Personal Time-Lapse

Nhan Tran
Cornell University
Ithaca, NY, USA
nhan@cs.cornell.edu

Ethan Yang
Cornell University
Ithaca, NY, USA
eey8@cornell.edu

Angelique Taylor
Cornell Tech
New York, NY, USA
amt298@cornell.edu

Abe Davis
Cornell University
Ithaca, NY, USA
abedavis@cornell.edu

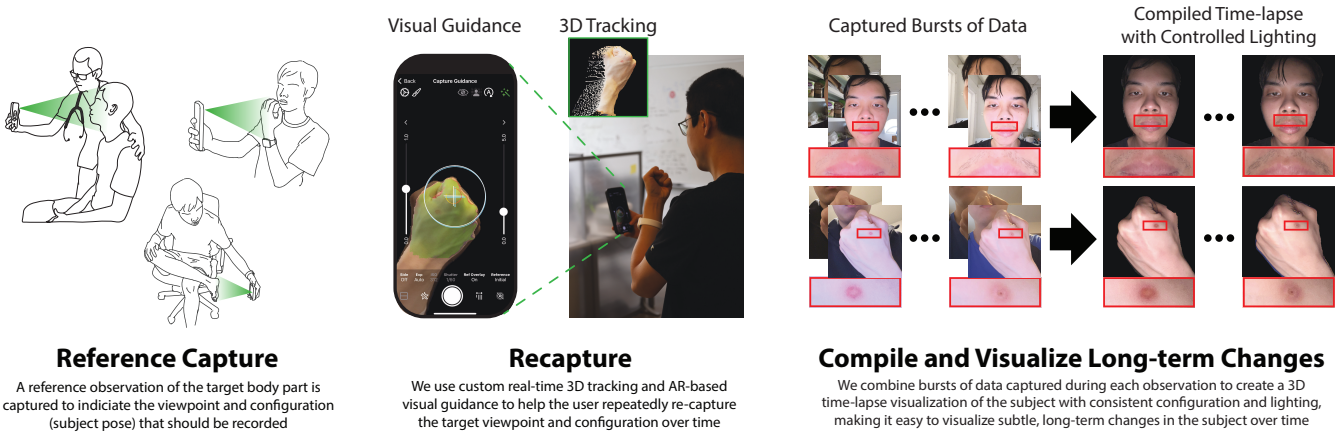

Figure 1: We present a mobile application for capturing personal time-lapse (PTL), which visualizes gradual changes of a subject over long periods of time. Recording PTL begins with the capture of a reference observation (left), which can be done by the subject of capture or by a specialist. During recapture (middle), our tool uses custom tracking and novel visual guidance to help the user reproduce the conditions of the reference observation. Users recapture new observations over an extended period of time, which we process with computational illumination to visualize the subject under controlled lighting conditions (right). Our final visualization is an interactive 3D visualization of long-term changes in the subject under consistent lighting. Our work offers a convenient way to document and visualize long-term changes in the body, with many potential applications in remote healthcare and telemedicine.

ABSTRACT

Our bodies are constantly in motion—from the bending of arms and legs to the less conscious movement of breathing, our precise shape and location change constantly. This can make subtler developments (e.g., the growth of hair, or the healing of a wound) difficult to observe. Our work focuses on helping users record and visualize this type of subtle, longer-term change. We present a mobile tool that combines custom 3D tracking with interactive visual feedback and computational imaging to capture *personal time-lapse*, which approximates longer-term video of the subject (typically, part of the capturing user's body) under a fixed viewpoint, body pose, and lighting condition. These personal time-lapses offer a powerful and detailed way to track visual changes of the subject over time. We begin with a formative study that examines what makes personal time-lapse so difficult to capture. Building on our findings, we motivate the design of our capture tool, evaluate this design with users, and demonstrate its effectiveness in a variety of challenging examples.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**.

KEYWORDS

Camera-based UIs, Virtual/Augmented Reality, Graphics / 3D

ACM Reference Format:
Nhan Tran, Ethan Yang, Angelique Taylor, and Abe Davis. 2024. Personal Time-Lapse. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3654777.3676383

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-0628-8/24/10
https://doi.org/10.1145/3654777.3676383

1 INTRODUCTION

The human body is constantly changing. Some changes are fast and temporary, like the motion involved in walking or breathing. Other changes are slower and more permanent, like those involved in growth and healing. These slower and often more subtle changes are important indicators of human health. Unfortunately, they are also some of the hardest to capture and visualize, as their short-term progress can be difficult to distinguish from transient changes

of viewpoint, body pose, and lighting. For example, as a scraped knee begins to heal, the wounded area will shrink. However, the shrinkage observed in a day may be small compared to changes in size caused by simply bending the knee, which stretches or loosens skin around the kneecap. The color of the scrape may also change over time, but that can be difficult to distinguish from the effects of viewing the knee under different lighting conditions—for example, cooler lighting in the morning versus warmer lighting at the end of the day. Controlling transient factors that can affect the appearance of a subject is especially important when capturing data for medical applications. For this reason, medical professionals employ a variety of tools and techniques to standardize the recording of patient data in clinical settings [1, 10]. However, as the recent COVID-19 pandemic highlighted, existing strategies do not extend to remote healthcare scenarios, creating an urgent need for new and more scalable tools. Our work addresses this need by developing a mobile application for capturing and visualizing long-term changes of the body. Our strategy is to help users control the conditions of repeated observations taken over long periods of time. To accomplish this, we make three key contributions:

- **Fast Geometry-Based Mobile 3D Tracking:** Existing mobile AR toolkits do not support tracking and registration for non-rigid surfaces. We implement fast geometry-based tracking on mobile devices to enable new types of real-time user feedback and guidance. While the basic algorithms behind our approach have been used for non-rigid tracking before, to our knowledge, ours is the first interactive implementation on mobile devices, which lets us explore a new space of interactive capture opportunities.
- **Augmented Reality (AR)-Based User Guidance:** Our task involves simultaneously controlling the viewpoint of capture and the configuration of the subject being captured. Building on our custom mobile tracking, we offer novel interactive guidance through the use of AR-based visual feedback.
- **Computational Imaging:** The ability to guide users back to a consistent viewpoint and subject configuration creates unique opportunities for computational imaging. For each observation, we capture a rapid burst of data that lets us control subject lighting, even across data captured in different environments.

Our resulting tool makes it easy for users to capture a subject under specified conditions over an extended period of time. We then compile these captured observations into an interactive visualization of long-term changes that we call *personal time-lapse* (PTL).

1.1 Personal Time-Lapse (PTL)

Our goal is to highlight long-term changes of a subject by visualizing it under controlled conditions over time. Our challenge then depends on the choice of subject and conditions we wish to control. For this paper, we primarily focus on subjects that are part of the capturing user's body (e.g., their face, hand, or foot), though our work can be used on other types of subjects as well (see Figure 7 and supplemental material for examples). To support a flexible range of downstream use cases, we derive the target conditions for capture from an initial reference observation, which can be captured by the user themselves or by a third party, such as a physician specifying observations for a patient to capture later from home.

We derive the conditions we wish to fix from a review of guidelines on medical and clinical photography (see Section 2.3), as well as texts on stop-motion animation [34, 38], and discussions with a professional filmmaker known for their work on time-lapse of the body [23]. We primarily focus on controlling three capture criteria for each observation:

- **Viewpoint:** The pose of the camera relative to the subject
- **Configuration:** The current shape of the subject. For example, in the case of articulated subjects like a human hand, this would include the angles of each joint.
- **Lighting:** The color, intensity, and angular distribution of light illuminating the subject.

Our challenge is to control these three factors in uncontrolled environments using only a handheld mobile phone. We developed our solution as an iOS application and show results from a user study with 14 participants demonstrating dramatic improvements in geometric accuracy for data captured with our tool of hands, feet, and faces ($p < 0.001$). We also validate our approach with two compelling longitudinal case studies. Additional results, as well as a link to our free app, can be found on our project website.

