

Stitching 3D Millimeter Wave Images of a Person in Motion

K. Parker Trofatter, Jonah N. Gollub, and David R. Smith

Abstract—3D millimeter wave imaging systems are indispensable for security screening applications. Available imagers are constrained to imaging only one person at a time and require people to strike and hold a pose during measurement. This is both inconvenient and introduces a security checkpoint bottleneck. Recently, an experimental millimeter wave imager based on computational imaging principles has been demonstrated with a near-video shutter speed. This allows people to move during imaging, promising to significantly boost screening throughput. However, applying automatic threat detection to images of a person in motion is complicated by deformation of the human body and specularly-limited coverage characteristic of millimeter wave imaging. One strategy to address both issues is to combine multiple images into a single composite stitch. Here, we develop a model to stitch images of a person in motion. We systematically demonstrate in experiment and simulation this approach and advocate that it can be applicable to any real-time 3D imaging system.

Index Terms—computer graphics applications, interaction techniques, sampling, registration, image models, motion, sensor fusion

I. INTRODUCTION

REAL-TIME 3D imaging systems measure and stream tremendous amounts of data about their environment. Taking advantage of these sensors at the same rate they produce raw data requires computationally efficient techniques to reconstruct, analyze, and represent high-level scene content. This challenge is compounded by the push for multi-sensor fusion in emerging technologies, such as sensor suites in smartphones and wearable devices [1]; depth cameras [2]; and video, radar, and lidar systems of autonomous vehicles [3].

An important application at the forefront of 3D imaging and data processing technology is millimeter wave security screening [4]. Millimeter waves are a form of electromagnetic radiation particularly suited for screening people because they safely penetrate clothing while strongly reflecting off skin and concealed threat objects [5]. Millimeter wave imagers require large apertures to form images of people with millimeter resolution. Commercially available systems accomplish this with a dense array of stationary antennas [6], or by synthetic aperture radar (SAR) with a mechanically scanned antenna array [7]. However, these approaches have been limited to imaging individual people and requiring that people strike and hold a pose during imaging, presenting a throughput bottleneck. Interest has grown in screen-while-walk systems that allow people to move freely through the imaging volume during measurement [8]. Recently, a screen-while-walk millimeter wave imager was experimentally demonstrated with

an inexpensive and scalable modular architecture based on computational imaging principles [9].

Modular screen-while-walk imaging systems invite a full reevaluation of security screening. Deployment options are expanded by relaxing assumptions about the configuration of the aperture and how people interact with the device. For instance, systems unobtrusively installed at portals and along corridors can opportunistically image passerbys. At a lower cost point, multiple checkpoints become feasible, which could cooperatively build models of people in an area. Increased screening throughput and distributed sensor networks would prevent vulnerable crowds from forming at checkpoints, and could move the secure area boundary outside critical infrastructure such as airport terminals.

Relaxing control over the imaging environment and allowing people to move comes at the price of complicating automatic threat detection (ATD). Of central concern is the deformation of the human body while it is in motion. A standard approach for ATD is training machine learning algorithms for anomaly detection [10]. Conceptually, ATD could be trained on data of people in different poses with and without threat objects, however direct application of this idea is intractable because the space of human poses is enormous. Additionally, active illumination millimeter wave imagers suffer from specularly-limited coverage in which reflective objects smooth on the scale of the probing wavelengths appear mirror-like. This limits the visibility of specular surfaces to highlights that satisfy the law of reflection with the probing antennas. As the human body and many threat objects are smooth on millimeter scales, this can cause significant reduction in scene coverage for any single image, which ATD must contend with. Commercial systems mitigate specularly-limited coverage with careful aperture design and human posing restrictions.

The observation can be made that the motion of a person through a screen-while-walk imager supplies many views of the person with different coverage. Intuitively, these images can be stitched together to produce a single image with better coverage, which may be easier to process with ATD. While the idea is simple, it is not clear how to combine multiple images of a deformable body into a single image in real-time. We address the stitching problem as it applies to the computational imager, and systematically develop a model for stitching people in motion through experiment and simulation. Conceptually the technique can be applied to other 3D imagers.

This document is divided into 6 sections. Section II introduces the experimental millimeter wave computational imaging system and imaging model, and general observations about reconstructed images are made. Section III discusses

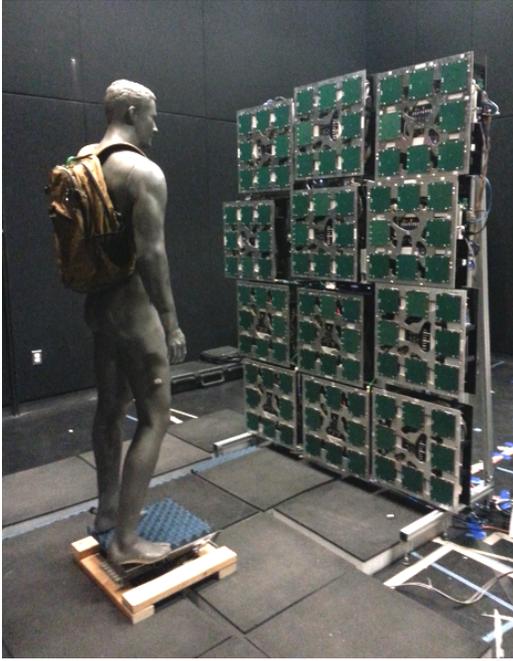


Fig. 1. Experimental Millimeter Wave Imaging System — A modular array of stationary antennas images a scene. The antenna radiation patterns have low spatial correlation as a function of frequency. Backscattered transceiver measurements between pairs of antennas encodes scene information related to a physical forward model. Images are reconstructed by inverting the model.

the fundamental stitching tasks of registration, calibration, and blending, and how they relate to the experimental imager. Section IV develops and experimentally demonstrates procedures to stitch scenes undergoing rigid body motion. Section V extends these ideas to stitch images of a person in motion using a skeleton model. Crude but promising experimental results are obtained with a depth camera, lending credibility to the technique. A more compelling stitching model incorporating deformation is then developed and demonstrated in simulation. Section VI summarizes the stitching techniques covered and identifies key areas of research going forward.

II. MILLIMETER WAVE COMPUTATIONAL IMAGER

We advocate that the stitching procedure covered in this paper has general application to the challenges of using a 3D sensor to monitor a scene. The motivation for this work stems specifically from research on an experimental millimeter wave imaging system that is notable for having a shutter speed fast enough to image people in motion, and the advent of reconstruction algorithms that can operate in real-time. A number of enabling technologies were developed to realize the millimeter wave system, unified by the theme of computational imaging. Understanding how the system functions is important to interpreting images and developing a stitching strategy.

Computational imaging permits rethinking the hardware layer of conventional imaging systems. Traditional imaging relies on the sensor hardware for both measuring information about a scene and forming an image, such as how a digital camera focuses light from a scene onto a pixel detector array which directly records an illuminance. In contrast, computational imaging relaxes the role of the sensor to only measuring

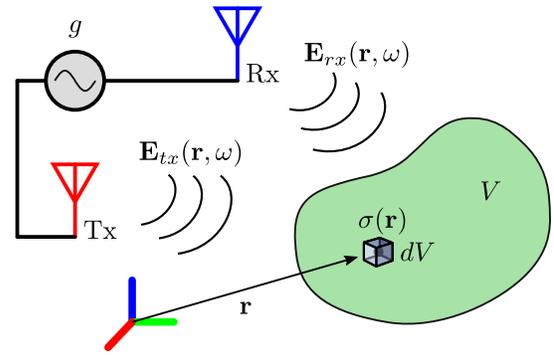


Fig. 2. Imaging Forward Model — Predict transceiver measurement g given a volume V with conductivity distribution $\sigma(\mathbf{r})$ excited by Tx electric field $\mathbf{E}_{tx}(\mathbf{r}, \omega)$ and sensed by Rx electric field $\mathbf{E}_{rx}(\mathbf{r}, \omega)$ at frequency ω

information about a scene, relying on post-processing to reconstruct an image. This dramatically increases the design space of sensors and measurement modalities, and deftly leverages the aggressive growth of computer technology [11]. For instance at optical frequencies, a coded aperture imaging system consisting of a single pixel detector that collects light through a reconfigurable mask can, with enough measurements using different masks, reconstruct images similar to those formed by a camera with a focal plane array [12], [13]. While the benefit and utility of simplifying inexpensive camera hardware in the optical regime may be questionable, for expensive microwave transceiver hardware the savings can be substantial and fundamentally enabling.

For our experimental millimeter wave imager, a coded aperture approach is used to implement a cost-effective and real-time imaging solution [14], [15], [16]. The key is to use antennas with radiation patterns that have low spatial correlation as a function of driving frequency [17]. A frequency swept transceiver can then rapidly measure scene information [18]. The number of antennas and frequency bandwidth is optimized for personnel security screening. The experimental imager uses 24 transmit (Tx) antennas, 72 receive (Rx) antennas, and 101 frequency points over K-band to give more than 175,428 unique measurement combinations to roughly match the number of resolution limited voxels a person occupies [19]. Serially measuring this many combinations is time consuming, so a multiplexed strategy based on binary phase-shift keying is employed to measure in parallel. Real-time shutter rates of 7 Hz have been experimentally demonstrated, and up to 100 Hz are possible [9].

Image reconstruction is cast in terms of an inverse scattering problem. The forward scattering problem is to predict transceiver measurements between Rx-Tx pairs for a given volumetric scene conductivity distribution (Figure 2). In general, fully simulating this problem is computationally prohibitive for real time applications, so simplifying assumptions are made. In particular, the scene is assumed to be composed of non-magnetic, isotropic, and non-dispersive material; and multi-scattering is ignored by invoking the first Born approximation.

