extraction, stereo matching, and three-dimensional reconstruction techniques [12].

Three-dimensional foot reconstruction presents novel and unique constraints to the traditional photogrammetric techniques. A majority of previous research efforts focuses large static reconstruction of landscapes and architecture, and with a smaller scale scale, many new challenges are presented [15]. The overall volume of the foot is small, with an uneven surface, and it is not a rigid body with an non-obvious texture. Combined with the under-arch gap that is critical to orthopedics, there remains many challenges to mitigate.

Previous studies and related works in the field of photogrammetry have often overlooked the constraints imposed by mobile computing platforms. While these investigations have made significant strides in advancing the capabilities of photogrammetric reconstruction techniques, they have typically been conducted in environments where computational resources are abundant and readily available [1, 2, 5, 7–9, 12–16]. However, the widespread proliferation of mobile devices as tools for capturing imagery has brought to light the inherent limitations of these platforms in terms of processing power, memory capacity, and computational efficiency [11]. Consequently, there exists a critical gap in the literature regarding the adaptation of photogrammetric methodologies to the constraints imposed by mobile computing environments. Understanding this gap is imperative for the development of practical and efficient photogrammetric solutions that cater to the growing demand for these mobile-centric applications.

9 CONCLUSION

Foot measurement is crucial for multiple industries, yet traditional methods such as ruler measurements or foot scanners present challenges like large measurement errors or high costs and inconvenience. This paper presents tests conducted on different mobile devices in different environmental conditions for close-range multi-image photogrammetry, with a focus on reconstructing an accurate 3D solid model of a human foot.

Utilizing tools outlined above, the comparison results reveal notable disparities among the camera system. Apple's Object Capture SDK demonstrates the best model reconstruction with the lowest error and variance of less than a centimeter. Monoscopic camera capture also demonstrated the worse reconstructed model with surface noise and inaccurate scale recapturing due to errors in calculating depth data. Despite these differences, all options demonstrate operational functionality in photogrammetry with no failed pipelines and arguably better performance than measuring with a ruler. Stereoscopic vision outperforms microscopic intuitively and adding auxiliary data such as IMU and accelerometers further improve the reconstruction.

By enabling non-experts to conduct measurements and leveraging photogrammetric and image processing techniques, this technology has the potential impact to reach beyond with the successful resolution of these constraints.

ACKNOWLEDGMENTS

This project was made possible in part by the support of Dr. Jean-Sébastien Blouin, Professor in the UBC School of Kinesiology and his team at the Sensorimotor Physiology Laboratory for their time and resources in understanding the industry best practices and current standards.

The diction and sentence construction in the technical writing in this paper has been assisted in part by large language models tools including ChatGPT.

REFERENCES

[1] Jasim Ahmed Ali AL-Baghdadi, Albert K. Chong, Peter Milburn, and Richard Newsham-West. Correlating video-captured 3d foot model with foot loading during walking. In *2013 IEEE International Conference on Signal and Image Processing Applications*, pages 161–166, 2013.

[2] Wei-Ta Chu and Cheng-Hsi Lin. 3d foot model construction from photos, model segmentation, and model alignment. In *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, pages 349–353, 2019.

[3] Thomas Dickinson, Jessica Liu, and Mark Roughley. The affordances of 3d imaging modalities for forensic facial identification, Nov 2022.

[4] Dieter Fritsch and Miguel Syll. Photogrammetric 3d reconstruction using mobile imaging. 2015.

[5] Carsten Griwodz, Simone Gasparini, Lilian Calvet, Pierre Gurdjos, Fabien Castan, Benoit Maujean, Gregoire De Lillo, and Yann Lanthony. Alicevision Meshroom: An open-source 3D reconstruction pipeline. In *Proceedings of the 12th ACM Multimedia Systems Conference - MMSys '21*. ACM Press, 2021.

[6] Apple Inc. Scanning objects using object capture, 2023. https://developer.apple.com/documentation/RealityKit/guided-capture-sample.

[7] Da-Un Jung, Yong-In Yun, and Jong-Soo Choi. A new interface using image-based foot tracking for motion sensing devices. *IEEE Transactions on Consumer Electronics*, 58(3):803–809, 2012.

[8] Shian-Ru Ke, Jenq-Neng Hwang, Kung-Ming Lan, and Shen-Zheng Wang. View-invariant 3d human body pose reconstruction using a monocular video camera. In *2011 Fifth ACM/IEEE International Conference on Distributed Smart Cameras*, pages 1–6, 2011.

[9] Satoshi Kondo, Yasuhiro Akagi, and Katsuhiro Kitajima. Study on a method of generating a 3d virtual foot model with use of a measuring technique based on multiple camera image data. In *SICE Annual Conference 2007*, pages 7–11, 2007.

[10] Min-Jae Lee, Seung-Hae Baek, and Soon-Yong Park. 3d foot scanner based on 360 degree rotating-type laser triangulation sensor. In *2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, pages 1065–1070, 2017.

[11] Silvia Menato, Giuseppe Landolfi, Marino Alge, and Marzio Sorlini. Empowering widespread shoe personalization via a 3d foot scanning app. In *2014 International Conference on Engineering, Technology and Innovation (ICE)*, pages 1–7, 2014.

[12] Lulu Niu, Gang Xiong, Xiuqin Shang, Chao Guo, Xi Chen, and Huaiyu Wu. 3d foot reconstruction based on mobile phone photographing. *Applied Sciences*, 11(9), 2021.

[13] Suto Setiyadi, Husneni Mukhtar, and Willy Anugrah Cahyadi. A comparative study of affordable photogrammetry software for reconstructing 3d model of a human foot. In *2021 IEEE 7th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, pages 286–291, 2021.

[14] Ali K. Thabet, Emanuele Trucco, Joaquim Salvi, Weijie Wang, and Rami J. Abboud. A dynamic 3d foot reconstruction system. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 599–602, 2011.

[15] Giuseppina Vacca. Overview of open source software for close range photogrammetry. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W14:239–245, 08 2019.

[16] Yuji Yoshida, Shunta Saito, Yoshimitsu Aoki, Makiko Kouchi, and Masaaki Mochimaru. Shape completion and modeling of 3d foot shape while walking. In *2012 International Symposium on Optomechatronic Technologies (ISOT 2012)*, pages 1–3, 2012.

A RECONSTRUCTED ARTIFACTS

In section 6, the model reconstructed are available here github.com/Joshua-Chiu/photogrammetry-characterization