1.2 Background & Motivation

Our work has many possible downstream uses, with medical applications (e.g., wound tracking and telemedicine) being our single biggest motivation. Whereas many previous efforts in this space have focused on supporting specific clinical tasks (discussed in Section 2.3), we instead focus on more general challenges related to geometry and usability. This focus is motivated by recent trends and observations in remote healthcare, which we review here for context.

1.2.1 Patient Photography in Remote Healthcare Settings. The recent COVID-19 pandemic greatly accelerated the increasing trend of using patient-captured data in healthcare and telemedicine [13, 25, 31, 35]. This acceleration has highlighted a need for more general tools to support remote and asynchronous healthcare. In particular, mobile apps that capture data for specific diagnostic tasks tend not to generalize well, which is why standard photography tools have remained the dominant way for patients to communicate visual information to specialists in remote scenarios [32].[1] With this observation in mind, we focus our effort on developing a general tool for capturing parts of the human body, which we evaluate on the basis of usability and geometric accuracy.

1.2.2 Current Limits of 3D Tracking On Mobile Devices. A major limitation of mobile photography for remote healthcare has been the inability to determine at capture time whether a user has recorded data correctly. Guidelines for making this determination often exist (see Section 2.2), but those guidelines depend on information that cannot be evaluated with existing mobile tracking APIs. Current 3D tracking APIs on mobile phones are designed to support AR applications where the goal is to render virtual objects into real scenes. For this, 3D pose estimation is based on visual SLAM that derives 3D information from static rigid features in a scene. However, humans

[1] This includes standard camera interfaces used in apps designed for HIPAA compliance.

are neither static nor rigid, which causes this tracking to fail on most parts of the human body. Specialized human body pose estimators do exist (e.g., in Apple's ARKit and Google's MediaPipe) but are based on approximate mappings to a canonical human model derived from frequently inaccurate priors. We address these limits by developing our own 3D mobile tracking system based on a fast variant of projective ICP [18, 36] that has been shown to work well with non-rigid surfaces [14, 26]. Our custom tracking module lets us overcome various limitations inherited by previous related efforts by exploring more general interactive capture strategies that were not possible with previous tracking APIs.

2 RELATED WORK

2.1 Augmented Reality & HCI

2.1.1 Guided Photography Systems. AR-based guidance has been explored for particular photography tasks in prior work. Adams et al. [2] visualize scene coverage during panoramic image capture. Bae et al. [4] explore the problem of re-photographing older photos, mostly focusing on architecture. More recently, Kim and Lee [21] explore the use of AR-based guidance for framing photographs, and works by E. et al use AR guidance to help users apply photography principles related to framing [11], decluttering of compositions [12], and to teach users about these principles. Yan et al. [42] address handheld time-lapse capture with mobile devices, which bears some resemblance to our own task, but focuses on subjects that are farther from the camera and fixed in their environment. This avoids some of the most significant technical challenges facing PTL capture as we examine further in our formative study.

2.2 Guidelines for Medical Photography

Medical and clinical photography have come to play critical roles in healthcare. As such, several efforts have explored recommendations and guidelines for photography in different medical settings, including general guidelines [1, 5, 17] as well as those more specific to photography on mobile devices [44], photography of skin surfaces [20], and cosmetic procedures [33, 39, 40]. While specific guidelines vary across specific settings, several features are common. Our work focuses on controlling three of the most common properties described: the **viewpoint** of the camera relative to the subject, the **configuration** of the subject (often referred to as "pose"), and the **lighting**. For a more detailed discussion of the role these play in existing guidelines, we refer to the work of Davda and Pasquali [10].

2.3 Mobile Apps for Medical Photography

Several previous works have explored the design of mobile applications for aiding specific monitoring or diagnostic tasks such as the early detection of melanoma [7, 41] or tracking of diabetic foot ulcers [6, 28, 43]. Other works have focused on the secure transfer of more general clinical image data through mobile applications apps, e.g. for HIPAA compliance [24]. We refer to several related surveys for additional information [7, 15, 25, 37]. Notably, most patient-captured data used in health contexts do not come from specialized capture apps [31, 35] which motivates our general purpose approach as discussed in Section 1.2.

2.4 Real-time 3D Tracking

Real-time 3D tracking involves continuously estimating the camera's position and orientation relative to a 3D representation of the scene. Most real-time mobile 3D tracking implementations are based on multi-threaded versions of visual SLAM (e.g., [22, 29]), which rely on static, rigid visual features. This typically causes pose estimation to ignore or fail on non-rigid subjects, a limitation that is acknowledged in documentation for common mobile APIs (e.g., Apple's ARKit [3]).

An alternative approach to 3D tracking makes use of geometric data (e.g., point clouds) like that provided by a depth camera. KinectFusion [18, 30] accomplish this by using a coarse-to-fine iterative closest point (ICP) algorithm to average depth images into a volumetric signed distance function (SDF) for representing the scene. This approach is quite effective and more robust to non-rigid subjects than most visual SLAM-based approaches, but it can be expensive. The SDF requires significant memory, and their pose estimation used a highly optimized CUDA implementation of projective ICP, which is likely why there are no known open-source implementations on mobile platforms. Our tracking implementation overcomes the memory issue by only registering against a fixed reference view of the subject, and the speed issue with our own highly optimized implementation of projective ICP for mobile GPUs on iOS devices.

3 FORMATIVE RESEARCH

With the exception of photo-a-day montages that align a sequence of photos on a specific person's face (e.g. [9, 19]), time-lapse of the human body is quite rare. Our formative findings in this work suggest that this rarity stems from the difficulty of capture, which our work is meant to help address.

3.1 Expert Interview

The closest examples to our vision of personal time-lapse we could find came from a YouTube channel specializing in time-lapse. We contacted the channel's creator, who turned out to be Bálint Kolozsvári, an accomplished filmmaker known for his expertise in time-lapse [16]. Kolozsvári agreed to an interview about his work for this paper, and his full responses can be found in our supplemental material. Our most informative discussion centered around a time-lapse he created of a wounded finger healing, which has been seen by millions of viewers across different video platforms. Despite capturing this video in a carefully controlled environment with fixed studio lighting, a tripod-mounted camera, and a plaster mold to control the positioning of his hand, Kolozsvári reported that he still found controlling the precise configuration of his finger for each frame to be challenging:

> *"...I made a plaster mold for the lower part of my hand. But with that, I still had to position my finger very carefully ... At times I was able to position and take the photo in less than a minute, and some other time it took me up to 15 minutes to find the right position."*

This illustrates the need for precise guidance over the viewpoint and configuration of subjects.

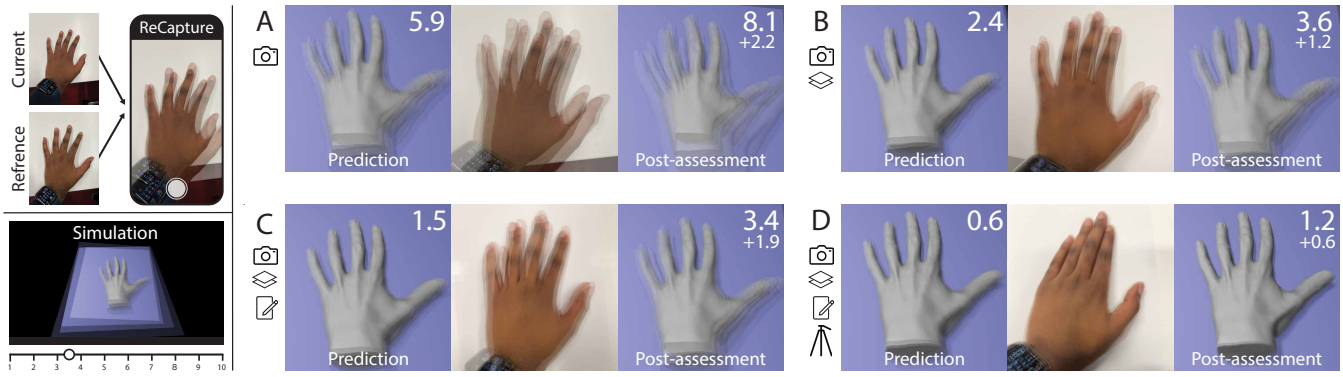

Figure 2: Formative Study: In the upper left, we show the overlay mode from ReCapture, used in Conditions B-D as a baseline, which overlays the reference image onto the current view during capture. Below it, we show our interactive simulator which visualizes the accuracy of recaptured images by blending renders of a hand from noisy camera poses where users adjust the noise variance with a slider. This allows users to score their accuracy before capturing as a *prediction*, and after recapturing as a *post-assessment* for the various conditions (Section 3.2.2) as shown on the right-hand side. For each condition, we report the average *prediction* (left) and *post-assessment* scores (right) and visualize them with our simulator as seen by users. We also include a representative result captured by a user (middle). Across all conditions, users generally scored their performance worse than their prediction, indicating they found capturing PTL harder than expected.