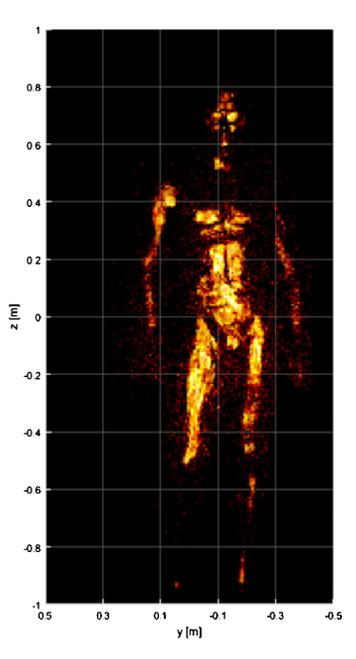


Fig. 3. Reconstructed 3D Image — A typical reconstructed image of a conductive mannequin. Images consist mostly of surfaces embedded within the imaging volume. Note the limited coverage due to specularity.

This results in the linear equation [20]

$$g_{ijk} = \int_V \mathbf{E}_{tx,i}(\mathbf{r}, \omega_k) \cdot \sigma(\mathbf{r}) \mathbf{E}_{rx,j}(\mathbf{r}, \omega_k) dV. \quad (1)$$

where g is the predicted transceiver measurement, V is the imaging volume, σ is the scene conductivity distribution, and \mathbf{E}_{rx} and \mathbf{E}_{tx} are the Rx and Tx antenna electric field radiation patterns respectively. Measurements are indexed by a set of Rx antennas $i \in R$, a set of Tx antennas $j \in T$, and a set of frequencies $k \in \Omega$. By discretizing the integral into voxels, rewriting the imager-dependent quantity $\mathbf{E}_{rx,i} \cdot \mathbf{E}_{tx,j}$ as a matrix \mathbf{H} , and vectorizing scene conductivity σ and measurement \mathbf{g} , the forward model is compactly written as the matrix equation

$$\mathbf{g} = \mathbf{H}\sigma. \quad (2)$$

For our imaging systems, directly relating the discretized scene configuration σ to transceiver measurements \mathbf{g} for people-sized volumes at centimeter resolutions can result in a matrix \mathbf{H} on the order of 10^5 measurements by 10^6 voxels. To reduce the size of this matrix, depth cameras provide prior information about the location of foreground objects within the scene, which routinely eliminates over 90% of the columns of \mathbf{H} and rows of σ . However, explicitly computing a pseudoinverse to even this reduced problem is prohibitively expensive. Instead, either FFT-based techniques [21] or regularized iteration must be employed to realistically invert (2). GPU accelerated matrix implementations using look-up tables and matrix factorization have achieved reconstruction times on the order of 10 s [19]. If the inverse scattering problem is instead cast as a stationary phase asymptotic series, reconstruction times on the order of 100 ms have been demonstrated [22].

The forward model (1) can be used to understand characteristics of reconstructed images (Figure 3). An immediate implication is that images are volumetric. However, built into the Born approximation is the assumption that the incident fields are not appreciably delayed by the scene, making it a backscatter model suitable only for highly conductive materials that strongly reflect at their surface and prevent radiation from penetrating their interiors. Consequently, images mostly consist of surfaces, and lack volumetric information about the interior of objects. For conductive media, like human skin and metal, this is a good assumption. For dielectrics, like plastic and glass, this assumption fails because light will propagate through these media at different speeds depending on the index of refraction, resulting in artifacts where surfaces occluded by dielectrics appear farther away in images than in reality. Another issue is that specular reflection limits scene coverage. Surfaces smooth on the order of the probing wavelengths are mirror-like and visible only when they satisfy the law of reflection with respect to the probing Rx-Tx pair. As people and many threats are smooth at our operational wavelengths (K-band 1.7-1.1 cm), specular reflection can significantly limit scene coverage.

Image characteristics influence the choice of antenna layout, which crucially determines scene coverage. Stationary apertures with a slow shutter speed maximize coverage by imaging people in a standardized pose from an optimal perspective. Mechanically scanned apertures go further by measuring the imaging volume from a diversity of perspectives, which increases the chance a specular surface is seen. However, neither of these options are available when imaging a person in motion with a stationary aperture. To maximize coverage and mitigate specularity, the frequency-diverse imager must exploit its layout and fast shutter speed to collect multiple images of people as they move through the imaging volume.

III. STITCHING

Image stitching seeks to combine multiple images of the same scene into a single composite image called a *stitch*. In general, stitching can be divided into three tasks: registration, calibration, and blending [23]. First, registration aligns images by modeling the geometric transformation relating images, establishing correspondences between images, and then estimating the transformation parameters between those correspondences. Next, calibration equalizes images to compensate for variations in the imaging process, making overlapping registered images locally similar. Finally, registered and calibrated images are blended to compute the resulting stitch by compositing. In this section we discuss how these tasks apply to millimeter wave images of people in motion, where specularity-limited coverage and deformation complicate the stitching process. The section concludes with a basic stitching algorithm.

A. Registration

Registration is perhaps the most involved stitching task, and the focus of this work. Humans are visually gifted by nature, and easily align overlapping images of a scene. However,

distilling what we do into an algorithm is surprisingly difficult. We briefly review the literature for inspiration.

A popular stitching problem is to combine multiple photographs into a panoramic stitch [24]. Due to user measurement patterns and camera image distortion, the geometric transformation between photos is usually parameterized by cylindrical or spherical projections [25]. Photo registration generally proceeds by correlation [26], neural network inference [27], or local feature analysis to detect correspondences and estimate the geometric transformation parameters between photos. Local feature analysis in particular efficiently finds, describes, and matches localized image features deemed "interesting". Feature algorithms such as SIFT [28] and SURF [29] coupled with bundle adjustment [30] or stochastic search algorithms like RANSAC [31] are effective at solving the photo registration problem. However, while local feature analysis seamlessly extends to 3D images [32], it is unclear if millimeter wave images characterized by specularly-limited coverage of mirror-like surfaces have sufficiently distinguishable features. Moreover, photo registration gives little insight into modeling a deformable scene.

Medical imaging is deeply concerned with the registration of deformable scenes. For instance, radiotherapy planning may use computational anatomy to accurately register anatomical structures against templates to safely predict and administer dosage [33]. However, sophisticated medical image analysis tends to be computationally prohibitive for real-time applications. A balance must be struck between accuracy and computational complexity.

To develop a registration model suitable for real-time millimeter wave imaging of people in motion, we start by investigating the subproblem of registering images of rigid body scenes that don't deform. The defining property of a rigid body is that the (Euclidean) distance between points on the body are constant in time. Therefore, the geometric relationship between a rigid body at two different times is a distance-preserving isometry transformation. Isometries of \mathbb{R}^3 are members of the Euclidean group $E(3)$, and include rotations, translations, reflections, and their compositions. Since a rigid body can't be continuously transformed into a reflection without deformation, we restrict the isometries under consideration to the Special Euclidean group $SE(3)$, which excludes reflections. These transformations are called *rigid transformations*. For our millimeter wave imaging system, the measurement and image formation process introduce no spatial distortion in the resulting volumetric images. Thus, images of rigid body scenes are geometrically related by rigid transformations.

A rigid transformation $A_{ba} \in SE(3) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ (Figure 4) can be interpreted as a change of basis relating how a geometric point represented by $\mathbf{r}_a \in \mathbb{R}^3$ with respect to basis a is represented by $\mathbf{r}_b \in \mathbb{R}^3$ with respect to basis b . Any rigid transformation can be uniquely represented by some proper rotation matrix $\mathbf{R}_{ba} \in \{\mathbf{M} \in \mathbb{R}^{3 \times 3} | \det(\mathbf{M}) = 1\} = SO(3)$ (Special Orthogonal group $SO(3)$ excludes reflections) and translation vector $\mathbf{t}_{ba} \in \mathbb{R}^3$

$$\mathbf{r}_b = A_{ba}(\mathbf{r}_a) \equiv \mathbf{R}_{ba}\mathbf{r}_a + \mathbf{t}_{ba}. \quad (3)$$

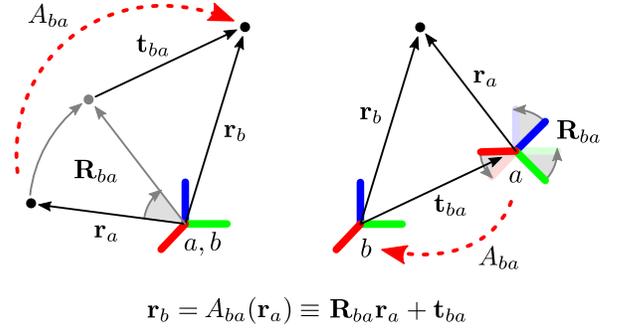


Fig. 4. Rigid Transformation — A rigid transformation $A_{ba} \in SE(3)$ can be interpreted as transforming points with respect to a fixed basis (left), or equivalently a change of basis describing fixed points (right). As a change of basis, A_{ba} maps point representation $\mathbf{r}_a \in \mathbb{R}^3$ with respect to basis a , to $\mathbf{r}_b \in \mathbb{R}^3$ with respect to basis b . A_{ba} can be represented by proper rotation matrix $\mathbf{R}_{ba} \in SO(3)$ and translation vector $\mathbf{t}_{ba} \in \mathbb{R}^3$.

The subscripts on A_{ba} indicate the transformation is from basis a to basis b . The inverse of (3) is easily verified to be

$$\mathbf{r}_a = A_{ba}^{-1}(\mathbf{r}_b) \equiv \mathbf{R}_{ba}^{-1}(\mathbf{r}_b - \mathbf{t}_{ba}), \quad (4)$$

and is equivalent to A_{ab} from basis b to basis a

$$A_{ab} = A_{ba}^{-1} \quad (5)$$

where $\mathbf{R}_{ab} = \mathbf{R}_{ba}^{-1} \in SO(3)$ and $\mathbf{t}_{ab} = -\mathbf{R}_{ba}^{-1}\mathbf{t}_{ba}$.

