Capturing the Human Figure Through a Wall

Fadel Adib, Chen-Yu, Hsu Hongzi, Mao Dina, Katabi, Frédo Durand
Massachusetts Institute of Technology

ACM Transactions on Graphics (TOG), 2015

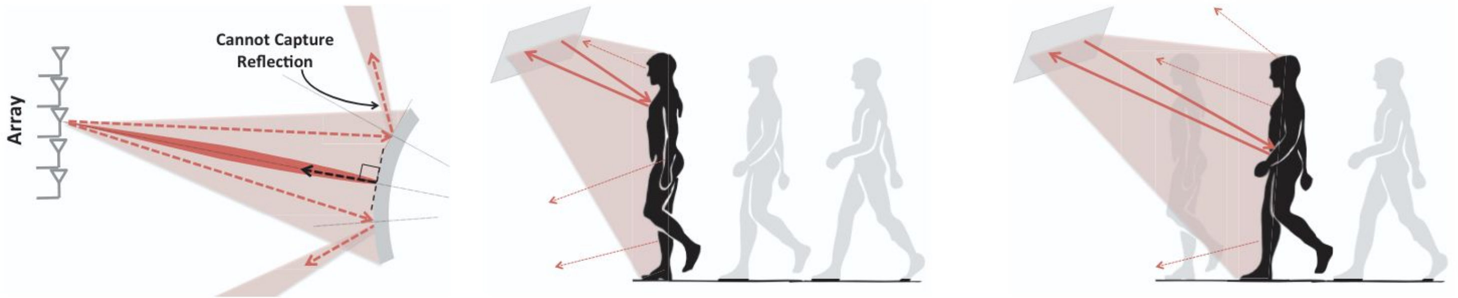
The setup

- Bandwidth : 5-7 GHz ≈ 5 cm
- Low power (1/1000 of Wi-Fi)

Millimeter waves	occlusion and shorter distances
Centimeter waves	lower resolution or simple point in space

Pipeline

1. Coarse-to-fine algorithm that efficiently scans 3D space looking for RF reflections of various human limbs and generating 3D snapshots of those reflections.
2. Motion-based : consecutive RF snapshots tend to expose different body parts and diverse perspectives of the same body part. – needs real time

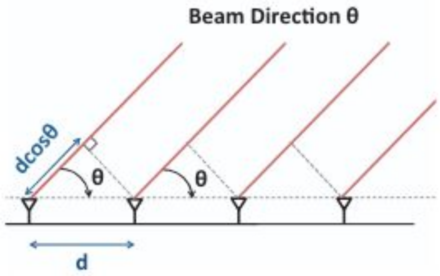


(a) Simple Reflector (b) Reflections off Human Body (c) Exploiting Motion to Capture Various Body Parts

Let’s measure!

Spatial direction - antenna arrays

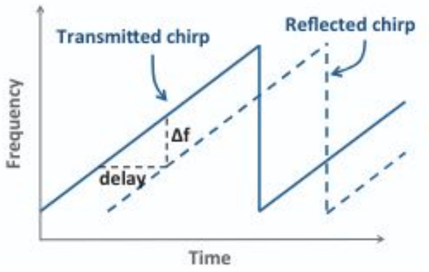
$$P(\theta) = \left| \sum_{n=1}^{N} s_n e^{j2\pi \frac{nd\cos\theta}{\lambda}} \right|$$



(a) Antenna Array

Depth - Frequency Modulated Carrier Wave (FMCW)

$$P(r) = \left| \sum_{t=1}^{T} s_t e^{j2\pi \frac{kr}{c} t} \right|$$



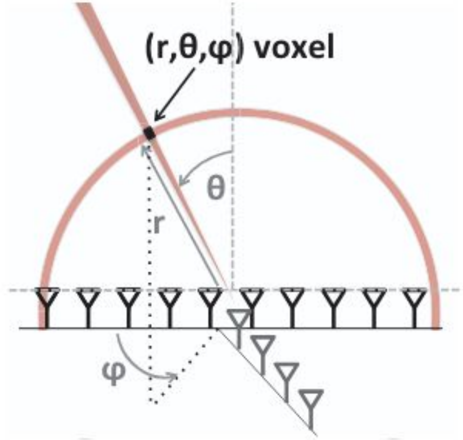
(b) Frequency Chirp

To remove background, simply subtract it from a known empty room.