3.2 Formative Study

To better understand what makes capturing PTL difficult, we conducted a formative study aimed at investigating two questions:

- **RQ1:** What difficulties do users encounter when capturing PTL with existing tools, and how does the level of difficulty compare with expectations?
- **RQ2:** What impacts do different types of guidance related to viewpoint and configuration have on the accuracy of PTL capture?

Our IRB-approved study (IRB0147654) involved five participants (2 males and 3 females ages 20-30) recruited through social connections and public forums. No participant had prior experience capturing time-lapses of body parts, but two reported experience capturing time-lapses of landscapes using existing commercial tools. Each participant was compensated with money for their effort.

3.2.1 Task. Participants were tasked with capturing four small time-lapses of their hands under four different capture conditions. The conditions were randomly ordered for each participant, who then captured a reference photograph for each condition. They then cycled through the conditions, re-capturing the reference each time, until they had cycled through the conditions four times.

3.2.2 Conditions. To address RQ2, we designed our conditions to include existing tools as a baseline, then added conditions with additional guidance related to viewpoint and subject configuration. ReCapture app refers to a capture mode offered in the ReCapture Time-Lapse mobile app for iOS [42] that shows a visual overlay of the reference image being recaptured (see Figure 2 upper left):

- **Condition A:** Capture with the standard iOS camera app as a baseline.
- **Condition B:** Capture using ReCapture app.
- **Condition C:** Same as Condition B, but users are also given paper and pen that can be used to trace the configuration of their hand for guidance of later captures.
- **Condition D:** Same as Condition C, but capture is performed with a dedicated camera that is fixed to a static tripod for all captures (i.e., the viewpoint is fixed).

Conditions A and B use existing tools as baselines. Conditions C and D progressively add additional constraints, with the paper and pen added in C designed to help users control configuration, and the tripod added in D to control viewpoint. We did not include a condition exploring the third capture criteria mentioned in Section 1.1, lighting, because we planned to pursue a computational approach to control lighting.

3.2.3 Evaluation. We designed an interactive metric for evaluating the quality of captures to help standardize the meaning of our observations across users. Our metric, shown in Figure 2, is based on rendered images of a simulated hand. Users are shown a blend of images sampled with a given variance around some central reference view of the hand. The amount of variance represents the accuracy of recaptured images: setting it to zero means all images are recaptured with perfect accuracy, which results in a blended image that is perfectly sharp. Increasing the variance leads to less consistent viewpoints, which create increasingly large ghosting artifacts when blended. We let users control the amount of variance with a slider to specify different levels of capture accuracy. Before the experiment, we explained each condition to users and asked them to predict their own accuracy for each condition using our metric (*prediction*). Then, after they finished capturing data, we blended the images they captured in each condition and asked them to approximately match the slider in our metric with the variance they observed for each blend of captures (*post-assessment*). This gave us an approximate measure of perceived accuracy for each

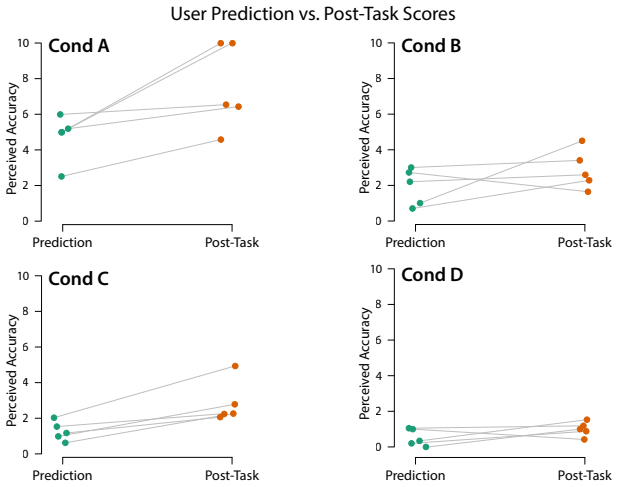

Figure 3: User Prediction vs. Post-Task Scores: The differences between the prediction and post-task assessment scores across all four conditions. Participants consistently found capturing PTL unexpectedly challenging.

condition, which we could compare with the predictions given prior to capture. We report the average prediction and post-assessment scores for each condition in the upper right-hand corners of Figure 2. We also surveyed participants on what they found challenging.

3.2.4 Findings. We observe improved overall accuracy as the level of guidance increases, with the most significant gap being between Condition A (standard camera app) and the other conditions. This indicates that offering some live visual comparison to the reference image being recaptured is very important. The simple overlay of Condition B had some issues: users reported sometimes having difficulty differentiating the overlay from the live camera image, and commented that attempting to align one local region of the hand sometimes led to misalignment in other regions. This second challenge can be attributed to the ambiguity of whether to use camera translation, rotation, or changes in configuration to align different parts of the image. Section 4 describes how our tool addresses these ambiguities with feedback based on a comparison of current and reference scene geometry.

Comparing the predicted and post-assessment accuracy of each condition (Figure 3) shows that capturing PTL is generally harder than people expect, especially when using existing tools. A paired t-test for Condition C shows that post-assessment scores increased beyond the predicted values, with $p = 0.008$, even with the added configuration guidance provided by the pen and paper.

One user noted that tapping the shutter button often caused the camera to shake, which decreased capture accuracy. Based on this observation, we implemented automatic capture triggering in our tool, which we describe in Section 4.

4 APPLICATION DESIGN

Building on our findings from the formative study, we developed a mobile-AR iOS application for capturing PTL.

4.1 PTL Data

A single PTL consists of a sequence of *samples* representing observations of the subject at different times. One sample is designated as a reference, and the criteria for capturing new samples is measured relative to this reference (i.e., users are guided to reproduce the conditions of this reference when capturing new data). Each sample consists of two RAW RGB images (explained in Section 4.7) and corresponding depth maps, as well as metadata about the time, location, and camera parameters used for each captured image.

4.2 Workflow

Our application opens to a gallery of thumbnails representing PTLs stored on the current device. From here, users can create a new PTL, or select an existing one to view or capture new samples. During capture, RGB and depth data are streamed from the front-facing sensors of the device. We feed depth images to our custom 3D tracker to compute camera pose, and visual guidance for users is overlaid on top of the live RGB camera feed, as depicted in Figure 4.

Creating New PTL: Users create new PTL by simply capturing an initial reference sample of their subject under desired conditions. The only guidance we provide for capturing this initial sample is a visualization of pixels that are too close to the camera for accurate focus or depth readings, which helps ensure the quality of captured data. After capturing a reference sample, users can optionally draw a mask on the captured image or set a range of depths to indicate what part of the captured scene is their subject (Section 4.5.1).

Recapturing Subsequent PTL Samples: Our tracker computes 3D pose by aligning the incoming depth image at each frame with the depth map of the current reference sample. We use this information to guide the user's control of the camera and subject, and to automatically trigger the capture of a new samples when certain viewpoint and configuration conditions are met. Figure 4 shows our guidance interface, which we describe later in this section.

Viewing PTL: To visualize a PTL, we first align all the collected samples with a robust offline ICP implementation, then we apply the computational illumination strategy described in Section 4.7 to calculate a time-lapse of the subject under fixed lighting conditions. Our final visualization is a 3D time-lapse of the subject under controlled lighting. Example interactive 3D visualizations can be found on our project website.