Rigid transformations are closed under functional composition. Given A_{ba} from basis a to basis b , and A_{cb} from basis b to basis c , we can compute A_{ca} from basis a to basis c by repeated application of definition (3)

$$\begin{aligned} \mathbf{r}_c &= A_{cb}(A_{ba}(\mathbf{r}_a)) \\ &= \mathbf{R}_{cb}(\mathbf{R}_{ba}\mathbf{r}_a + \mathbf{t}_{ba}) + \mathbf{t}_{cb} \\ &= (\mathbf{R}_{cb}\mathbf{R}_{ba})\mathbf{r}_a + (\mathbf{R}_{cb}\mathbf{t}_{ba} + \mathbf{t}_{cb}) \\ &= \mathbf{R}_{ca}\mathbf{r}_a + \mathbf{t}_{ca} \\ &= A_{ca}(\mathbf{r}_a), \end{aligned}$$

where $\mathbf{R}_{ca} = \mathbf{R}_{cb}\mathbf{R}_{ba} \in SO(3)$ and $\mathbf{t}_{ca} = \mathbf{R}_{cb}\mathbf{t}_{ba} + \mathbf{t}_{cb}$. This is compactly written using functional composition notation as

$$A_{ca} = A_{cb} \circ A_{ba}. \quad (6)$$

Notice how the basis b subscripts "cancel" in (6); in a sense, basis b helps us compute A_{ca} , but is not fundamental to the description of A_{ca} .

With the language of rigid transformations we can develop a registration model for rigid scenes (Figure 5). The geometric transformation parameters, or *pose*, of a rigid scene is fully described by a rigid transformation with respect to some basis. We seek to take images of a rigid scene in different poses relative to global basis g and align them relative to stitch basis s . Suppose basis b is attached to the rigid scene. We can experimentally measure pose A_{gb} for each image, and define a constant *rest pose* A_{bs} to locate the scene relative to basis s . Then pose

$$A_{gs} = A_{gb} \circ A_{bs} \quad (7)$$

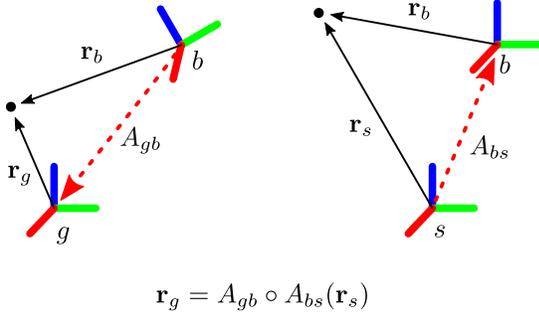


Fig. 5. Rigid Registration Model — A rigid scene is imaged relative to global basis g (left). Images are aligned relative to stitch basis s (right). Suppose basis b is attached to the scene. Pose A_{gb} is experimentally measured each image. Rest pose A_{bs} is a constant defined to locate the stitch relative to basis s . Pose $A_{gs} = A_{gb} \circ A_{bs}$ then relates stitch points r_s to global points r_g .

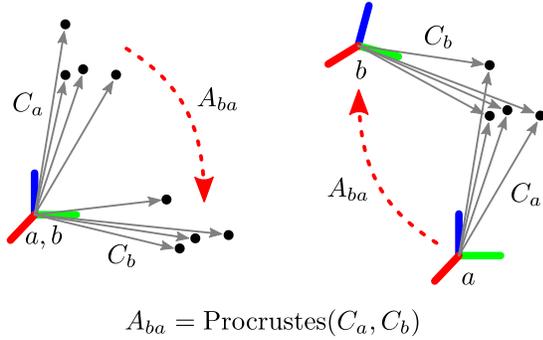


Fig. 6. Orthogonal Procrustes Analysis — Computes the optimal rigid transformation A_{ba} that aligns constellation C_a onto constellation C_b (left). Equivalently, A_{ba} is interpreted as a change of basis from a to b (right).

relates stitch points r_s to global points r_g . Thus for rigid scene registration the main task is to measure pose A_{gb} .

In a controlled imaging environment, one way to experimentally estimate the pose of an object with respect to a sensor is to measure a known set of points $C \subset \mathbb{R}^3$, called a *constellation*, and apply Orthogonal Procrustes Analysis (OPA) (Figure 6). OPA computes the optimal rigid transformation A_{ba} that transforms constellation C_a onto a similar constellation C_b (see supplementary materials for derivation)

$$A_{ba} = \text{Procrustes}(C_a, C_b). \quad (8)$$

Equivalently, pose A_{ba} can be interpreted as a change of basis from basis a to basis b . To use OPA in practice, it is assumed constellation C_a is known and recorded relative to basis a in a virtual template file before measurement. The constellation is assumed to consist of three or more points distributed asymmetrically to uniquely constrain a pose. A physical version of the constellation is then affixed to a rigid scene and measured as constellation C_b relative to some sensor system basis b . Application of equation (8) results in pose A_{ba} that encodes the pose from the rigid scene to the sensor system. This is used in several contexts throughout the work.

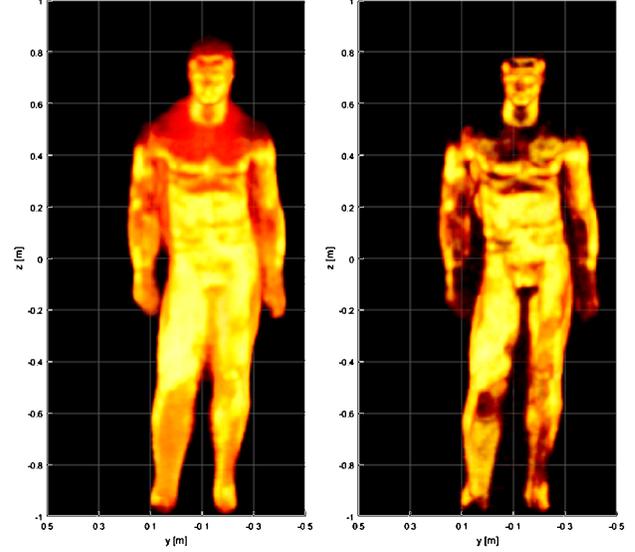


Fig. 7. Blending — Stitched data from a rotation experiment is blended in two ways. Magnitude summation blending (left) is sensitive to noise and exposure time, while maximum magnitude blending (right) is not.

Section IV experimentally applies the rigid registration model and OPA to rigid scenes undergoing translational and rotational motion. Section V extends the rigid registration model to deformable scenes by incorporating a pose skeleton, which is used to approximate the deformation of a scene using a small number of rigid transformations.

B. Calibration

Calibration makes overlapping registered images locally similar by correcting for variations in the imaging process. For photographic stitching, the primary concern is limited dynamic range, in which the dynamic range of a panorama exceeds the dynamic range of the camera, such as when stitching photos of bright sky and dark ground. Tone mapping is typically used to stitch these types of high dynamic range panoramas.

For millimeter wave images, the interpretation of signal magnitude between different images is more difficult to ascertain, which affects blending. The magnitude of the reconstructed images depends in a complicated way on scene configuration, dynamic range, signal-to-noise ratio (SNR), prior information, and regularized reconstruction algorithms. The magnitude of each image is effectively determined up to a different unknown scaling factor.

Experimental experience indicates that similar scenes have similar normalized magnitudes. Satisfactory preliminary stitching results have been obtained without extensive calibration. However, calibration must be done to properly blend dissimilar images for high quality ATD input. This remains an open avenue of research.

C. Blending

The last stitching task is to blend registered and calibrated images by compositing to produce the final stitch. For every

stitch voxel, the associated stitch point \mathbf{r}_s is transformed to global point \mathbf{r}_g , the source images are sampled using interpolation, the sampled data is combined by some blending operation, and the result is accumulated as stitch voxel data.

There is no one correct blending operation, and a variety of choices exist to varying effect (Figure 7). A plausible rule is to add the sampled image magnitudes together. This works but produces a stitch that is sensitive to exposure time and low-level noise. For data sets of unknown size, such as streaming data, a better rule is to assign the maximum magnitude of the sampled images, which mitigates the issues with addition. For the remainder of this paper all results use the maximum value blending strategy.

D. Algorithm

To stitch an image set, we apply the following algorithm. Given input image set Σ , image reconstruction points $\mathbf{r}_g \in \mathbb{R}^3$ with respect to the global basis g , stitch points $\mathbf{r}_s \in \mathbb{R}^3$ with respect to stitch basis s , and rest pose A_{bs} , we wish to output a stitch. First initialize the stitch to zero. For each image, solve the registration problem to estimate pose A_{gb} . Then transform \mathbf{r}_s with equation (7) to obtain query points \mathbf{r}'_g . Next, sample the image at the query points using interpolation to obtain registered image data. Finally, blend registered data with the current stitch to update the stitch. This algorithm is suited for image set Σ of unknown size.

Algorithm 1: Stitching

Input: $\Sigma, \mathbf{r}_g \in \mathbb{R}^3, \mathbf{r}_s \in \mathbb{R}^3, A_{bs} \in \text{SE}(3)$

Output: stitch

stitch = 0

for image $\in \Sigma$ **do**

$A_{gb} = \text{register}(\text{image})$

$\mathbf{r}'_g = A_{gb} \circ A_{bs}(\mathbf{r}_s)$

 data = interpolate(image, $\mathbf{r}_g, \mathbf{r}'_g$)

 stitch = blend(stitch, data)

end

IV. RIGID BODY STITCHING

Due to the complexity of stitching millimeter wave images of a person in motion, we first apply the rigid registration model (7) developed in Section III to perform rigid body stitching. Conceptually, a rigid scene and rigid imager move relative to each other along different trajectories while an image set is measured. In this section we present two rigid-body experiments. The first experiment models a person being conveyed past an imaging system by using a mannequin on a 2D translation stage. The second experiment models a common SAR scanning configuration used in airport security by using a mannequin on a rotation stage.

