ORIGIN™

WiFi can do more.

Proprietary & Confidential
WiFi-Based Robust Human and Non-human Motion Recognition With Deep Learning

Guozhen Zhu, Beibei Wang, Weihang Gao, Yuqian Hu, Chenshu Wu, K. J. Ray Liu
Origin Research, Rockville, MD, USA
{guozhen.zhu, beibei.wang, weihang.gao, yuqian.hu, chenshu.wu, ray.liu}@originwirelessai.com

Guozhen Zhu
March 15th 2024
Content

• Introduction
• Framework Design
  • WiFi Signal Preprocessing
  • Deep Neural Networks
• Evaluation
• Conclusion
WiFi-based human and non-human recognition

In many indoor environments, the moving subjects can be:

- Human motion
- Non-human motion

Non-human motion introduce high false alarm rates to intelligent systems and applications:
- Intruder detection
- Occupancy monitoring
- Smart buildings
- Activity recognitions...
WiFi-based human and non-human recognition

- Pets, vacuum machines, and electrical appliances are essential parts of families and extensively exist in various environments.
  - About 85 million families in US have pets in 2022 (American Pet Products Association).
  - The global robotic cleaners market is $5.59 billion in 2021.
  - Electrical appliances such as fans and washing machines are common in households.

→ Distinguishing human and non-human subjects is crucial for practical indoor intelligent applications and systems.
WiFi-based human and non-human recognition

Previous work using hand craft features

Feature vector
- Physical features:
  - existence of gait, stride length, stride cycle time, average speed, speed variance, speed 25 percentile, speed 75 percentile
- Statistical features:
  - ACF peaks mean, ACF valleys mean, ACF peaks interval distance, ACF valleys interval distance, motion statistic mean, motion statistic variance
WiFi-based human and non-human recognition

Motivation

- Challenging cases:
  - Features cannot be derived from data:
    - Device limitations:
      - Coverage
      - Noise
    - Motion is complex:
      - Intrusion
      - Multiple people
  - Features show overlap:
    - Pet too large

Need more powerful models that can directly extract features from data!
WiFi-based human and non-human recognition

Framework Overview
**System Design**

**Input preparation**

1. Take channel power response of CSI
   
   \[ G(t, f) \triangleq |\bar{H}(t, f)|^2 \]
   
   \[ = |H(t, f)|^2 + 2 \text{Re}\{n^*(t, f)H(t, f) \exp(-j(\alpha(t) + \beta(t)f))\} + |n(t, f)|^2 \]

2. ACF of channel power response
   
   \[ \rho_C(\tau, f) = \frac{\text{cov}[G(t, f), G(t+\tau, f)]}{\text{cov}[G(t, f), G(t, f)]} \]
   
   \[ = \frac{\sigma_x^2(f)}{\sigma^2_x(f) + \sigma^2_n(f)\rho_x(\tau, f) + \frac{\sigma^2_n(\tau)}{\sigma^2_x(\tau) + \sigma^2_n(f)}} + \delta(\tau) \]

3. Apply Maximum Ratio Combine (MRC)
   
   \[ \hat{\rho}_x(\tau) = \sum_{i=1}^{N_x} \rho_C(\tau = \frac{1}{F_n}, f_i)\rho_C(\tau, f_i) \]

4. Take the differential
   
   \[ \Delta\hat{\rho}_x(\tau) \]
WiFi-based human and non-human recognition

Motion Detection

The autocorrelation function (ACF) of the CSI $H(t, f)$ with the time lag $\tau$:

$$
\rho_C(\tau, f) = \frac{\text{cov}[G(t, f), G(t + \tau, f)]}{\text{cov}[G(t, f), G(t, f)]}
= \frac{\sigma_s^2(f)}{\sigma_s^2(f) + \sigma_n^2(f)} \rho_s(\tau, f) + \frac{\sigma_n^2(f)}{\sigma_s^2(f) + \sigma_n^2(f)} \delta(\tau)
$$

motion statistic

$\phi_G(f) \triangleq \rho_G \left( \tau = \frac{1}{F_s}, f \right)$

WiFi-based human and non-human recognition

Deep Learning Models

- Categories:
  - Convolutional neural network
    - LeNet
    - ResNet (-18, -50, -101)
  - Recurrent neural network
    - RNN
    - GRUNet
    - LSTM

Pro: extract spatial hierarchies with 2D kernel
Con: focus on local

Pro: good at extract temporal dependencies
Con: neglect patterns of A-ACF within a time stamp
WiFi-based human and non-human recognition

Input Preparation
WiFi-based human and non-human recognition

Evaluation Environments

Floor plan of (a) Scenario I, an apartment, (b) Scenario II, a townhouse, and (c) Scenario III, a single family house.

Evaluation Metrics

Recall (True Positive Rate) = \( \frac{True\ Positive}{Total\ P ossitive} \)

Precision (Positive Predictive Values) = \( \frac{True\ P ossitive}{True\ P ossitive + False\ P ossitive} \)

False Alarm Rate (False Positive Rate) = \( \frac{False\ P ossitive}{Total\ N egative} \)

Accuracy = \( \frac{True\ P ossitive + True\ N egative}{Total} \)
Evaluation

Dataset

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Human</th>
<th>Pet</th>
<th>Cleaning robot</th>
<th>Fan</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>F1, M1, M2, M3, M4, M5, M6, M7, M8, F2, F3</td>
<td>D1, D2, D3, D4, D5, D6, D7, D8</td>
<td>iRobot V3</td>
<td>Rotation</td>
</tr>
<tr>
<td>II</td>
<td>F1</td>
<td></td>
<td>iRobot V3</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>M2, M3, M9</td>
<td>C1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Human:**
  - **12 individuals** (9 males, 3 females)
  - Age: 23 to 34 years old
  - Height: 154 cm to 194 cm
  - Participants are free to walk, run, sneak, or stop and have small motions during data collection such as using their phones while walking.