Coarse-to-fine

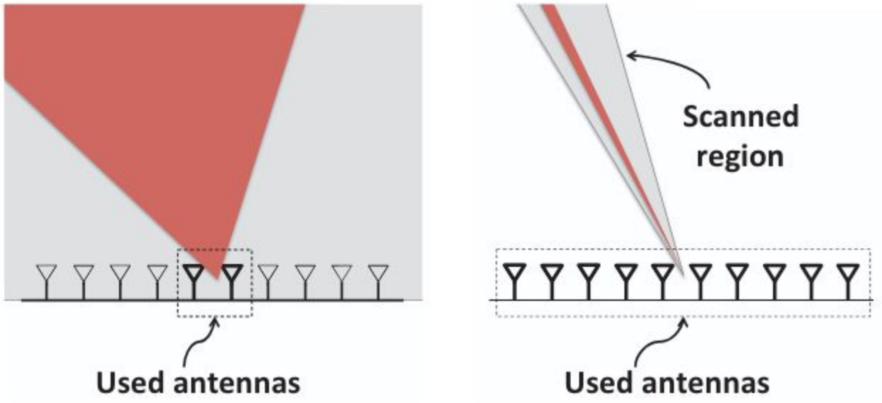
Power from a voxel

$$P(r,\theta,\phi) = \left|\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{t=1}^{T} s_{n,m,t}\, e^{j2\pi\frac{kr}{c}t}\, e^{j\frac{2\pi}{\lambda}\sin\theta(nd\cos\phi + md\sin\phi)}\right|$$



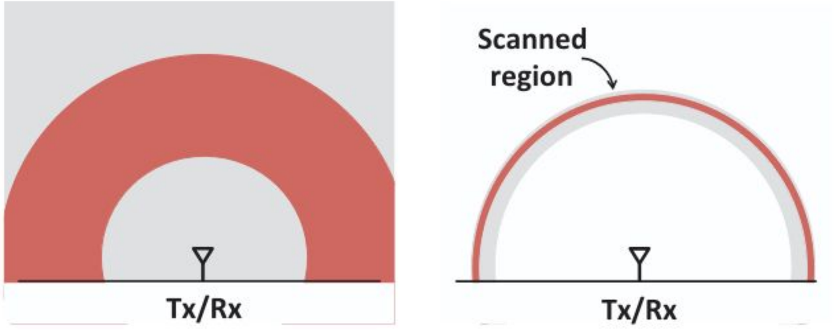
Scanning all voxels would take too long…

Coarse-to-fine



(a) Coarse Estimate (b) Finest Estimate

Resolution of array of length L $\Delta\theta = 0.886\frac{\lambda}{L}$
[Orfanidis 2002]

