Synthesized Millimeter-Waves for Human Motion Sensing

Xiaotong Zhang\textsuperscript{1,2}, Zhenjiang Li\textsuperscript{1}, Jin Zhang\textsuperscript{2}

\textsuperscript{1}Computer Science, City University of Hong Kong
\textsuperscript{2}Computer Science and Engineering, Southern University of Science and Technology
Sensing Human Motion

• Meaning of this activity?
  Activity Recognition

• How does the body move?
  Skeleton Tracking

• Which solution to use?
  mmWave, Wi-Fi, Wearable...
Deep Learning Model
Scarcity Issue of Training Data

Quality Label Collection
- Expensive devices

Data-Modalities
- Multi-modality data

User-Engagement
- Over engagement

Vicon, OptiTrack, etc.

Vision-based data

mmWaves

“Generate (synthesize) mmWave sensing signals using vision-based datasets, allowing these signals to inherit motion labels directly from the datasets.”

- FMCW Radar
Our Solution – SynMotion

• **Module-1**: signal synthesis
• **Module-2**: sensing signatures
• **Module-3**: sensing services
Module-1: Signal Synthesis

- **Signal synthesis pipeline**

1. **Transmitted FMCW Signal**
2. **Reflected & Received Signal**
3. **Intermediate Frequency (IF)**
4. **Reflections from Body**

\[
S(t) = \sum_{k \in K} \sum_{i=1}^{n} S_{i}^{k}(t), \\
\text{s.t.} \quad \text{Tx and Rx signals are not blocked by body, } \\
\sigma \in \mathcal{H}(\text{ellipsoid}),
\]
Module-1: Signal Synthesis

• Signal synthesis pipeline

1. Transmitted FMCW Signal

2. Reflected & Received Signal

3. Intermediate Frequency (IF)

4. Reflections from Body

• Inheriting labels

Synthesized signals

Radar

Coordinate system of the vision-based dataset

Motion labels, e.g., skeleton coordinates

\[ S(t) = \sum_{k \in K} \sum_{i=1}^{n} S_{IF}(t), \quad \text{s.t.} \quad \text{Tx and Rx signals are not blocked by body, } \sigma \in \mathcal{H}(\text{ellipsoid}). \]
Module-2: Sensing Signatures

- Popular signatures
  - Micro-Doppler spectrum
  - Activity recognition
- Heat map
  - Skeleton tracking

from synthesized signals
from real signals
Module-3: Compensation

**Synthetic-to-Real Problem**
- **Fine-tuning** (with labeled data)

**Training Framework**
- **Variant tracker** & **actual tracker**
- **Three** training steps
Two Trackers

- **Variant Tracker**
  - Attention
  - ResNet-18
  - LSTM-Block

- **Actual Tracker**
  - Attention
  - ResNet-18
  - LSTM-Block

Initial Pose

Heat maps

Offsets

$\ldots x_{t-2}, x_{t-1}, x_t \ldots$
Two Trackers

- **Variant Tracker**
  - Initial Pose
  - Heat maps
  - ResNet-18 → LSTM-Block → Attention

- **Actual Tracker**
  - Initial Pose
  - Heat maps
  - ResNet-18 → LSTM-Block → Attention

- Offsets

- Body shape
- Skeleton movement
Three Training Steps

1. **Step 1**: train both trackers using synthesized signals
2. **Step 2**: fine-tune variant tracker
3. **Step 3**: fine-tune actual tracker

- **Variant Tracker**
  - Attention
  - ResNet-18
  - LSTM-Block

- **Actual Tracker**
  - Attention
  - ResNet-18
  - LSTM-Block

**Good user-independent feature**

- Labels (coordinates of skeleton points)
- Pseudo ground truth (estimated coordinates)

- Real mmWaves

- Heat maps

- Initial Pose

- “Good user-independent feature” chart with error values for Axis X, Axis Y, and Axis Z for not fine-tuned and fine-tuned conditions.
Evaluation

• Devices
  o TI IWR1443BOOST radar
  o OptiTrack

• Datasets
  o Our dataset
    ▪ Group-a): 10 users
    ▪ Group-b): 10 users
  o Public datasets
    ▪ NTU RGB+D
    ▪ CMU MoCap
Evaluation

| Skeleton Points | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | Average |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------|
| Axis X          | 1.9 | 2.0 | 2.0 | 2.2 | 2.2 | 2.5 | 2.6 | 5.3 | 2.2 | 2.5 | 3.5 | 5.3 | 2.0 | 2.6 | 2.6 | 2.1 | 4.3 | 3.4 | 2.9    |
| Axis Y          | 1.4 | 1.4 | 1.6 | 1.5 | 1.8 | 1.5 | 4.1 | 4.5 | 1.6 | 2.3 | 4.7 | 5.9 | 1.4 | 0.6 | 0.6 | 1.4 | 2.1 | 2.5 | 2.5    |
| Axis Z          | 2.7 | 2.7 | 2.9 | 3.2 | 3.5 | 3.1 | 3.1 | 3.7 | 4.8 | 3.2 | 3.3 | 4.4 | 4.2 | 2.6 | 2.7 | 2.7 | 2.6 | 2.9 | 3.2    |
| Overall         | 4.5 | 4.6 | 4.8 | 5.2 | 6.0 | 5.1 | 5.4 | 7.0 | 8.8 | 5.2 | 5.8 | 7.9 | 9.1 | 4.6 | 4.3 | 4.3 | 4.6 | 6.5 | 6.4 | 5.8    |

- **SynMotion**
  - Overall error: 5.8 cm
- **RF-Pose3D [2]**
  - Overall error: 5.3 cm

[2] RF-Based 3D Skeletons, in ACM SIGCOMM, 2018
Evaluation

• Different datasets

• Our dataset
  • Per-axis error: 2.5 – 3.2 cm

• NTU RGB+D
  • Per-axis error: 3.2 – 3.6 cm

• CMU MoCap
  • Per-axis error: 3.2 – 3.8 cm
Conclusion 1, 2, 3

1. One goal:
   - Solve the scarcity issue of training data

2. Two aspects of significance:
   - Bootstrap mmWave sensing at low cost
   - Enhance interpretability of mmWave sensing

3. Three modules:
   - Signal synthesis
   - Sensing signatures
   - Synthetic-to-real training
Thank you
Q&A