The key task of rigid registration is to measure rigid scene pose A_{gb} with respect to the global imager basis g . For the controlled experiments in this section we choose to use photogrammetry with OPA (Figure 8). A Creaform MaxSHOT 3D camera is used to measure the location of point-like

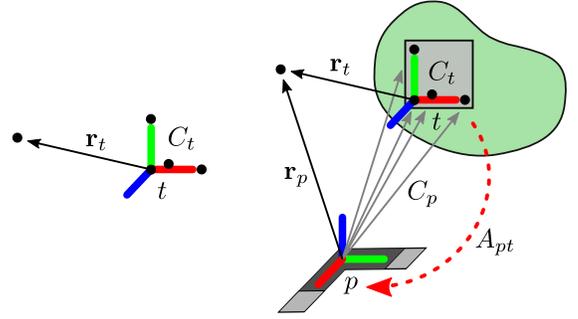


Fig. 8. Photogrammetry Registration — Constellation C_t is designed relative to template basis t . A physical constellation is affixed to a rigid body and measured relative to photogrammetry basis p . OPA is applied to estimate pose A_{pt} .

reflective sticker fiducials relative to a photogrammetry prop basis p to produce a point cloud accurate to within imager tolerances [34] [35]. Three or more fiducials are patterned in an asymmetric constellation on an object to unambiguously encode the scene pose. To isolate the constellation in the presence of other points, a special numbered fiducial is placed at the center of the constellation to group points by radius. The first time the constellation C_p is measured with respect to basis p , the points are used to construct a template basis t into which they are transformed and recorded as constellation C_t in a template file. Subsequent measurements of constellation C_p are compared against the template file using OPA to determine pose A_{pt} from template basis t to photogrammetry basis p

$$A_{pt} = \text{Procrustes}(C_t, C_p) \quad (9)$$

The pose for every antenna is determined in the same manner. Choosing a particular antenna to serve as the global basis g , the pose A_{pg} is defined. Using the photogrammetry prop as an intermediate basis, the pose A_{gt} is computed as

$$A_{gt} = A_{gp} \circ A_{pt} \quad (10)$$

If we affix a constellation to a stage, we can rename template basis t to scene basis b to measure pose A_{gb} .

There is nothing special about photogrammetry, and in general any sensor system capable of imaging points in \mathbb{R}^3 may be used to experimentally measure poses.

A. Translation stitching

In this experiment, a mannequin is placed on a 2D translation stage. The stage can be programmed with a sequence of stage parameter vectors $\mathbf{x} \in \mathbb{R}^2$ to follow any trajectory in a plane. To simplify the discussion we restrict our attention to linear trajectories modeling a person being conveyed past an imaging system. The experiment proceeds by stepping a mannequin along a linear trajectory and collecting image data at each step.

The stage is equipped with a photogrammetry constellation so pose A_{gb} can be experimentally registered by (10). This is a time consuming process, so we choose to measure the pose

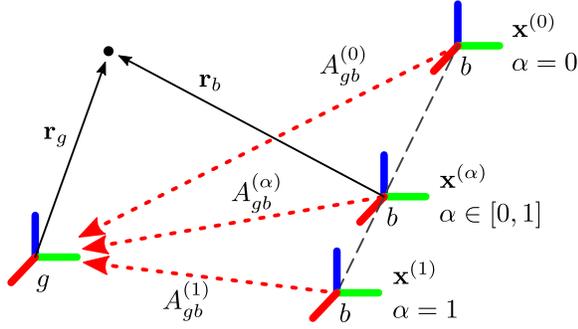


Fig. 9. Translation Registration — A linear stage affixed with basis b is translated relative to global imager basis g . Terminal stage configurations $\mathbf{x}^{(0)}$ and $\mathbf{x}^{(1)}$ and associated poses $A_{gb}^{(0)}$ and $A_{gb}^{(1)}$ are measured, and linearly interpolated with parameter α to compute $\mathbf{x}^{(\alpha)}$ and $A_{gb}^{(\alpha)}$.

for a small number of stage configurations and to compute arbitrary poses from those measurements.

A linear trajectory can be constructed by linearly interpolating a pair of stage parameter vectors and their corresponding poses (Figure 9). The linear interpolation, or *lerp*, of two properties p_0 and p_1 by parameter $\alpha \in \mathbb{R}$ is defined as

$$\text{lerp}(p_0, p_1, \alpha) \equiv p_0 + (p_1 - p_0)\alpha. \quad (11)$$

Given stage parameter vectors $\mathbf{x}^{(0)}$ and $\mathbf{x}^{(1)}$ with associated measured poses $A_{gb}^{(0)}$ and $A_{gb}^{(1)}$ and interpolation parameters $\alpha = 0$ and $\alpha = 1$, we compute $\mathbf{x}^{(\alpha)}$ and $A_{gb}^{(\alpha)}$ for any α as

$$\mathbf{x}^{(\alpha)} = \text{lerp}(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \alpha) \quad (12)$$

$$A_{gb}^{(\alpha)} = \text{lerp}(A_{gb}^{(0)}, A_{gb}^{(1)}, \alpha). \quad (13)$$

For a linear trajectory with $2 \leq N \in \mathbb{Z}$ points and equal-sized steps starting at $A_{gb}^{(0)}$ and ending at $A_{gb}^{(1)}$, let $\alpha = (n - 1)/(N - 1)$ where $n \in [1, N] \subset \mathbb{Z}$.

With image registration solved, we apply the stitching algorithm. Experimental image data and a stitch of a mannequin undergoing a linear translation trajectory past an imaging system are shown in Figure 10. It is clear that individual images suffer from specularly-limited coverage, while the stitch improves coverage on the side of the mannequin closest to the imager. This experiment indicates that a flat imaging system from a single perspective will be limited by specularly even when images are stitched, which suggests that different antenna configurations and trajectories should be considered.

B. Rotation stitching

A more interesting rigid body scene trajectory is pure rotational motion. Rotation mitigates specularly-limited scene coverage by viewing a scene from all sides. Imaging a rotating scene with a fixed aperture is equivalent to imaging a fixed scene with an aperture counter-rotating around the scene. This mimics a common configuration found in commercial mechanically scanned SAR systems, permitting a fair comparison between imagers to be made. However, it should be noted that SAR coherently reconstructs a single image based on

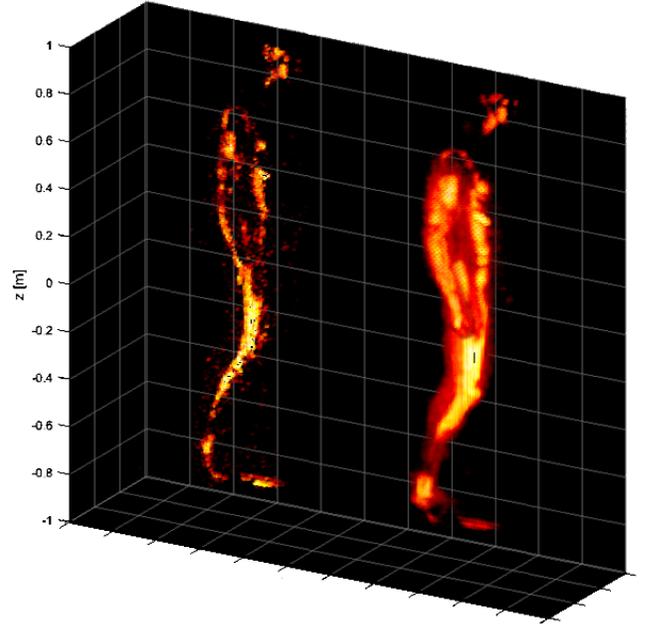


Fig. 10. Experimental Translation Stitching — A single image (left) is compared against the stitch (right). Note the gains in coverage on the limbs.

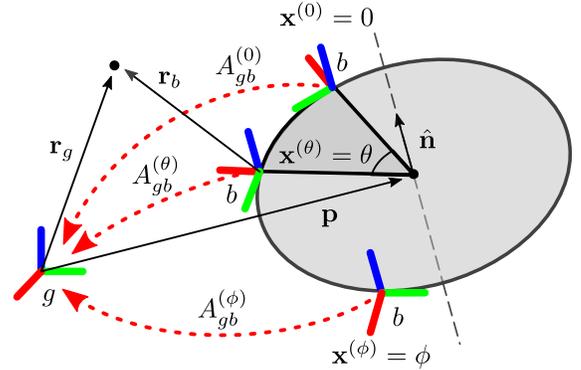


Fig. 11. Rotation Registration — A rotation stage affixed with basis b is rotated relative to global basis g . For stage configurations $\mathbf{x}^{(0)}$ and $\mathbf{x}^{(\phi)}$ associated poses $A_{gb}^{(0)}$ and $A_{gb}^{(\phi)}$ are measured. From these poses the stage axis $\hat{\mathbf{n}}$ and a point on that axis \mathbf{p} are deduced, and pose $A_{gb}^{(\theta)}$ for any stage configuration $\mathbf{x}^{(\theta)}$ is computed.

all collected measurements, while a real-time imager coherently reconstructs a sequence of images that are incoherently stitched into a single image.

In this experiment, a mannequin is placed on a rotation stage. The stage can be programmed with a sequence of stage configuration angles $\mathbf{x}^{(\theta)} = \theta \in [0, 2\pi) \subset \mathbb{R}$ to follow any rotational trajectory. The experiment proceeds by stepping a mannequin along a rotation trajectory and collecting image data at each step.