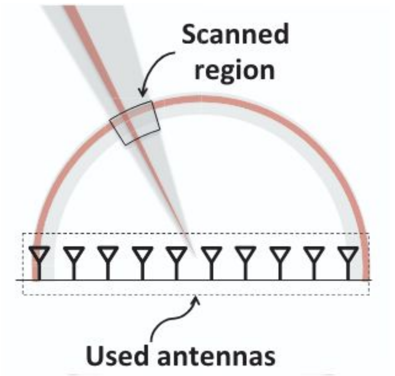


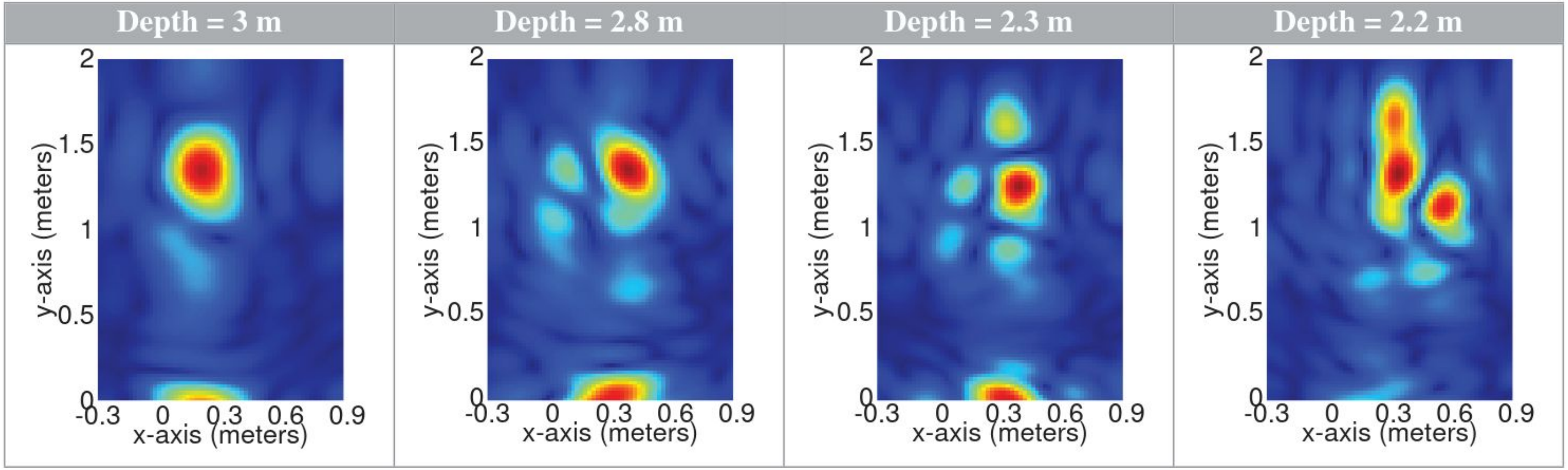
(a) Coarse Estimate (c) Finest Estimate

Depth resolution with a bandwidth B $\Delta r = \frac{c}{2B}$
[Mahafza 2013]

Coarse-to-fine

75ms on Nvidia Quadro K4200 GPU for a 80 ms acquisition



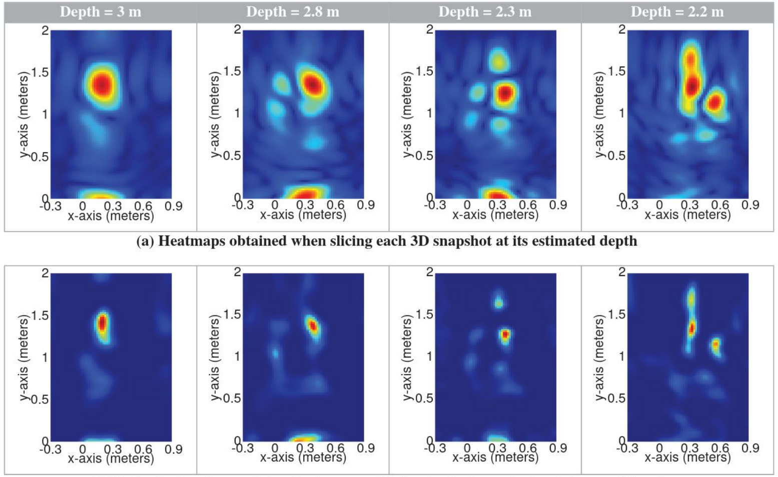


Motion-based Figure Capture

Let’s suppose the user is walking toward the RF-Capture device.

1. Compensation for Depth
   - Deconvolution : Lucy-Richardson algorithm

2. Compensation for Swaying
   - Detection of torso - dominant reflector
   - Alignment



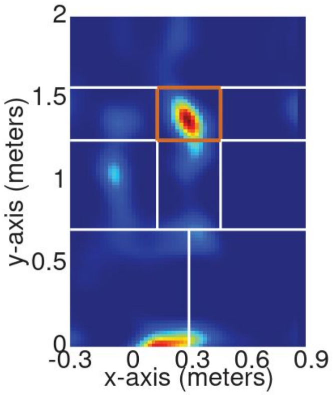
(a) Heatmaps obtained when slicing each 3D snapshot at its estimated depth

(b) Output after deconvolving the images with the depth-dependent point-spread function

