mmFER: Millimetre-wave Radar based Facial Expression Recognition for Multimedia IoT Applications

Xi Zhang†,1,2, Yu Zhang†,1, Zhenguoh Shi1, Tao Gu1

1 Macquarie University, Sydney, Australia
2 RMIT University, Melbourne, Australia
† The first two authors contributed equally to this work
Facial Expression Recognition (FER)

- **Emotional awareness by FER** for interaction (CHI), communication (feedback), and well-being (healthcare)

- Deliver a valuable assessment of audience’s preference, interest level, engagement and reactions, etc.

- Enabling a fundamental capability that IoT system can “better understand” users, actively create more personalized and responsive user experiences
State-of-the-Arts

Vision-based approach:
- Privacy concerns
- Ambient light conditions (e.g., in the dark)
- Blocking (e.g., wearing masks)

Wearable based approach (PPG, EEG, earphones):
- Discomfort to users for long-time wearing
- One device for each user

Wireless sensing approach (Ultrasound, Wi-Fi):
- Fail by impact of body motions
- Short detection range (e.g., <= 60cm)
- Poor support for multiple users
Our Solution: mmWave Radar Sensing

- **High robustness**: robust to work in different environment conditions, e.g., dark
- **Large bandwidth**: high resolution for detecting objects and tiny motions
- **Long-range detection**
- **Fine spatial resolution**: fine spatial resolution enabled based on the MIMO
- **Wide Field of View (FOV)**: cover a large area with a single sensor
- **Penetration**: can easily penetrate materials such as glasses, masks
- **Privacy-preserving manner**
Challenges

- Default point cloud approach **fails** to detect user’s face due to **highly sparse point clouds** generated.

- **Subtle facial movements**: facial muscle movements by expressions are in **millimetre levels**.

- **Massive ambient noise** contains in raw mmWave signals, *e.g.*, body motion, walking people, appliance, and ambient noise reflected by walls.

- **Limited mmWave dataset**: facial data collection is **costly** due to labelling efforts and **privacy concerns**.
Key Ideas

- What if we could “partially locate” user’s face for capturing subtle facial movements from noisy raw mmWave signals?
  
  - **Step-1**: using unique biometric features to locate users and eliminate ambient noise
  - **Step-2**: using spatial facial features to locate faces and remove irrelevant body motions

- What if we could use a public image FER dataset (i.e., large-scale) and its pre-trained models to “transfer” knowledge from image domain to mmWave domain to effectively enable the learning with much less data collection?
  
  - Using cross-domain transfer learning to enable optimal model performance with small-scale mmWave dataset
Our Contributions

• A first-of-its-kind mmWave radar based FER system that detects subtle facial muscle movements associated with raw mmWave signals for multimedia IoT applications

• A novel dual-locating approach to accurately locate on subjects’ faces in space based on MIMO technology

• A novel cross-domain transfer pipeline to enable an effective and safe model knowledge transformation for mmWave-based FER model learning

• An off-the-shelf mmWave radar based implementation with extensive experiments

• This pioneering system mitigates concerns over privacy concerns and lighting constraints, and has strong adaptability to fit a number of real-world scenarios with high accuracy
System Working Flow

Raw data collection

- Reflected mmWave Signals
- Spatial Facial Information
- Dual-locating Approach
  - Face Localization
  - Biometric Verification
  - Candidate Range Localization
  - Dynamic Object Removal

Cross-domain learning

- Cross-transfer Pipeline
  - mmWave Model
  - Image Model

Face localization

- Face Image Input
- Facial Expressions
  - 😊  😊
  - 😒 😒
Main Technique 1: Dual-locating Approach

Step-1: eliminating ambient noise

Noise removal pipeline (3-process)

Step-2: removing body motions

Face-matching mechanism using Gaussian Mixture Model (GMM)

Heart rate (left) and respiration (right) verification

Located facial mmWave raw data
mmFER Demo

Normal speed: facial muscle movement within seconds

Slow-motion
Main Technique 2: Cross-transfer Pipeline

- Inspired by the principle of cross-domain transfer learning, uniquely using a pre-trained FER image model to “teach” training our mmWave model.

- Proposing an autoencoder based feature alignment mechanism to reduce the impact of data heterogeneity of image to mmWave data.

- Proposing a hybrid learning loss function:
  1) A supervised loss;
  2) A Kullback–Leibler (KL) divergence loss;
  3) A contrastive loss based on positive-negative correlation, largely improve model performance.
System Implementation and Setup

• **Setup**: TI IWR1843BOOST sensor board operating at 77-81GHz ($299) and a TI DCA1000EVM data capture board ($599)

• **Upright RX antenna array** in elevation for face localization

• **Data collection**: recruiting 10 subjects

• **Use scenario**: tested in different scenarios with different noise setup, e.g., body motions, postures, subject-to-radar distance, face orientation, wearable accessories
Evaluation: Dual-locating Performance

- Fig. (a) shows that our approach can effectively enable face localization at different subject-to-radar distances with minor error drift.

- Fig. (b) shows that our approach can locate face accurately for multiple targets by removing ambient noise.
Evaluation: Cross-transfer Performance

- Comparing to 3 baselines:
  - Knowledge distillation (KD)
  - Facial landmark based image-to-sensing transformation (Keypoint)
  - Unsupervised cross learning approach (S-cross)

- Our approach outperforms baselines, achieving highest accuracy
Evaluation: Overall Performance

- **84.48% accuracy** in a subject-to-radar distance between 0.3 and 1.5m
- **80.57% accuracy** when distance increases to 2.5m
- No major accuracy drop in a scenario with **wearing accessories**
Thank you!
Questions