4.3 Representing Viewpoint & Configuration

Mathematically, we can define the viewpoint V_t and configuration C_t of an observation in terms of the transformation that relates our observed geometry at time t to the geometry of our reference sample. V_t represents the rigid component of this transformation, while C_t represents any residual non-rigid component of the transformation. This factorization has three key advantages. First, it aligns well with the colloquial meaning of viewpoint and configuration. Second, it is convenient to compute; we find the current viewpoint by using ICP to solve for the rigid transformation that minimizes the sum of squared distances between corresponding points in the current observed geometry and our reference sample geometry. Finally, the third advantage of our factorization has to do

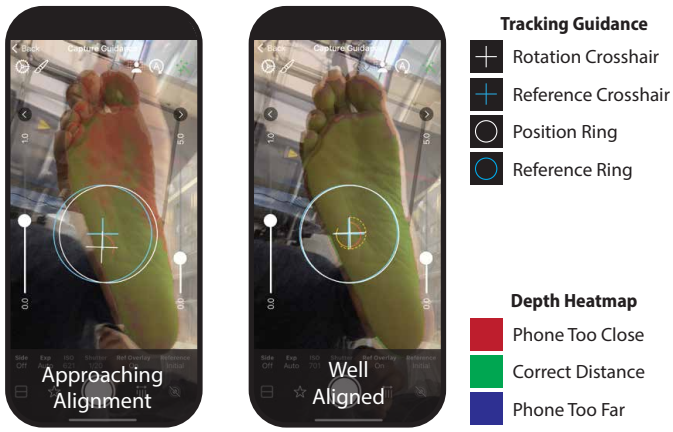

Figure 4: User Interface: Here, we show an annotated interface of what is displayed to the user during capture. Guidance for configuration is conveyed through the overlay and heatmap, while guidance for viewpoint is conveyed through the crosshairs and rings. Note that our app offers other configurable options not shown here, including the ability to hide the background.

with user guidance—more specifically, it separates alignment into a viewpoint component with fixed degrees of freedom (DOF) and a spatially-varying residual. This lets us use high-precision symbolic feedback to guide viewpoint alignment, and spatially-varying visualizations for guiding configuration.

4.4 Configuration Guidance

The degrees of freedom that govern configuration vary across different subjects, and are generally unbounded for very deformable surfaces. With this in mind, our strategy for guiding configuration focuses on spatially-varying visualizations that convey alignment on a per-pixel basis.

4.4.1 Overlay and Masking. Our formative study found that a simple overlay of the reference sample image is quite useful for guidance. With this in mind, we use such an overlay as the starting point for configuration guidance. Since each sample may be captured in a different environment with a different background, we also allow users to specify a mask of their subject, which we can use to remove the background from our overlay and pose estimation. Users can create a mask by painting it directly onto the reference subject view, and by using a slider to mask all pixels within a selected range of depths in the scene.

4.4.2 Depth Heatmap. Simple overlay guidance helps with aligning silhouettes and visible edges of objects, but is less useful for conveying misalignment in texture-less surfaces, or displacements along the optical axis of the camera. To help with this, we also use a heatmap to visualize the difference in depth between our current observation and our reference observation at each pixel. This heatmap is blended with the overlay as shown in Figure 4. The meanings of colors in this map are illustrated in Figure 5: surface points that are too close to the current camera are colored red,

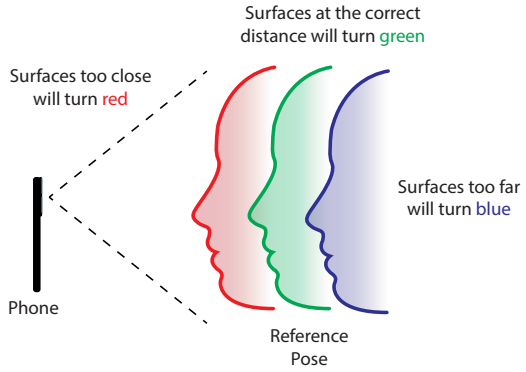

Figure 5: Depth guidance: Colors change depending on the distance between the current and reference subject.

points that are too far are colored blue, and points approaching the correct depth turn green.

4.5 Viewpoint Guidance

Viewpoints have six degrees of freedom (DOF), which can be further factored into three degrees of rotation and three degrees of translation. Previous work (e.g., Yan et al. [42]) and our formative study both found that users often struggle to differentiate between local offsets in translation and rotation, which motivates the use of visual feedback that makes this distinction clear. However, our representation of viewpoints is a relative one that does not differentiate between rigid transformations of our camera and rigid transformations of our subject, which makes decomposition into rotation and translation somewhat ambiguous. We can resolve this ambiguity by choosing an anchor point in our reference observation of the subject. By default, we choose a point along the optical axis of our reference sample view, at the average depth of points observed in our reference subject geometry.

4.5.1 Rotation. We found that overlay guidance is quite effective at helping users point their camera toward the subject. This leaves a rotational ambiguity that amounts to orienting a subject that is fixed in the center of the current view. To address this, we offer visual guidance in the form of two crosshairs. A white crosshair is fixed in the center of the screen for reference, and a blue crosshair moves according to the current relative orientation of our subject. From the user's perspective, the correct orientation is achieved when the blue crosshair matches up with the white one. Geometrically, we accomplish this behavior by positioning the blue crosshair based on a projection of a vector from our anchor point to the camera location in our reference viewpoint onto the current image plane (see Figure 6). The rotation of the blue crosshair is then set to the rotation of the reference viewpoint's xy axes relative to the current view. We experimented with several different designs for guiding rotation and found this one to be particularly effective for its simplicity and near-invariance to camera motion that is parallel to the current camera's image plane.

4.5.2 Translation Guidance. Our visual guidance for translation focuses on motion parallel to the current camera's image plane,

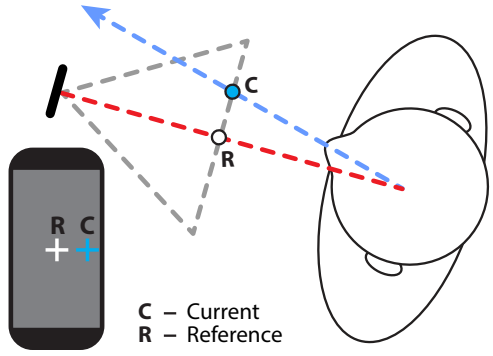

Figure 6: Rotation Guidance. To configure rotation, the user's goal is to align the blue crosshair with the white one which acts as a visual indicator of the user's current relative orientation. Geometrically, the position of the blue crosshair on the screen is determined based on a projection depicted here as the intersection of the blue arrow with the camera frustum.

as translation along the z-axis of our camera (i.e., changing the distance to a subject) is already visualized in our depth heatmap. We use two rings to guide translation along the xy plane: one stationary ring at the center of the view used as a target, and a second ring that is translated proportional to the current xy offset relative to the reference camera (see Figure 4). Aligning the two rings indicates that the camera has reached the target xy translation.

4.6 Automatic Recapture

Our ability to continuously evaluate viewpoint and configuration lets us automatically trigger capture when these criteria are within a threshold of our target reference sample. This addresses an issue discovered in our formative study, where users found that the need to manually press the camera shutter tended to ruin alignment right at the moment of capture. For subjects that are more physically difficult to capture, like the bottom of the foot, this effect can be particularly significant. We trigger automatic recapture based on viewpoint thresholds for translation and rotation, and a configuration threshold for the net depth disparity between the current masked depth and the reference masked depth. Additional details about these thresholds can be found in our supplemental material.