Once again we choose to use photogrammetry to experimentally register scene pose A_{gb} for a small number of stage configurations to compute arbitrary poses (Figure 11). A rotation is fully described by axis direction $\hat{\mathbf{n}} \in S^2$, axis point $\mathbf{p} \in \mathbb{R}^3$, and angle θ . $\hat{\mathbf{n}}$ and \mathbf{p} are constant for any stage

pose, so we estimate these to construct new poses. Given stage configuration angles $\mathbf{x}^{(0)} = 0$ and $\mathbf{x}^{(\phi)} = \phi \neq 0$ with associated measured poses $A_{gb}^{(0)}$ and $A_{gb}^{(\phi)}$, we compute pose $A_{gg}^{(\phi)}$ which encodes a rotation of ϕ about the stage axis relative to global basis g

$$A_{gg}^{(\phi)} = A_{gb}^{(\phi)} \circ A_{gb}^{(0)}. \quad (14)$$

To understand this equation consider scene basis b as fixed while basis g is transformed, and recognize this is equivalent to transforming points relative to fixed basis g by Figure 4. We extract from $A_{gg}^{(\phi)}$ rotation matrix $\mathbf{R}_{gg}^{(\phi)}$ and convert it to axis-angle representation [36] to isolate axis direction $\hat{\mathbf{n}}$. However, the translation vector $\mathbf{t}_{gg}^{(\phi)}$ is not necessarily an axis point \mathbf{p} . This is because $\mathbf{R}_{gg}^{(\phi)}$ represents a rotation about the global origin and not the stage axis. To compute axis point \mathbf{p} , we observe \mathbf{p} must be unchanged by pose $A_{gg}^{(\phi)}$

$$\mathbf{p} = A_{gg}^{(\phi)}(\mathbf{p}) = \mathbf{R}_{gg}^{(\phi)}\mathbf{p} + \mathbf{t}_{gg}^{(\phi)}. \quad (15)$$

An axis is a line in 3D, so an infinite number of points satisfy this equation. To guarantee a unique solution we solve for \mathbf{p} in the 2D planar subspace perpendicular to axis direction $\hat{\mathbf{n}}$. The projection of $\mathbf{t}_{gg}^{(\phi)}$ on this subspace is

$$\mathbf{t}_{gg}^{(\phi)'} = \mathbf{t}_{gg}^{(\phi)} - (\mathbf{t}_{gg}^{(\phi)} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}. \quad (16)$$

Recasting (15) in 2D (denoted by primes), we solve for \mathbf{p}'

$$\mathbf{p}' = (\mathbf{I}' - \mathbf{R}_{gg}^{(\phi)'})^{-1}\mathbf{t}_{gg}^{(\phi)'}. \quad (17)$$

where \mathbf{I}' is the identity matrix. Expanding \mathbf{p}' relative to the 3D global basis g yields axis point \mathbf{p} . We can now construct scene pose $A_{gb}^{(\theta)}$ for arbitrary stage configuration angle $\mathbf{x}^{(\theta)} = \theta$. First construct pose $A_{gg}^{(\theta)}$ by moving the scene axis to the global origin by subtracting \mathbf{p} , then rotate about axis direction $\hat{\mathbf{n}}$ by angle θ , and finally restore the axis to its original position by adding \mathbf{p}

$$A_{gg}^{(\theta)}(\mathbf{r}_g) = \mathbf{R}(\hat{\mathbf{n}}, \theta)(\mathbf{r}_g - \mathbf{p}) + \mathbf{p}. \quad (18)$$

Functional composition with $A_{gb}^{(0)}$ yields the desired pose

$$A_{gb}^{(\theta)} = A_{gg}^{(\theta)} \circ A_{gb}^{(0)}. \quad (19)$$

For a rotation trajectory with $1 \leq N \in \mathbb{Z}$ points and equal-angle steps starting at $A_{gb}^{(0)}$, let $\mathbf{x}^{(\theta)} = \theta = 2\pi(n-1)/N$ where $n \in [1, N] \subset \mathbb{Z}$.

With registration solved, the stitching algorithm is applied. Experimental image data and a stitch of a mannequin undergoing a full rotation trajectory in front of an imaging system are shown in Figure 3 and Figure 12. The stitch exhibits excellent scene coverage comparable to SAR systems [5] limited only by occlusion and specular surfaces significantly aligned with the axis. This results emphasizes the benefits of tightly controlling the imaging environment and imaging a scene from multiple perspectives. Unfortunately we must give up some of that control to stitch people in motion.

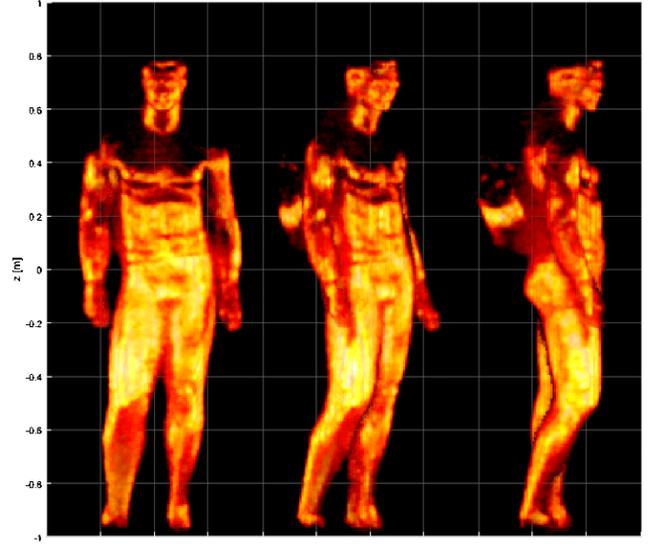


Fig. 12. Experimental Rotation Stitch — A stitch of a mannequin wearing a backpack containing a pressure cooker is shown from three perspectives. Compare to raw image Figure 3. Coverage is excellent, except for the top of the torso due to specularities. Images are comparable to SAR systems.

V. SKELETON STITCHING

The primary difficulty with stitching people in motion is that the scene is not a rigid body, but instead one that deforms, dramatically complicating the registration process. A single rigid transformation is no longer sufficient to fully characterize the scene pose. Instead, we segment the scene into a set of parts whose poses can be locally approximated by rigid transformations and combined with a deformation model.

It is impractical for security applications to affix photogrammetry constellations to people being imaged, therefore a different approach from section IV is needed to measure the pose of people. A serendipitous solution to this problem is to repurpose the depth cameras used to constrain the inverse imaging problem. Our imager incorporates Kinect depth cameras which were originally designed as interfaces for video game systems and can fit skeleton armatures to people to approximate their pose.

The Kinect skeleton is organized as a set of *bones*, $b \in B$ whose individual poses are described by rigid transformation (Figure 13). The Kinect skeleton contains only about 25 bones, which is clearly not anatomically correct, reflecting the Kinect's intended usage as an entertainment device rather than a precision instrument. Thus, the Kinect skeleton is only a first order approximation of the scene pose. Regardless, it is a strong prior that can bootstrap more sophisticated registration in future research. Poses measured relative to depth camera basis c must be mapped to global basis g to perform stitching, so depth camera pose A_{gc} must first be measured by registration.

Registration of a depth camera with the imager was tacitly assumed when describing how depth information is used to constrain the inverse imaging problem (2). The registration process is a permutation of the ideas used to register the

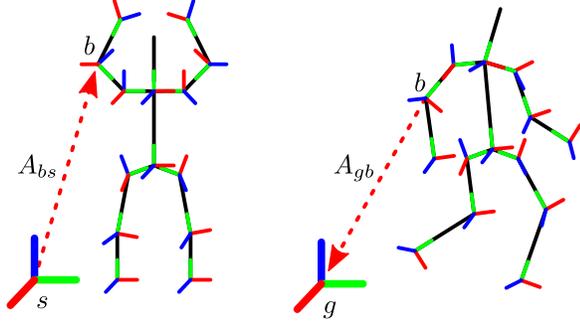


Fig. 13. Skeleton Registration — A Kinect skeleton models a person’s pose with a set of bones $b \in B$ and associated rigid transformations. The skeleton can take on any rest pose A_{bs} relative to stitch basis s (left). For each image, the skeleton pose A_{gb} is measured relative to global basis g (right).

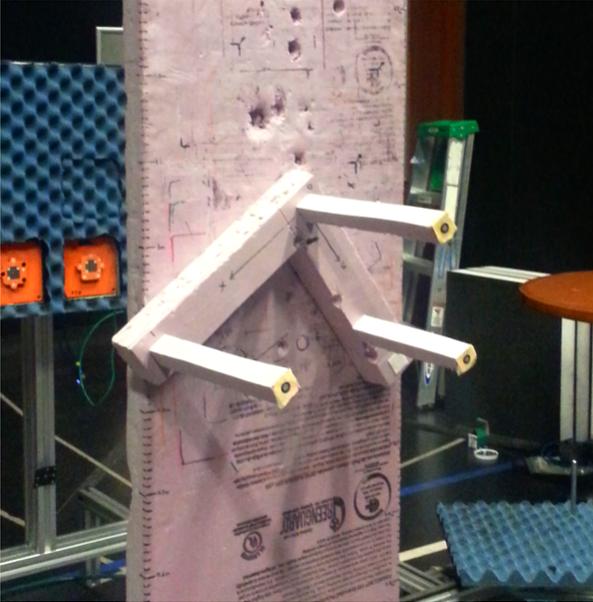


Fig. 14. Multisensor Registration Constellation Target — A constellation distinguishable in microwave conductivity and geometric depth is used to register the imager with depth cameras. Points consist stalks tipped with metal foil. Masking tape covers the foil to improve depth camera visibility. Photogrammetry fiducials are included to generate an accurate template file.

stages in section IV. A multisensor registration constellation target is constructed with points distinguishable in microwave conductivity and geometric depth (Figure 14). Each point of the target consists of a long narrow stalk tipped with a layer of metal foil, masking tape, and a photogrammetry fiducial. The foil acts like a fiducial for the microwave imaging system, while the stalk acts like a fiducial for the depth camera. The masking tape provides a matte surface to improve the visibility of the tip to the depth camera, which is confused by specular surfaces like the shiny foil. Initially the target is measured with photogrammetry to create a highly accurate template file constellation C_m relative to the multisensor target basis m . The target is then measured simultaneously with the imager and depth camera to yield constellations C_g and C_d relative to the global basis g and depth camera basis d respectively. OPA