- **Pets:**
  - **9 different pets**, including 8 dogs and a cat
  - Weight: **17 lb to 85 lb**
  - Pets are allowed to move freely.

- **Cleaning robot**
  - iRobot V3 vacuum machine

- **Fan:**
  - A rotation fan
WiFi-based human and non-human recognition

Evaluation – Classification Performance

CLASSIFICATION ACCURACY OF DEEP LEARNING MODELS FOR HUMAN AND NON-HUMAN MOTION IDENTIFICATION

<table>
<thead>
<tr>
<th>Model</th>
<th>Exp. I</th>
<th>Exp. II</th>
<th>Exp. III</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>LeNet</td>
<td>91.29%</td>
<td>88.69%</td>
<td>98.19%</td>
<td>92.72%</td>
</tr>
<tr>
<td>ResNet18</td>
<td>94.60%</td>
<td>92.78%</td>
<td>99.34%</td>
<td>95.57%</td>
</tr>
<tr>
<td>ResNet50</td>
<td>93.28%</td>
<td>92.46%</td>
<td>98.68%</td>
<td>94.81%</td>
</tr>
<tr>
<td>ResNet101</td>
<td>93.66%</td>
<td>90.52%</td>
<td>98.59%</td>
<td>94.25%</td>
</tr>
<tr>
<td>GRUNet</td>
<td>75.66%</td>
<td>72.20%</td>
<td>97.65%</td>
<td>81.84%</td>
</tr>
<tr>
<td>RNN</td>
<td>75.38%</td>
<td>79.31%</td>
<td>97.86%</td>
<td>84.18%</td>
</tr>
<tr>
<td>LSTM</td>
<td>86.46%</td>
<td>82.76%</td>
<td>96.71%</td>
<td>88.64%</td>
</tr>
</tbody>
</table>

- LeNet achieves an average accuracy of 92.72%.
- The ResNet family of models demonstrated superior performance, especially ResNet18, which achieved the highest average accuracy of 95.57%.
- Other deep learning models like RNN, GRUNet, and LSTM performed comparatively lower, with average accuracies of 84.18%, 81.84%, and 88.64%, respectively.

EVALUATION RESULTS OF FOUR-CLASS MOTION RECOGNITION

<table>
<thead>
<tr>
<th>Model</th>
<th>Average Precision</th>
<th>Average Recall</th>
<th>Average F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>LeNet</td>
<td>92.00%</td>
<td>91.96%</td>
<td>91.92%</td>
</tr>
<tr>
<td>ResNet18</td>
<td>94.23%</td>
<td>93.52%</td>
<td>93.70%</td>
</tr>
<tr>
<td>ResNet50</td>
<td>91.42%</td>
<td>91.71%</td>
<td>91.31%</td>
</tr>
<tr>
<td>ResNet101</td>
<td>91.82%</td>
<td>90.74%</td>
<td>90.83%</td>
</tr>
<tr>
<td>GRUNet</td>
<td>68.67%</td>
<td>81.66%</td>
<td>73.90%</td>
</tr>
<tr>
<td>RNN</td>
<td>73.90%</td>
<td>81.86%</td>
<td>77.12%</td>
</tr>
<tr>
<td>LSTM</td>
<td>87.73%</td>
<td>87.85%</td>
<td>87.50%</td>
</tr>
</tbody>
</table>

- Overall, the results demonstrate the superiority of our proposed deep learning-based framework in accurately identifying human and non-human subjects through-the-wall with WiFi.

The remarkable performance of the ResNet models affirms the efficacy of the deep learning approach in complex detection tasks.
WiFi-based human and non-human recognition

Evaluation – Robustness

- Followed the leave one-environment-out methodology and thus get three testing experiments.
- Each experiment contains a balanced dataset that can reflect the real-world conditions, where the data size ratio of human to non-human targets is 5:1.

- ResNet101 achieved an impressive average testing accuracy of 87.09%.
- All models surpassed the 80% mark in average testing accuracy.
- These outcomes validate that our proposed architecture minimizes the impact of factors such as environment, position, or direction on recognition performance.

This enables effective recognition of human and various non-human subjects in new environments without requiring additional training or parameter adjustments.
WiFi-based human and non-human recognition

Evaluation – Model Parameter

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>14.7</td>
<td>42.7</td>
<td>81.3</td>
<td>158.0</td>
<td>0.082</td>
<td>0.24</td>
<td>0.055</td>
</tr>
<tr>
<td>Size (MB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For implementation on edge devices, we need to balance the trade-off between memory requirement and recognition performance.
WiFi-based human and non-human recognition

Conclusion

• We introduce a framework for recognizing human and various non-human subjects using a single pair of WiFi devices with deep learning.
• Our system addresses the common challenge of non-human interference by leveraging a unique and context-agnostic feature A-ACF as input for deep neural networks, which ensures robust performance regardless of environment, location, or direction.
• Leverage a deep neural network to extract features from a designed statistic, A-ACF, Rigorous tests across diverse scenarios not only affirm the high accuracy of the proposed system on four types moving subjects recognition in complex settings but also guide the choice of the most suitable deep learning architectures for WiFi sensing tasks.