4.7 Computational Lighting

The ability to capture data in different lighting environments is crucial to supporting regular and convenient capture. Our key insight is that the only illumination we can reliably control during capture is the light that comes from our capturing device. Our strategy to control illumination works by isolating this lighting that we can control and computationally removing all the environmental lighting that we cannot control from each observation. To do this, we capture a rapid burst of two images for each sample: one with the camera flash turned off, and another with the camera flash turned on. Our no-flash image contains only the uncontrolled environment lighting in the scene, while the flash image controls the same environment lighting in addition to the light coming from our

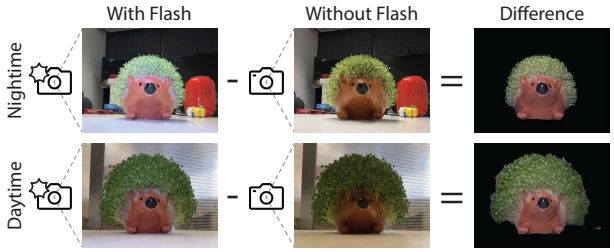

Figure 7: Computational Illumination: Our technique controls for lighting by taking the difference between two images taken with and without the flash in rapid succession. The final computed results have consistent lighting, despite being captured in drastically different lighting conditions (night vs day), which clearly shifts the focus to the change in growth as shown in the rightmost column.

device. Taking the difference between these two images then gives us an image showing only the controlled light coming from our device (Figure 7). We found this approach extremely effective when applied to calibrated RAW photos, which we capture by default in our tool. Figure 13 shows several examples of samples captured under different lighting conditions that are rendered with consistent lighting using our approach.

5 IMPLEMENTATION

Our application runs on iOS devices, and is written in Swift and Metal. Our computational illumination pipeline runs offline in Python, and the resulting PTL can be re-loaded into the app or viewed in a browser with a separate WebGL-based viewer. Our mobile app also offers an interactive preview of each PTL without any offline computation.

5.1 Geometry-Based Mobile 3D Tracking

Our 3D tracker is inspired by the ICP-based tracking used in KinectFusion [18, 30], which calculates the current camera pose by performing multi-scale projective ICP to align the current depth image with a reference model of the scene. Current open-source implementations of KinectFusion are not suitable for mobile devices for two reasons: first, they are written in CUDA to run on machines with dedicated GPUs, and second, they extract reference geometry from an expensive volumetric SDF that needs to be updated every frame. To run on mobile devices, our tracker uses the depth map from our reference sample in place of one extracted from a SDF, which saves significant memory and removes the cost of updating and extracting geometry at each frame. Our remaining challenge was to refactor the fast projective ICP algorithm to work efficiently on mobile hardware. For this, we perform fast projective data association on the GPU. Using shaders written in Metal, we construct a three-level, coarse-to-fine representation of each depth maps and their corresponding normals. For each iteration, projective data association for point correspondences is performed on the GPU, and then an optimal transformation based on the point-to-plane metric [8, 27] is performed on the CPU, leveraging Apple's Accelerate framework for high-performance vector and matrix computations.

The algorithmic differences between our tracker and ICP-based tracking in previous work are made solely to enable efficient execution on mobile devices, which enables the exploration of exciting new directions in guided capture.

6 USER STUDY

We evaluated our capture tool in an IRB-approved (IRB0147654) user study. The main goals of our user study were to evaluate the effectiveness of our capture tool relative to an existing baseline (simple overlay-based guidance) for capturing PTL of different parts of the body, and to gather feedback from users about different types of guidance. We conducted a within-subjects experiment with 14 participants (11 males, 3 females, ages 20-30) recruited through message boards. None of our participants had experience capturing time-lapses of their body, but four had prior experience creating other types of time-lapses, including landscape time-lapse and stop-motion video. Participants were compensated monetarily for their efforts. As this study focuses on new users of our tool, we also conducted two longer-term case studies, described in Section 7.

6.1 Task Design

We tasked each user with capturing PTL of three different body parts—their face, hand (fist), and foot—using two different capture modes. The first mode, *MeCapture*, includes all the features of our capture tool. The second mode, *OverlayOnly*, includes only the RGB overlay of the reference sample for guidance, similar to the Overlay Mode used by Yan et al. [42]. The study had four parts:

(1) **Tutorial:** Participants were shown how to use each capture mode and given a chance to practice re-capturing data with each mode.
(2) **Reference Capture:** The experiment administrator captured reference samples of participants' face and hand, and the participant captured a reference sample of their own foot.
(3) **Recapture:** Participants rotated through different locations, capturing each of the three body parts at each station before moving to the next. We randomized which of the two capture modes each participant used at even stations, and which they used at odd stations.
(4) **Post-study:** Participants reviewed their captured PTL data and answered a series of questions. This included comparing data from each capture mode without knowing which data came from which mode.

6.2 Quantitative Evaluation

Depth Variation: One way to measure the success of configuration guidance is by analyzing geometric variation across captured samples. When a subject's underlying geometry changes between samples, this creates a lower bound for such variation, but ideal configuration guidance should still be the best way to achieve this lower bound. To measure *depth variation* we first compute the standard deviation of depth values corresponding to each pixel in our subject mask. We then average these per-pixel values to get a global value. Lower values indicate more consistent configurations across captured samples. Figure 8 compares depth variation achieved with each of our capture modes. We see that our full tool, MeCapture,

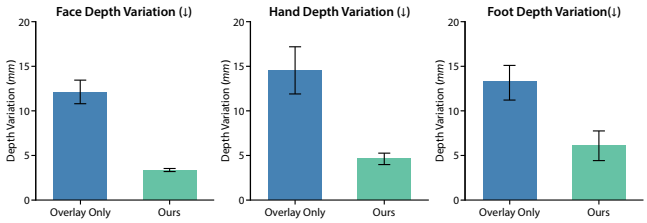

Figure 8: Depth Variation: Here, we plot the average and standard error of our depth variation metric for each capture target. Lower values indicate users were more accurately able to return to consistent configurations. Users consistently performed better using our method over the baseline across all targets.

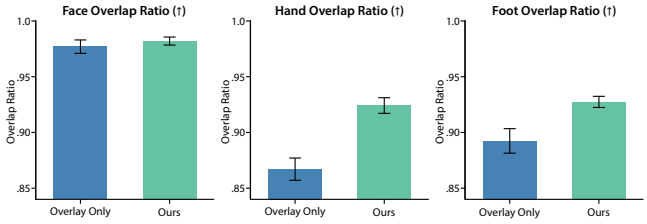

Figure 9: Overlap Ratio: A higher overlap ratio (OR) indicates better consistency and alignment between the depth captures. Plot of face, hand, and foot with mean overlap ratio and standard error. The difference in the face OR is not statistically significant, suggesting that participants are proficient in re-capturing selfies, while for other body parts like the hand and foot, MeCapture statistically improves the overlap ratio metric.

achieves significantly better (lower) depth variation than Overlay-Only ($p < 0.001$ for each of the three body parts). Figure 10 shows some examples visualizing this metric on data from our user study.

Overlap Ratio: While depth variation is effective at measuring accuracy along the optical axis of captured data, this metric can fail to capture misalignment in orthogonal directions when the subject is locally planar. To address this, we also evaluate *overlap ratio*, which measures how well-aligned samples are along the image plane of captured viewpoints. To calculate this, we first compute a cleaned subject mask for each sample by taking the mask used during capture and further refining it with a manually tuned depth range to remove any background pixels. The overlap ratio ϕ_i for recaptured sample i is then:

$$\phi_i = \frac{N_r - X_i}{N_r} \tag{1}$$

where N_r is the number of subject pixels in the reference sample, and X_i is the number of subject pixels in a recaptured sample that do not overlap with subject pixels in the reference sample. We average ϕ_i over all of the recaptured samples in a PTL to calculate an aggregated overlap ratio. Higher ratios indicate more accurate capture. MeCapture led to improved overlap ratios for all three subject types. However, the improvement for hands and feet was significantly higher ($p < 0.001$) than for faces ($p = 0.120$). We

