EarBuds Gesture Recognition with TinyML Using Novel Short-Range Radar Sensors

Michele Magno, Andrea Ronco
D-ITET Center for Project-based Learning and Embedded Systems, ETH Zurich

Gesture Recognition Based on Short-Range Radar

This work proposes a low-power high-accuracy embedded hand-gesture recognition targeting battery-operated wearable devices using low power short-range radar sensors. A 2D Convolutional Neural Network (CNN) using range-frequency Doppler features is combined with a Temporal Convolutional Neural Network (TCN) for time sequence prediction. The final algorithm has a model size of only 19855 parameters, yielding a memory footprint of only 39.5kB. A dataset with 11 challenging gestures plus an idle class has been recorded, containing 50 samples per gesture acquired by 19 people, and a single-user dataset of 700 samples per gesture from a single person. On these datasets an accuracy of 92% (20 users) has been achieved. The network has been integrated in an embedded system based on ARM Cortex-M4 with 270mm², suitable for wearable devices. The system can perform real-time gesture recognition with an active power consumption of ~17mA and of ~400uA in idle, which opens the door to extended activity in battery operated, small wearable devices.

Increasing research on radar for gesture recognition

Google developed micro-radar for gesture recognition

Good results on difficult hand-gestures: 92% accuracy on 11 gestures and 20 people

Short Range Radar

- Radar technology: FMCW
- RX signal mixed with TX to obtain Intermediate Frequency signal. Distance of the target encoded in the frequency of IF signal.
- Aviatiopn Radar: 1-12GHz (λ=2.5-30cm)
- Infineon’s Radar: 60GHz (λ=5mm)
- Sensor data: samples of IF signal over one frame
- Frame: 32 chips spaced 700µs, 16 samples per chip
- 13 FPS (frames per second)

Range-Doppler Maps

- Strong feature based on research1,2
- Efficient computation based on CMSIS-DSP
  1. FFT on the sample axis (Fast Time) → Resolve Range
  2. FFT on the chirps axis (Slow Time) → Resolve velocity

Technical Background

The concept is to combine strong features of 2D-Convolutional Neural Networks (CNN) and Temporal Convolutional Neural Networks (TCN) for real-time gesture recognition in embedded systems.

- 1. 2D Convolutional Neural Network (CNN)
- 2. Temporal Convolutional Neural Network (TCN)
- Dense Classifier:
- Sequence of stacked representation vectors of length 32, leverage temporal information for more accurate predictions.

Results

Achieved an accuracy 92%, only two very similar gestures are confused often (pinch index and pinch pinky)

Conclusion

This work presented a high-accuracy and low-power hand-gesture recognition system based on short-range radar. Two large datasets with 11 challenging hand-gestures performed by 20 different people containing a total of 20210 gesture instances are recorded, on which the final algorithm reaches an up to 92%. The model size is below 40kB and the hardware implementation shows that the application is feasible in a Cortex-M4 based, battery operated wearable device. The Earbud can be used also as vital sign monitoring.

Implementing Radar-Based Hand Gesture Recognition in Embedded Systems

- Created a dataset with challenging fine-grained gestures (1250 samples per class, 20 users)
- Real-Time algorithm that runs on well-know hardware (ARM Cortex-M4)
- Accuracy similar to what seen in Interacting With Soli (92% with 20 users)
- Network at least 700x smaller than I.w.Soli (40kB)

Contribution of This Work

- Strong feature based on research1,2
- Efficient computation based on CMSIS-DSP
- Sensor data: samples of IF signal over one frame
- Frame: 32 chips spaced 700µs, 16 samples per chip
- 13 FPS (frames per second)