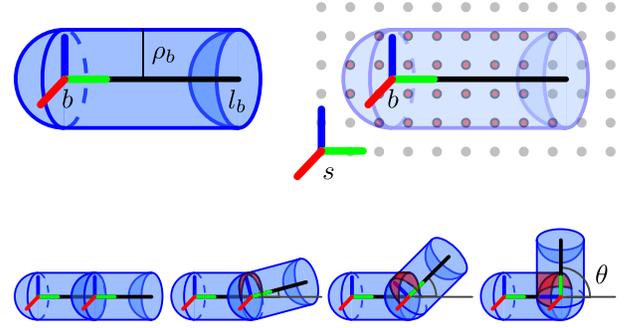


Fig. 15. Rigid Subvolume — For each bone $b \in B$, a rigid pill characterized by radius ρ_b and length l_b is defined (top left). A grid relative to stitch basis s is partitioned by the rest pose subvolumes (top right). Transformed subvolumes overlap (red) when adjacent bones are not colinear (bottom).

compares C_g and C_d to C_m to yield poses A_{mg} and A_{md} . Pose A_{gd} is then computed as

$$A_{gd} = A_{mg} \circ A_{md}. \quad (20)$$

With the depth cameras registered, depth and skeleton information are transformed into the global basis g to inform imaging and stitching. The skeleton by itself is not immediately useful for stitching. We must first associate the skeleton with a geometric “skin”. Here we experimentally demonstrate a simple model that segments the scene into rigid subvolumes attached to bones. We then extend this model to account for deformation by using techniques from computer graphics, and demonstrate its application in simulation.

A. Rigid Subvolume Stitching

A simple way to associate a skeleton with its surrounding volume is to define a rigid subvolume attached to each bone (Figure 15). An obvious candidate is an axis-aligned cylinder of some predetermined radius ρ_b and length l_b . Points are easily tested to be within this cylinder by checking if their projection on the axis falls within the interval of the bone and if their distance from the axis is less than the radius.

While a cylinder is a good subvolume for a single bone, cylinders do not smoothly connect at skeletal joints. A remedy is to cap the cylinder with spheres to create a pill-shaped subvolume. In this way, joints are much better represented. However, using a pill causes significant subvolume overlap around joints, double counting a large number of points. To mitigate this issue, the spherical cap belonging to the non-basis end of a bone is made to exclude points, resulting in a “dented” pill. When these subvolumes are mapped to a skeleton the overlap at joints is minimized.

Before stitching a person in motion, an arbitrary skeleton must be selected to serve as the rest pose which defines the pose of the stitched image (Figure 13). Conveniently, any experimental skeleton suffices, although rest poses with limbs spread apart are likely best suited for ATD. A grid of points is then defined relative to the stitch basis s . For each bone $b \in B$ with associated rest pose A_{bs} , a list of grid points $\mathbf{r}_{s,i} \in \mathbb{R}^3$ within the bone subvolume is tabulated, where the

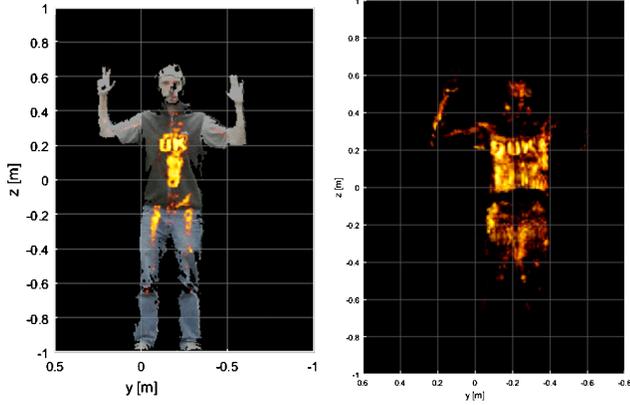


Fig. 16. Experimental Rigid Subvolume Stitch — In no single image are the letters all visible (left); note the camera overlay. The the word "DUKE" is clearly read in the stitch (right). Note the crude stitching around joints caused by the simplistic registration model, particularly around the hips and thighs.

index $i \in I(b) \subset \mathbb{Z}$ enumerates points belonging to bone b . To stitch image data of a person in motion, skeleton data is measured for each frame to obtain the measured bone poses A_{gb} with respect to global basis g . For every bone $b \in B$, the associated list of grid points are transformed by

$$\mathbf{r}_{g,i} = A_{gb} \circ A_{bs}(\mathbf{r}_{s,i}); \quad i \in I(b) \quad (21)$$

solving the registration task. The stitching algorithm can then be applied.

Experimental demonstration of rigid subvolume stitching is presented in Figure 16. Here the author was imaged with a set of metal letters spelling "DUKE" across his chest and hidden under a vest. The subject faced the imager with their feet planted, and rotated their torso over their full range for 60 frames. In no single image are all 4 letters visible, however when stitched the full word is seen, lending credibility to the skeleton technique. However, it is clear that the Kinect skeleton is crude, the bone subvolumes are overly simplistic, and overlap between subvolumes is problematic. A more sophisticated deformation registration model is needed to realistically stitch people in motion.

B. SKD+SSD Stitching

Accurately modeling the deformation of a solid is computationally demanding. Approximations must be made for real-time imaging systems. A simple deformation model used to great effect in computer graphics is a combination of Shape Key Deformation (SKD) and Skeleton Space Deformation (SSD) [37] (Figure 17). This model utilizes a skeleton conceptually compatible with the Kinect skeleton. SSD naturally extends rigid subvolume registration to account for realistic deformation. However, extreme deformations with SSD can cause unnatural results. SKD solves this issue by first modifying the underlying geometry as a function of skeleton pose before SSD is applied. While neither model is physically motivated, with enough effort, arbitrarily realistic deformations are possible.

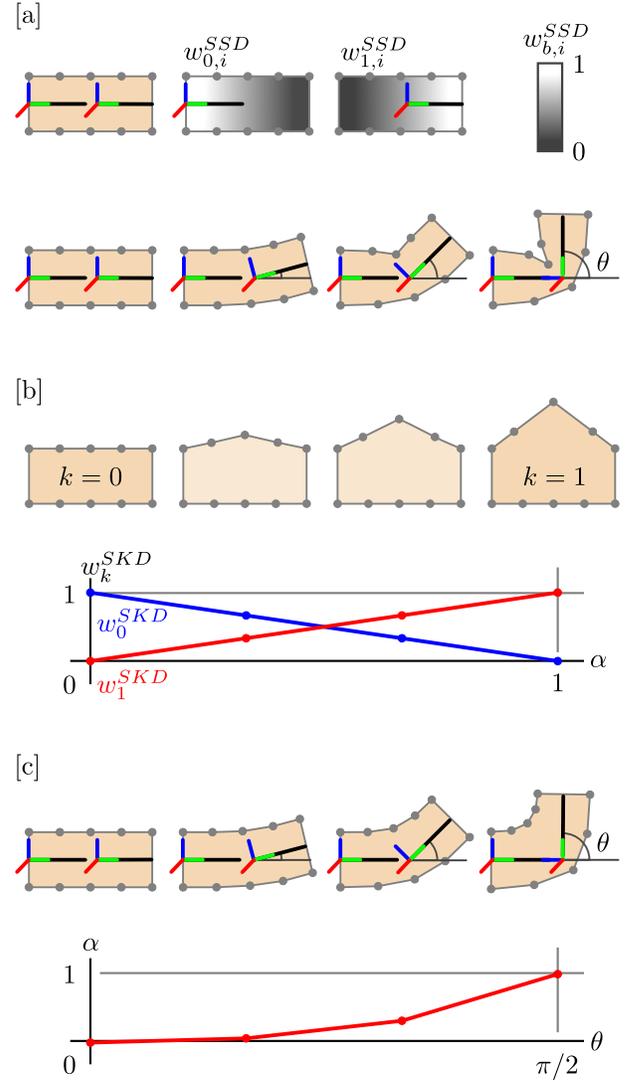


Fig. 17. SKD+SSD — [a] SSD is implemented by defining rest geometry and assigning weights $w_{b,i}^{SSD}$ for each bone and point (top). Deformation is achieved by rigidly transforming points with each bone and taking a weighed average (bottom). Unrealistic deformation occurs for extreme poses. [b] SKD forms new shapes from shape keys $k \in K$ that share the same topology by taking a weighted average of corresponding points with weights w_k^{SKD} (top). Weighting can be determined as a function of external variables, such as α in this example (bottom). [c] SKD is used to modify rest geometry before application of SSD to address problematic deformation (top). Pose variables, such as an angle θ between bones, can be mapped to SKD weight variables, such as α in this example (bottom).

SKD interpolates different *shape keys* with the same vertex topology to produce new shapes. Each shape key $k \in K$ defines a list of vertices $\mathbf{r}_{k,i}$ labeled with index $i \in I$. Stitch vertices $\mathbf{r}_{s,i}$ are computed by taking the weighted average of corresponding shape key vertices over all keys

$$\mathbf{r}_{s,i} = \sum_{k \in K} w_k^{SKD} \mathbf{r}_{k,i}; \quad \sum_{k \in K} w_k^{SKD} = 1 \quad (22)$$

where w_k^{SKD} are the shape key weights. The weights can be made to be a function of the skeleton pose. At one extreme, for every pose a different shape key could be defined,

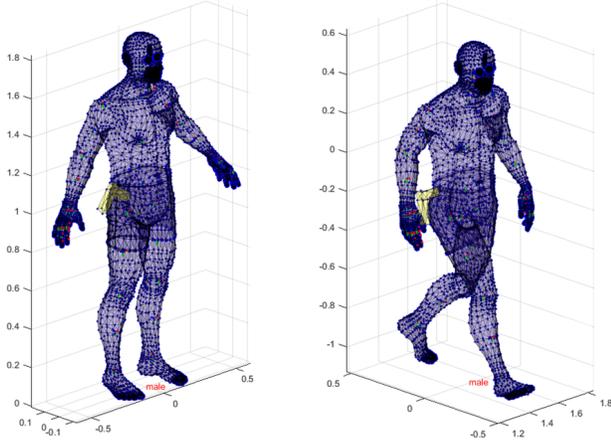


Fig. 18. SKD+SSD Puppet — A life-like human model is posed with a skeleton and deformed with SKD+SSD. Note the threat object on the hip.

which would essentially model deformation by lookup table. However this would take a prohibitive amount of memory. Instead, additional shape keys are defined only when memory-efficient SSD results in problematic deformations.