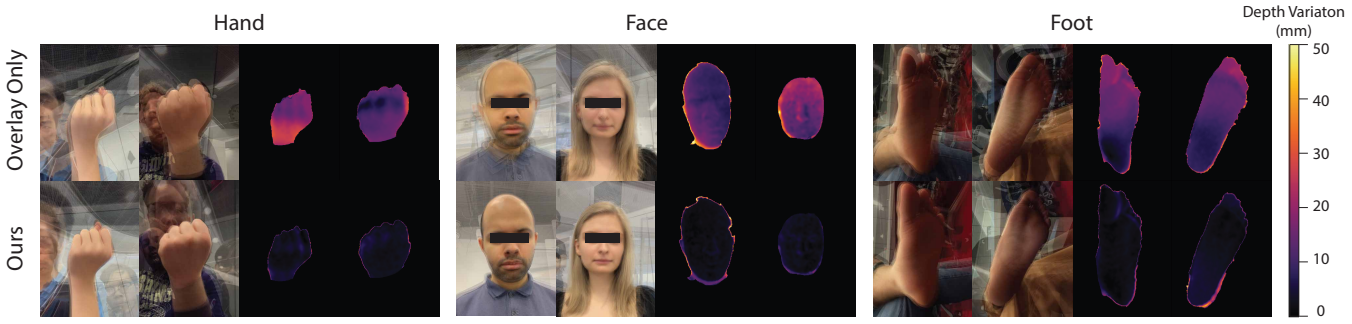

Figure 10: Depth Variation Comparisons from Data Captured in Our User Study: for each body part, we show a grid of results visualizing the consistency of data captured in our user study. The RGB composite images are shown on the left of each grid (hand, face, foot). Each composite is a blend of images captured in a single PTL. On the right of each grid, the heatmaps visualize per-pixel depth variation for each of the corresponding PTLs. In each grid, the top row of results comes from data captured with the baseline OverlayOnly guidance, and the bottom row of results comes from data captured with our full MeCapture guidance. Darker regions of heatmaps correspond to lower depth variation. The composite blend visualizations show that PTL captured with MeCapture is much sharper, and the heatmaps show significantly lower depth variation.

speculate two possible reasons for this: first, the face is probably the most commonly captured part of the human body, so it is a familiar task; and second, when capturing the face, the user's eyes are coincident with the subject of capture, which likely makes alignment easier. These possible explanations aside, our observation here is consistent with our observation in Section 3 that far more time-lapse of faces exist than for other body parts.

Capture Time: As has also been observed in previous work, there is often a trade-off between the speed and quality of captured data. To better understand this, we also recorded the amount of time users took to capture each sample. Time was recorded from when guidance began to when a capture action was triggered. On average, users did take longer to capture data with MeCapture than with OverlayOnly (face: +24.02s, hand: +35.83s, foot: +89.36s). This increase in capture time was accompanied by significant improvements to both objective and subjective accuracy measures, illustrating the trade-off between speed and quality. However, we note that the increase in capture time appears to diminish with extended use. For example, after extended use, our longer-term case study participants were able to recapture data significantly faster than the participants of our user study.

6.3 Qualitative Evaluation

User Confidence: Participants rated their confidence in achieving accurate alignment with each of the two capture modes on a 5-point Likert scale. Participants reported significantly higher confidence with MeCapture ($M = 4.214, SD = 0.699$) compared to OverlayOnly ($M = 2.5, SD = 1.019$). A paired samples t-test revealed a statistically significant difference between these two conditions ($p < 0.001$).

Feedback on Individual Features Participants were also asked to rate how useful they found individual features on a 5-point Likert scale. Figure 12 summarizes the responses. All of our added features were rated on the positive side of the scale by a majority of participants, but with variation in the specific scores for each feature. We note that these values reflect the perceived usefulness of each feature, which does not necessarily reflect the impact that each feature has on capture quality. However, it does tell us something about how users experienced each feature.

Subjective Comparison of Captured Data: After completing all capture tasks, participants were asked to compare blended composites of data captured with each capture mode. For each subject, we generated a blended composite for each of the two capture modes and showed these composites to the user side-by-side. The composites were unlabeled, and the ordering of capture modes was randomized so that users could not tell which composite came from which mode. Users were then asked to select the more consistent composite (i.e., the one with better-aligned images and therefore fewer ghosting artifacts). Participant responses (plotted in Figure 11) show a significant overall preference for data captured with MeCapture for hand ($p = 0.006$) and foot ($p = 0.006$) captures. For face captures, participants still preferred MeCapture, but only 71% of the time ($p = 0.090$). This difference echos our findings when analyzing overlap ratio.

6.4 Open-Ended User Feedback

Participants were also prompted to provide open-ended feedback about their experience, from which we highlight several themes. Full responses can be found in our supplemental material.

Coarse and Fine Guidance: All users found some form of visual guidance useful, with some commenting that the basic 2D overlay was particularly helpful for coarse alignments, *"The overlay made initial alignment easy..."*. For finer adjustments, the color heatmap was particularly effective: "*The color heatmap was useful when making fine-tuned adjustments, especially in the fist example, where it helped convey how I needed to move my hand and individual fingers to best line up with the reference scan.*"

Disambiguating Rotation and Translation: As noted in previous work and our formative study, disambiguating rotation, translation, and configuration, is one of the hardest parts of fine-scale

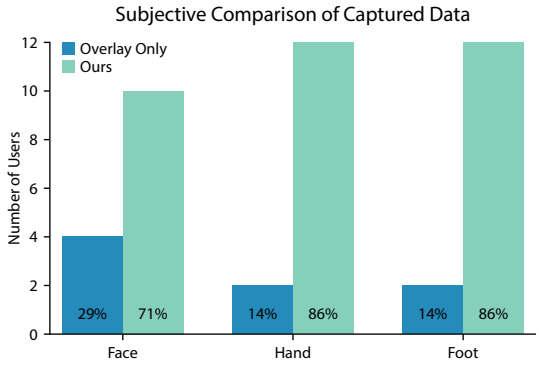

Figure 11: Subjective Comparison of Captured Data: In the post-task assessment, participants reviewed pairs of captured images representing their face, hand, and foot, with the left and right images randomized without knowing which came from which mode. They were then asked to select the image that demonstrated better alignment consistency.

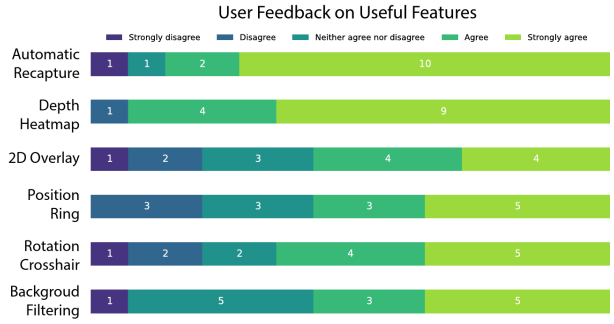

Figure 12: User Feedback on Useful Features: In the post-task assessment, participants were asked, *"For each of the following features, how much do you agree with the statement 'I found the [feature] to be useful?'"*. Their responses, ranging from "Strongly disagree" to "Strongly agree", are presented in this chart. Features are sorted by average approval, providing a clear snapshot of user preferences.

alignment. While our added guidance helps a great deal with this, it remains a significant challenge. As one participant noted, *"It's hard to manage all three dimensions at once. Adjusting one throws off the others."* Part of this challenge may also relate to physical control of the camera and subject, which cannot be addressed with guidance alone. Understanding the limits of how much visual guidance can help with alignment could be an interesting question for future work.

Tension between Speed and Accuracy: While our quantitative analysis shows a significant improvement in the quality of data captured with our tool, it also shows that capturing this data tended to take longer. This makes sense, given that much of our guidance amounts to highlighting alignment errors that might otherwise go unnoticed. Most users recognized the value in this; as one user said *"It takes practice, but the 3D guidance allows much more precise alignment."* However, some saw this guidance as a burden, interpreting the guidance as placing higher requirements on capture. One participant wrote, *"I got tired relatively quickly trying to satisfy all the alignment constraints."* One possible strategy to avoid frustrating users who favor speed over accuracy could be to adjust the guidance visualizations and automatic recapture threshold to be less sensitive to alignment errors.

Automatic Recapture: Users overwhelmingly found automatic recapture to be a useful feature. Much of this opinion seemed to come from negative experiences with manually triggering capture in OverlayOnly mode, *"It was nice not having to press the button and risk misaligning the camera."* However, some users found it difficult to trigger automatic recapture at times, *"It was difficult to know when automatic capture would trigger."* Our study used the same accuracy thresholds to trigger automatic recapture for all participants, but we could change this to something more adaptive. For example, we could gradually relax the threshold if we detect that a user is taking a particularly long time to capture.