Motion-based Figure Capture

3. Body Part Segmentation

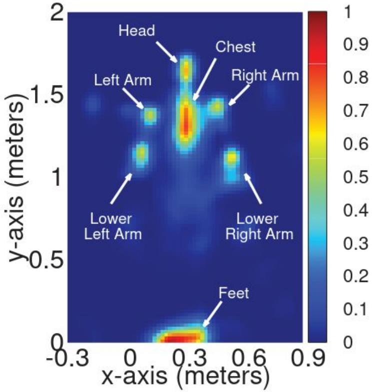
- Empiric measurements of each body part
- Detect blobs



(a) Segmentation

4. Skeletal Stitching

- Rigid parts (head and torso) : summed
- Deformable parts : highest SNR

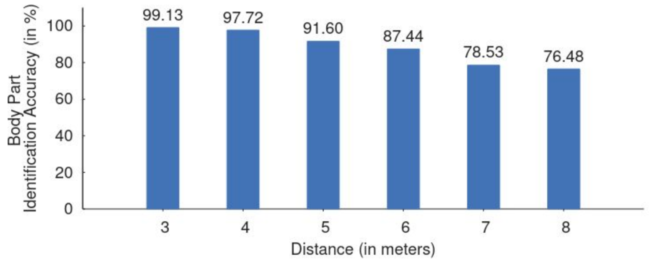


(b) Captured Figure

Results

15 participants

Body Part Identification



Estimated (columns) / **Acutal** (rows)

	Left Hand	Right Hand	Left Leg	Right Leg	Head	Undetected
Left Hand	91.6	0.0	5.6	0.0	2.8	0.0
Right Hand	0.0	90.2	0.0	9.4	0.4	0.0
Left Leg	0.0	0.0	89.7	0.0	0.0	10.3
Right Leg	0.0	0.0	0.0	86.8	0.0	13.2
Head	0.0	0.0	0.0	0.0	90.5	9.5

Body Part Tracking

- Draw a letter in the air

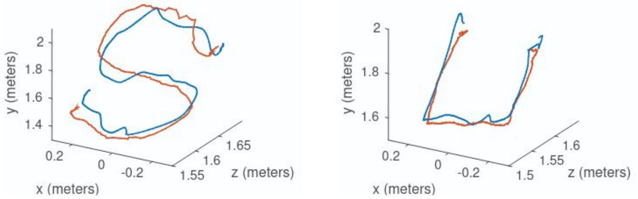


Figure 14—Writing in the Air. The figure shows the output of RF-Capture (in blue) and Kinect (in red) for two sample experiments were the subject wrote the letters "S" and "U" in mid-air.

Results

Human Figure Capture and Identification

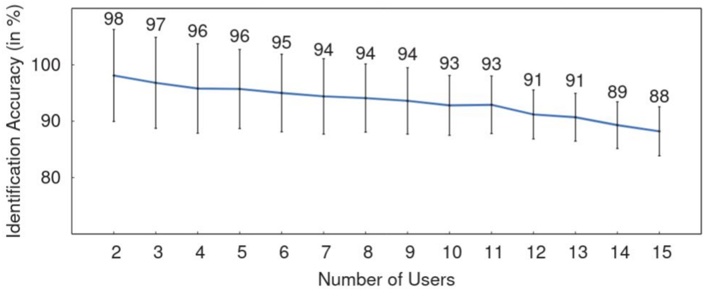


Figure 17—Identification Accuracy as a Function of the Number of Users. RF-Capture uses the captured figures to identify users through classification. The accuracy decreases as the number of users we wish to classify increases.

Acutal \ Estimated	1	2	3	4	5	6	7	8	9	10
1	99.7	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	96.0	0.4	1.3	0.3	0.0	0.9	0.0	0.4	0.7
3	0.0	0.0	99.9	0.0	0.0	0.1	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	2.1	0.0	90.8	0.0	4.8	2.3	0.0	0.0
6	0.0	0.0	5.1	0.0	0.0	94.2	0.0	0.7	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.1	86.9	0.0	13.0	0.0
8	0.0	1.0	0.0	0.0	0.0	0.0	0.8	97.6	0.0	0.6
9	0.0	0.0	0.0	0.0	0.0	0.0	8.8	0.0	91.2	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.8	0.0	98.1

Table 2—Confusion Matrix of Human Identification. The table shows the classification accuracy for each of our subjects.

Limitations

- The system needs movements (and consistent motion!)
- Cannot perform full skeletal tracking
- Simple model for human body