SSD transforms geometry as a function of skeleton pose which effectively reduces the number of degrees of freedom needed to describe deformation. SSD is a modification of rigid registration equation (7). For every vertex $\mathbf{r}_{s,i}$ with respect to stitch basis s , the associated vertex $\mathbf{r}_{g,i}$ with respect to global basis g is computed by taking the weighted average of rigid transformations over all bones

$$\mathbf{r}_{g,i} = \sum_{b \in B} w_{b,i}^{SSD} A_{gb} \circ A_{bs}(\mathbf{r}_{s,i}); \quad \sum_{b \in B} w_{b,i}^{SSD} = 1 \quad (23)$$

where $w_{b,i}^{SSD}$ are the deformation weights for each bone $b \in B$ and vertex index $i \in I$. SSD strikes a good balance between computational complexity and realism. However, geometry influenced by a bone that is twisted or bent to an extreme degree relative to other bones causes pinching and collapsing of the resulting deformation. SKD is therefore applied before SSD to supply geometry suitable for deformation by the given skeleton pose.

While the governing equations of the SKD+SSD model are extremely simple, generating high quality data to feed the model is not a trivial task. The open source animation software Blender is used to facilitate this process. Blender is used to prepare a representation of a person by modeling surface geometry, defining a skeleton, and weighting the geometry to the skeleton. Taken together, this data is called a *puppet* (Figure 18). The puppet geometry can be realistically deformed by specifying a skeleton pose. The skeleton pose can be animated with Blender in simulation, or experimentally measured with depth cameras. For this pioneering work, we restrict our research to simulation where the skeleton pose can be perfectly known.

Puppet and animation data are exported from Blender and imported into the imaging software. Imported geometric data is composed of vertices and triangular faces. Vertices are used

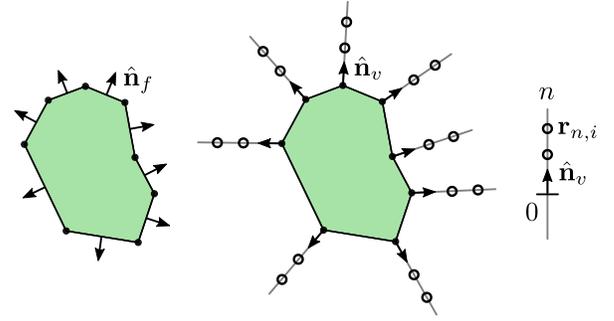


Fig. 19. Cactusing — Geometry produced by computer graphics modeling software is usually an orientable surface composed of faces with normals $\hat{\mathbf{n}}_f \in S^2$ (left). To detect “worn” features above the surface, points are added along vertex normals $\hat{\mathbf{n}}_v \in S^2$ computed from face normals (center). The distribution of points is defined by a 1D point profile $\mathbf{r}_{n,i}$ (right).

to model point scatterers in simulation. Faces are supposed to represent surfaces, however if face vertices are farther apart than half the wavelength of the probing radiation, the face must be subdivided to correctly model a surface. In addition, the faces form an orientable surface. To stitch the volume immediately surrounding the body where worn objects are found, points are added to the geometry at each vertex along the average normal of adjacent faces in a “cactusing” procedure (Figure 19).

A stitching simulation proceeds by selecting a frame of animation, assigning the corresponding skeleton pose to the puppet, transforming the puppet geometry to global basis g , and then applying the forward model (2). The inverse problem is then solved to reconstruct an image. To stitch, the skeleton pose can be reused to solve the registration problem. The stitching algorithm can then be applied. Stitch data can either be accumulated per vertex, or mapped to a regular grid relative to stitch basis s .

A simulated SKD+SSD stitch of a male puppet walking past an imager configured like the experimental system is shown in Figure 20. The puppet is equipped with a gun on his right thigh. When compared to the rigid subvolume stitch, the quality of the SKD+SSD stitch is much better: overlap is eliminated; joints are realistically deformed and stitched; the improved skeleton more accurately represents the pose of the scene; and the improved geometry better matches the human body. Whereas ATD could miss the threat object in any single image, it is plausible that ATD would easily identify the threat on the persons thigh in the stitched image.

With perfect knowledge of the skeleton pose, this kind of simulation represents the gold standard for SKD+SSD stitching. The problem of experimentally estimating human pose has been extensively studied in the computer vision literature [38], but is still an active area of research. Combining the SKD+SSD deformation model with available experimental skeleton pose estimation is the next step in research.

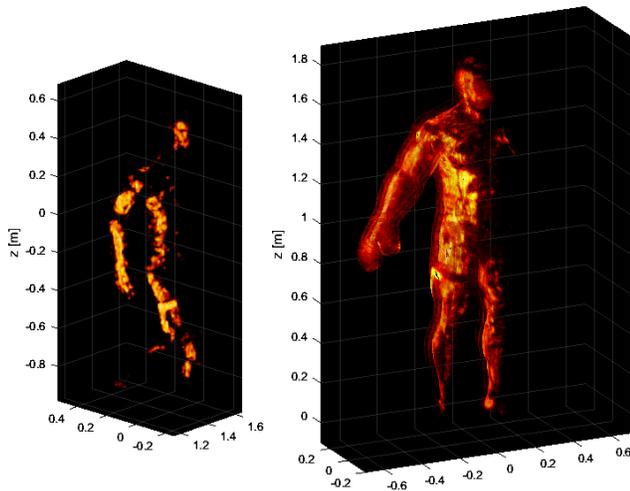


Fig. 20. Simulated SKD+SSD Stitch — A male puppet wearing a threat object was animated walking past a simulated copy of the experimental imager. Specularity-limited coverage critically restricts the number of frames the threat is visible (left). Stitching with the same puppet greatly improves coverage, making the threat easily visible in a standardized pose (right).

VI. CONCLUSION

In this work we have demonstrated the ability to stitch images of people in motion (and repose them in a standardized pose) from a sequence of images from a screen-while-walk millimeter wave imager. By leveraging the motion of a person moving through the imaging volume, a diversity of images with different perspectives and coverage are obtained which can be stitched together into a single image of the scene with better coverage. The stitching concepts of registration, calibration, and blending were discussed in relation to the experimental imager. To handle the complexities of stitching a deforming body, a Kinect depth camera was used to experimentally estimate the skeleton pose of a person, and a model subdividing the scene into rigid subvolumes produced a stitch that demonstrated the technique was viable. A more realistic SKD+SSD deformation model was formulated, and it was shown in simulation to effectively stitch a person in motion assuming the pose skeleton can be estimated.

With the introduction of SKD+SSD deformation and its incorporation into a full imaging system simulation suite, a host of future research questions can be answered without the need to construct physical experiments. For instance, the layout of the imager and trajectory a person takes through the imaging volume must be optimized to maximize coverage while minimizing system costs. Hallways, U-turns, dog-legs, and arches are all being considered, and can be rapidly evaluated on a computer. In addition, a calibration solution must be developed so images can be properly combined. We hypothesize a calibration object present in every image could be used to normalize the magnitude of the images. With good coverage of the scene established, a process that matches a person to a puppet from a library must be developed to stitch a wide variety of body shapes. The most challenging unsolved task is to estimate the skeleton pose of a real person in the scene which is beyond the limitations of the Kinect skeleton

functionality utilized in experiment.

The ultimate metric for the usefulness of stitching images of people in motion is how ATD on raw images compares to ATD on stitches, or images processed in some other way. We suggest stitching as a preprocessing step may significantly simplify and improve ATD analysis, but this remains an open question.

ACKNOWLEDGMENT

This work was supported by the Department of Homeland Security, Science and Technology Directorate (Contract No. HSHQDC-12-C-00049). The published material represents the position of the author(s) and not necessarily that of the DHS