7 CASE STUDIES

The purpose of PTL is to visualize long-term changes of a subject, which is difficult to assess in a typical short user study. To better understand long-term use, we also present two longer-term case studies involving users who captured part of their body regularly over an extended period. In both cases, the users captured data on their personal phones from a variety of locations (at home, work, etc). Samples from these case studies can be seen in Figure 13, and interactive visualizations of each captured PTL can be found on our project website.

7.1 Case Study 1: Facial Hair

User 1 used our app to capture the growth of their facial hair. They captured a reference sample shortly after shaving their face, then recaptured additional samples one or more times a day as their facial hair grew back. In total, they captured 60 samples over a period of 40 days. This user commented that as they became more familiar with the routine of recapturing new samples, they found it easier to do quickly and in a greater variety of settings, *"It definitely got easier over time. I could capture images while walking, sitting in the garage, or outdoors. I can also capture in the dark using the heat map."*. They also reported that the ability to preview a PTL visualization in the app helped motivate regular use, as it let them see the progress of capture over time.

7.2 Case Study 2: Wound Healing Tracking

User 2 began using our app after an accident that led to a burn on the back of their hand. They recaptured new samples of the burn one or more times a day as it healed, totaling 42 samples over a period of 35 days. The capture PTL shows the wound close and scab as it heals. User 2 commented positively on several aspects of the app, *"The rotation guidance was useful in helping me rotate my fist correctly. I captured images at school, home, the office, and other places. This app helped me maintain consistent poses. It's also satisfying to*

Figure 13: Long-term results of Personal Time-Lapse using our system: Case Study 1: Hair Growth Monitoring over 40 days, showcasing the gradual growth of facial hair. Case Study 2: Wound Healing Tracking over 35 days, illustrating the progression of a burn wound's healing process.

have the time-lapse visualizer right in the app, allowing me to enjoy how many images I have taken and observe the alignment."

8 DISCUSSION

Our work offers a powerful, general tool for documenting and visualizing long-term changes of the body. Our longitudinal case studies, in particular, demonstrate the ability to create unique and compelling visualizations using a convenient capture process that is simple to integrate in daily routines.

8.1 Limitations

Our work makes capturing PTL much more practical than it was with existing tools, but it does not always make capturing PTL easy.

8.1.1 Degrees of Freedom. Capture remains especially difficult for subjects with many degrees of freedom in their configuration. For example, hands can become very difficult to recapture if the fingers take on a very non-standard configuration. In such cases, every finger joint becomes an additional degree of freedom that must be aligned, making recapture a much higher-dimensional problem for users to navigate. This limitation is partially a consequence of our tool being general-purpose. One possible solution would be to build tools that focus on specific parts of the body with known degrees of freedom. This would make it easier to design visual guidance that helps disambiguate individual degrees for the user.

8.1.2 Precise Control of Subject & Camera. Even with ideal visual guidance, the ease and accuracy of recapture can be limited by the user's ability to precisely control the camera and subject. For example, small motions in the camera and subject can be caused by handshake, breathing, or even heartbeats. We can think of two possible ways to address this in future work. One would be to use some physical device to help constrain part of the subject or camera. The second would be to use some form of lucky imaging, where a stream of redundant observations are captured for each sample, and only the lowest-error data from this stream is used.

8.2 Future Work

Our current pipeline focuses on repeatedly capturing one viewpoint of a subject, but this could be extended to scan a more complete surface representation. The main challenge here is how to control the configuration and lighting of a subject as it is being scanned.

Our current tool is relatively general-purpose in that it does not target any one specific body part or diagnostic task. However, one could develop specialized variants of our tool for more specific use cases, and in doing so improve performance on those use cases.

8.3 Potential Downstream Applications

8.3.1 Remote Healthcare: Our tool could enable new ways for patients and physicians to interact in remote healthcare settings. In particular, the ability to specify reference samples offers physicians and other experts a powerful way to collect information in outpatient settings.

8.3.2 Field Sciences: While we primarily focus on the human body in this work, our tool can also be used for other types of subjects. This could be especially useful in scientific fieldwork, where there is often interest in observing long-term changes in specimens (e.g., plants, animals, or structures).

9 CONCLUSION

Our work fills an important need in remote healthcare that gained significant attention during the pandemic. We derive criteria for capturing and visualizing long-term changes in the body and identify the weaknesses in existing mobile tracking APIs that limit the ability to provide guidance for these criteria. To address these weaknesses, we developed a custom 3D mobile tracker and used it to design and build a novel guidance interface for capturing personal time-lapse. We combine this with a computational illumination

method to control the lighting of captured subjects as well. We validated our work with a user study and two long-term case studies. Our work is the first to tackle the challenging and important problem of personal time-lapse and has a high potential for real-world impact.

ACKNOWLEDGMENTS

This work was partially supported by a National Science Foundation Faculty Early Career Development Grant under award #2340448. It was also partially supported by a generous gift from Meta. We also thank our study participants and testers, especially Xinrui Liu, for their help and feedback in developing our app.

REFERENCES

[1] 2020. *Photography in Clinical Medicine.* Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-24544-3

[2] Andrew Adams, Natasha Gelfand, and Kari Pulli. 2008. Viewfinder alignment. In *Computer Graphics Forum*, Vol. 27. Wiley Online Library, 597–606. https://doi.org/10.1111/j.1467-8659.2008.01157.x

[3] AppleDev. 2024. Saving and Loading World Data. https://developer.apple.com/documentation/arkit/arkit_in_ios/data_management/saving_and_loading_world_data

[4] Soonmin Bae, Aseem Agarwala, and Frédo Durand. 2010. Computational Rephotography. *ACM Trans. Graph.* 29, 3, Article 24 (jul 2010), 15 pages. https://doi.org/10.1145/1805964.1805968

[5] Cagatay Barut and Hakan Ertilav. 2011. Guidelines for standard photography in gross and clinical anatomy. *Anatomical Sciences Education* 4, 6 (2011), 348–356. https://doi.org/10.1002/ase.247

[6] Ross Brown, Bernd Ploderer, Leonard Si Da Seng, Peter Lazzarini, and Jaap van Netten. 2017. MyFootCare: a mobile self-tracking tool to promote self-care amongst people with diabetic foot ulcers. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction (OzCHI '17)*. Association for Computing Machinery, New York, NY, USA, 462–466. https://doi.org/10.1145/3152771.3156158

[7] Elizabeth Chao, Chelsea K. Meenan, and Laura K. Ferris. 2017. Smartphone-Based Applications for Skin Monitoring and Melanoma Detection. *Dermatologic Clinics* 35, 4 (Oct. 2017), 551–557. https://doi.org/10.1016/j.det.2017.06.014

[8] Yang Chen and Gérard Medioni. 1992. Object modelling by registration of multiple range images. *Image and vision computing* 10, 3 (1992), 145–155.

[9] Hugo Cornellier. 2017. https://youtu.be/65nfbW-27ps?si=7mp5dSxrlBhW4Kj1

[10] Mayur Davda and Paola Pasquali. 2020. *Standardization in Photographic Documentation.* Springer International Publishing, Cham, 211–230. https://doi.org/10.1007/978-3-030-24544-3_14

[11] Jane L. E, Ohad Fried, Jingwan Lu, Jianming Zhang, Radomír Měch, Jose Echevarria, Pat Hanrahan, and James A. Landay. 2020. Adaptive Photographic Composition Guidance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, 1–13. https://doi.org/10.1145/3313831.3376635

[12] Jane L. E, Kevin Y. Zhai, Jose Echevarria, Ohad Fried, Pat Hanrahan, and James A. Landay. 2021. Dynamic Guidance for Decluttering Photographic Compositions. In *Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. ACM. https://doi.org/10.1145/3472749.3474755

[13] Zemichael Gizaw, Tigist Astale, and Getnet Mitike Kassie. 2022. What improves access to primary healthcare services in rural communities? A systematic review. *BMC Primary Care* 23, 1 (2022), 313.

[14] Dirk Hahnel, Sebastian Thrun, and Wolfram Burgard. 2003. An extension of the ICP algorithm for modeling nonrigid objects with mobile robots. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence* (Acapulco, Mexico) *(IJCAI'03)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 915–920.

[15] Xin He, Xi Zheng, and Huiyuan Ding. 2023. Existing Barriers Faced by and Future Design Recommendations for Direct-to-Consumer Health Care Artificial Intelligence Apps: Scoping Review. *Journal of Medical Internet Research* 25 (Dec. 2023), e50342. https://doi.org/10.2196/50342

[16] IMDB. 2024. https://www.imdb.com/name/nm2404488/

[17] IMI. 2024. https://www.imi.org.uk/resources/professional-resources/national-guidelines/

[18] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. 2011. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 559–568.

[19] Noah Kalina. 2020. https://www.youtube.com/watch?v=wAIZ36GI4p8

[20] Jeannie Khavkin and David A. F. Ellis. 2011. Standardized Photography for Skin Surface. *Facial Plastic Surgery Clinics of North America* 19, 2 (May 2011), 241–246. https://doi.org/10.1016/j.fsc.2011.04.001

[21] Minju Kim and Jungjin Lee. 2019. PicMe: Interactive Visual Guidance for Taking Requested Photo Composition. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300625

[22] Georg Klein and David Murray. 2007. Parallel tracking and mapping for small AR workspaces. In *2007 6th IEEE and ACM international symposium on mixed and augmented reality*. IEEE, 225–234.

[23] Bálint Kolozsvári. 2024. Kolo / Time Lapse. https://www.youtube.com/@KoloTimeLapse

[24] Adam Landman, Srinivas Emani, Narath Carlile, David I. Rosenthal, Simon Semakov, Daniel J. Pallin, and Eric G. Poon. 2015. A Mobile App for Securely Capturing and Transferring Clinical Images to the Electronic Health Record: Description and Preliminary Usability Study. *JMIR mHealth and uHealth* 3 (Jan 2015), e3481. https://doi.org/10.2196/mhealth.3481

[25] Katie J. Lee, Anna Finnane, and H. Peter Soyer. 2018. Recent trends in teledermatology and teledermoscopy. *Dermatology Practical & Conceptual* 8, 3 (Jul 2018), 214–223. https://doi.org/10.5826/dpc.0803a13

[26] Hao Li, Robert W. Sumner, and Mark Pauly. 2008. Global correspondence optimization for non-rigid registration of depth scans. In *Proceedings of the Symposium on Geometry Processing* (Copenhagen, Denmark) *(SGP '08)*. Eurographics Association, Goslar, DEU, 1421–1430.

[27] Kok-Lim Low. 2004. Linear least-squares optimization for point-to-plane icp surface registration. *Chapel Hill, University of North Carolina* 4, 10 (2004), 1–3.

[28] Andrew J. Marek, Emily Y. Chu, Michael E. Ming, Zeeshan A. Khan, and Carrie L. Kovarik. 2018. Impact of a smartphone application on skin self-examination rates in patients who are new to total body photography: A randomized controlled trial. *Journal of the American Academy of Dermatology* 79, 3 (Sept. 2018), 564–567. https://doi.org/10.1016/j.jaad.2018.02.025

[29] Raúl Mur-Artal, J. M. M. Montiel, and Juan D. Tardós. 2015. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics* 31, 5 (2015), 1147–1163. https://doi.org/10.1109/TRO.2015.2463671

[30] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. 2011. Kinectfusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE international symposium on mixed and augmented reality*. Ieee, 127–136.

[31] Bernd Ploderer, Atae Rezaei Aghdam, and Kara Burns. 2022. Patient-Generated Health Photos and Videos Across Health and Well-being Contexts: Scoping Review. *Journal of Medical Internet Research* 24, 4 (Apr 2022), e28867. https://doi.org/10.2196/28867

[32] Bernd Ploderer, Atae Rezaei Aghdam, and Kara Burns. 2022. Patient-Generated Health Photos and Videos Across Health and Well-being Contexts: Scoping Review. *Journal of Medical Internet Research* 24, 4 (April 2022), e28867. https://doi.org/10.2196/28867

[33] Mark P. Pressler, Mikaela L. Kislevitz, Justin J. Davis, and Bardia Amirlak. 2022. Size and Perception of Facial Features with Selfie Photographs, and Their Implication in Rhinoplasty and Facial Plastic Surgery. *Plastic and Reconstructive Surgery* 149, 4 (April 2022), 859. https://doi.org/10.1097/PRS.0000000000008961

[34] K.A. Priebe and H. Selick. 2011. *The advanced art of stop-motion animation.* Course Technology. https://books.google.com/books?id=2J8VkgAACAAJ Citation Key: priebe2011advanced tex.lccn: 2010922093.

[35] Rajeev V. Rikhye, Grace Eunhae Hong, Preeti Singh, Margaret Ann Smith, Aaron Loh, Vijaytha Muralidharan, Doris Wong, Rory Sayres, Michelle Phung, Nicolas Betancourt, Bradley Fong, Rachna Sahasrabudhe, Khoban Nasim, Alec Eschholz, Yossi Matias, Greg S. Corrado, Katherine Chou, Dale R. Webster, Peggy Bui, Yuan Liu, Yun Liu, Justin Ko, and Steven Lin. 2024. Differences Between Patient and Clinician-Taken Images: Implications for Virtual Care of Skin Conditions. *Mayo Clinic Proceedings: Digital Health* 2, 1 (Mar 2024), 107–118. https://doi.org/10.1016/j.mcpdig.2024.01.005

[36] S. Rusinkiewicz and M. Levoy. 2001. Efficient variants of the ICP algorithm. In *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*. 145–152. https://doi.org/10.1109/IM.2001.924423

[37] Hannah Schmitz, Carol L Howe, David G Armstrong, and Vignesh Subbian. 2018. Leveraging mobile health applications for biomedical research and citizen science: a scoping review. *Journal of the American Medical Informatics Association* 25, 12 (Dec. 2018), 1685–1695. https://doi.org/10.1093/jamia/ocy130

[38] S. Shaw. 2004. *Stop Motion: Craft Skills for Model Animation.* Focal Press. https://books.google.com/books?id=kuUAWIvPaAgC

[39] Doron D. Sommer and Martyn Mendelsohn. 2004. Pitfalls of nonstandardized photography in facial plastic surgery patients. *Plastic and Reconstructive Surgery* 114, 1 (Jul 2004), 10–14. https://doi.org/10.1097/01.prs.0000127791.31526.e2

[40] Florine Vegter and J. Joris Hage. 2000. Standardized Facial Photography of Cleft Patients: Just Fit the Grid? *The Cleft Palate Craniofacial Journal* 37, 5 (Sept. 2000), 435–440. https://doi.org/10.1597/1545-1569_2000_037_0435_sfpocp_2.0.co_2

[41] Dan E. Webster, Christine Suver, Megan Doerr, Erin Mounts, Lisa Domenico, Tracy Petrie, Sancy A. Leachman, Andrew D. Trister, and Brian M. Bot. 2017. The Mole Mapper Study, mobile phone skin imaging and melanoma risk data collected using ResearchKit. *Scientific Data* 4, 1 (Feb. 2017), 170005. https://doi.org/10.1038/sdata.2017.5 Publisher: Nature Publishing Group.

[42] Ruyu Yan, Jiatian Sun, Longxiulin Deng, and Abe Davis. 2022. ReCapture: AR-Guided Time-lapse Photography. In *ACM Symposium on User Interface Software and Technology (UIST)*. https://doi.org/10.1145/3526113.3545641

[43] Moi Hoon Yap, Katie E. Chatwin, Choon-Ching Ng, Caroline A. Abbott, Frank L. Bowling, Satyan Rajbhandari, Andrew J. M. Boulton, and Neil D. Reeves. 2018. A New Mobile Application for Standardizing Diabetic Foot Images. *Journal of Diabetes Science and Technology* 12, 1 (Jan 2018), 169–173. https://doi.org/10.1177/1932296817713761

[44] Timothy Zoltie, Sigrid Blome-Eberwein, Sarah Forbes, Mike Theaker, and Walayat Hussain. 2022. Medical photography using mobile devices. *The BMJ* 378 (Aug 2022), e067663. https://doi.org/10.1136/bmj-2021-067663