REFERENCES

- [1] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges," *Information Fusion*, vol. 35, pp. 68–80, 2017.
- [2] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE multimedia*, vol. 19, no. 2, pp. 4–10, 2012.
- [3] H. Cho, Y.-W. Seo, B. V. K. V. Kumar, and R. R. Rajkumar, "A multi-sensor fusion system for moving object detection and tracking in urban driving environments," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1836–1843, IEEE, 2014.
- [4] D. M. Sheen, D. L. McMakin, and T. E. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 9, pp. 1581–1592, 2001.
- [5] D. M. Sheen, D. L. McMakin, H. D. Collins, T. E. Hall, and R. H. Severtsen, "Concealed explosive detection on personnel using a wideband holographic millimeter-wave imaging system," in *SPIE Signal Processing, Sensor Fusion, and Target Recognition V*, vol. 2755, pp. 503–514, International Society for Optics and Photonics, 1996.
- [6] Rohde and Schwarz, "Rohde and Schwarz QPS quick personnel security scanners." Brochure, 2016.
- [7] D. M. Sheen, T. E. Hall, D. L. McMakin, A. M. Jones, and J. R. Tedeschi, "Three-dimensional radar imaging techniques and systems for near-field applications," in *SPIE Radar Sensor Technology XX*, vol. 9829, International Society for Optics and Photonics, 2016.
- [8] F. Gumbmann and S. S. Ahmed, "Walk through screening with multistatic mmW technology," in *SPIE Millimeter Wave and Terahertz Sensors and Technology IX*, vol. 9993, p. 999306, International Society for Optics and Photonics, 2016.
- [9] A. Pedross-Engel, D. Arnitz, J. N. Gollub, O. Yurduseven, K. P. Trofater, M. F. Imani, T. Sleasman, M. Boyarsky, X. Fu, D. L. Marks, et al., "Orthogonal coded active illumination for millimeter wave, massive-mimo computational imaging with metasurface antennas," *IEEE Transactions on Computational Imaging*, vol. 4, no. 2, pp. 184–193, 2018.
- [10] K. J. Liang, G. Heilmann, C. Gregory, S. O. Diallo, D. Carlson, G. P. Spell, J. B. Sigman, K. Roe, and L. Carin, "Automatic threat recognition of prohibited items at aviation checkpoint with x-ray imaging: a deep learning approach," in *SPIE Anomaly Detection and Imaging with X-Rays (ADIX) III*, vol. 10632, p. 1063203, International Society for Optics and Photonics, 2018.
- [11] S. K. Nayar, "Computational cameras: approaches, benefits and limits," *Columbia University Computer Science Technical Reports*, p. 6, 2011.
- [12] E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Applied optics*, vol. 17, no. 3, pp. 337–347, 1978.
- [13] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, "A single-pixel terahertz imaging system based on compressed sensing," *Applied Physics Letters*, vol. 93, no. 12, p. 121105, 2008.
- [14] J. Hunt, T. Driscoll, A. Mrozack, G. Lipworth, M. Reynolds, D. Brady, and D. R. Smith, "Metamaterial apertures for computational imaging," *Science*, vol. 339, no. 6117, pp. 310–313, 2013.
- [15] G. Lipworth, A. Mrozack, J. Hunt, D. L. Marks, T. Driscoll, D. Brady, and D. R. Smith, "Metamaterial apertures for coherent computational imaging on the physical layer," *Journal of the Optical Society of America A*, vol. 30, no. 8, pp. 1603–1612, 2013.

- [16] T. Sleasman, M. F. Imani, J. N. Gollub, and D. R. Smith, "Dynamic metamaterial aperture for microwave imaging," *Applied Physics Letters*, vol. 107, no. 20, p. 204104, 2015.
- [17] D. L. Marks, O. Yurduseven, and D. R. Smith, "Cavity-backed metasurface antennas and their application to frequency diversity imaging," *Journal of the Optical Society of America A*, vol. 34, no. 4, pp. 472–480, 2017.
- [18] O. Yurduseven, J. N. Gollub, D. L. Marks, and D. R. Smith, "Frequency-diverse microwave imaging using planar mills-cross cavity apertures," *Optics Express*, vol. 24, no. 8, pp. 8907–8925, 2016.
- [19] J. N. Gollub, O. Yurduseven, K. P. Trofatter, D. Arnitz, M. F. Imani, T. Sleasman, M. Boyarsky, A. Rose, A. Pedross-Engel, H. Odabasi, et al., "Large metasurface aperture for millimeter wave computational imaging at the human-scale," *Scientific Reports*, vol. 7, p. 42650, 2017.
- [20] G. Lipworth, A. Rose, O. Yurduseven, V. R. Gowda, M. F. Imani, H. Odabasi, P. Trofatter, J. Gollub, and D. R. Smith, "Comprehensive simulation platform for a metamaterial imaging system," *Applied Optics*, vol. 54, no. 31, pp. 9343–9353, 2015.
- [21] L. Pulido-Mancera, T. Fromenteze, T. Sleasman, M. Boyarsky, M. F. Imani, M. Reynolds, and D. Smith, "Application of range migration algorithms to imaging with a dynamic metasurface antenna," *Journal of the Optical Society of America B*, vol. 33, no. 10, pp. 2082–2092, 2016.
- [22] D. L. Marks, O. Yurduseven, and D. R. Smith, "Fourier accelerated multistatic imaging: a fast reconstruction algorithm for multiple-input-multiple-output radar imaging," *IEEE Access*, vol. 5, pp. 1796–1809, 2017.
- [23] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.
- [24] S. Mann and R. W. Picard, "Virtual bellows: Constructing high quality stills from video," in *Proceedings of 1st International Conference on Image Processing*, vol. 1, pp. 363–367, IEEE, IEEE Comput. Soc. Press, 1994.
- [25] R. Szeliski, H.-Y. Shum, H.-Y. Shum, and H.-Y. Shum, "Creating full view panoramic image mosaics and environment maps," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pp. 251–258, ACM Press/Addison-Wesley Publishing Co., 1997.
- [26] L. G. Brown, "A survey of image registration techniques," *ACM computing surveys (CSUR)*, vol. 24, pp. 325–376, Dec. 1992.
- [27] T. A. A. Alzohairy, E. El-Dein, et al., "Image Mosaicing based on Neural Networks," *International Journal of Computer Applications*, vol. 975, p. 8887, feb 2016.
- [28] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 7th IEEE International Conference on Computer Vision*, vol. 2, pp. 1150–1157, Ieee, 1999.
- [29] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: speeded up robust features," in *European Conference on Computer Vision*, pp. 404–417, Springer, 2006.
- [30] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment – a modern synthesis," in *International Workshop on Vision Algorithms*, pp. 298–372, Springer, 1999.
- [31] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in *Readings in Computer Vision*, pp. 726–740, Elsevier, 1987.
- [32] P. Scovanner, S. Ali, and M. Shah, "A 3-dimensional sift descriptor and its application to action recognition," in *Proceedings of the 15th ACM international conference on Multimedia*, pp. 357–360, ACM, ACM Press, 2007.
- [33] V. R. S. Mani and S. Arivazhagan, "Survey of medical image registration," *Journal of Biomedical Engineering and Technology*, vol. 1, pp. 8–25, mar 2013.
- [34] H. Odabasi, M. F. Imani, G. Lipworth, J. Gollub, and D. R. Smith, "Investigation of alignment errors on multi-static microwave imaging based on frequency-diverse metamaterial apertures," *Progress In Electromagnetics Research*, vol. 70, pp. 101–112, 2016.
- [35] Creaform, "MaxSHOT-3D optical coordiante measuring system." Brochure, 2014.
- [36] D. Eberly, "Rotation representations and performance issues." Self, 2002.
- [37] J. P. Lewis, M. Corder, and N. Fong, "Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 165–172, ACM Press/Addison-Wesley Publishing Co., 2000.

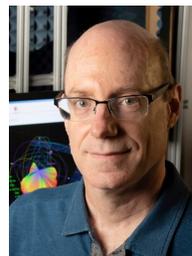
- [38] M. B. Holte, C. Tran, M. M. Trivedi, and T. B. Moeslund, "Human pose estimation and activity recognition from multi-view videos: comparative explorations of recent developments," *IEEE Journal of selected topics in signal processing*, vol. 6, no. 5, pp. 538–552, 2012.



K. Parker Trofatter Graduated from the University of Tennessee Knoxville with a B.S in Engineering Physics and a B.S. in Computer Science. He worked briefly at Duke University as an associate in research in David Smith's laboratory on metamaterial microwave imaging, and has since joined David's group to continue research as a Ph.D. student studying Electrical Engineering. His interests are broad, and span physics, mathematics, computer science, and computer graphics.



Jonah N. Gollub Received a B.A. degree in physics from Reed College in 2000, and a Ph.D., also in physics, from the University of California, San Diego, in 2009. His thesis work covered a variety of metamaterial focused topics, including the hybridization of metamaterials with magnetic materials and the demonstration of magnetic surface plasmon polaritons on metamaterials. From 2010 to 2013 he worked as a lead modeling and simulation scientist at a startup company developing surface metamaterials, with applications for imaging and biological detection under DARPA, MDA, Army, and NSF funded efforts. He joined Duke University as a research scientist in 2013 to focus on the development of a real-time millimeter wave imaging system utilizing frequency diverse metasurface antennas and computational imaging techniques.



David R. Smith r. David R. Smith is the James B. Duke Distinguished Professor of the Electrical and Computer Engineering Department at Duke University, where he also serves as Director for the Center for Metamaterial and Integrated Plasmonics. He holds a secondary faculty appointment in the Physics Department at Duke University and is a Visiting Professor of Physics at Imperial College, London. Dr. Smith received his Ph.D. in 1994 in Physics from UCSD. Dr. Smith's research interests include the theory, simulation and characterization of unique structures across the electromagnetic spectrum, including photonic crystals, metamaterials and plasmonic nanostructures. Smith has over 300 publications in the area of metamaterials, and more than 100 patents and patent filings. Smith and his colleagues demonstrated the first left-handed (or negative index) metamaterial at microwave frequencies in 2000. In 2006, Smith and colleague Sir John Pendry reported a new electromagnetic design approach, now termed transformation optics, and suggested the possibility of a metamaterial invisibility cloak. Smith's group subsequently demonstrated a metamaterial invisibility cloak later in 2006. Dr. Smith was part of a five member team that received the Descartes Research Prize in 2005, awarded by the European Union, for contributions to metamaterials and other novel electromagnetic materials. Continually since 2009, Dr. Smith has been named a "Citation Laureate" by Clarivate Analytics Web of Science, for having among the most number of highly cited papers in the field of Physics. Dr. Smith is a co-recipient of the McGroddy Prize for New Materials, awarded by the American Physical Society, for the discovery of metamaterials (2013). In 2016, Dr. Smith was elected to the National Academy of Inventors. Dr. Smith has recently been active in transitioning metamaterial concepts for commercialization, being a co-founder of Evolv Technology, Echodyne Corporation, Pivotal Communications, and advisor to Kymeta Corporation and Lumotive Corporation. All companies devoted to developing metamaterial products. More recently, Dr. Smith has led efforts to apply metasurface apertures for use in computational imaging, with an emphasis on security screening using millimeter waves.