A Thesis for the Degree of Ph.D. in Engineering

Person Identification with Deep Neural Networks: A Comparative Study of CW Radar, LiDAR, and IR Array Sensors

September 2025

Graduate School of Science and Technology
Keio University

Kai Liu

別表６
（３）

Keio University

Thesis Abstract

No.

Registration Number	√ “KOU” □ “OTSU” No. *Office use only	Name	Kai Liu

Thesis Title

Person Identification with Deep Neural Networks: A Comparative Study of CW Radar, LiDAR, and IR Array Sensors

Thesis Summary

Person identification is a foundational task in biometrics, enabling systems to recognize individuals based on their physiological or behavioral traits. It plays a critical role in applications such as surveillance, healthcare monitoring, secure access control, and human–computer interaction. Conventional identification systems typically rely on RGB facial images or physical credentials, which perform well under constrained conditions. However, in real-world scenarios involving low lighting, occlusion, disguise, motion blur, or low-resolution inputs, these systems often fail to deliver reliable performance.

To overcome these limitations, there is a growing need for non-contact, robust identification systems that can operate across a wide range of environments. This dissertation addresses this challenge by investigating alternative sensing modalities beyond conventional RGB inputs. Specifically, it explores the use of radar-extracted physiological signals, infrared (IR) array sensors with LiDAR, and thermal facial imaging, each combined with deep neural network architectures. The overarching goal is to evaluate and improve the robustness, interpretability, and generalization of person identification systems under practical, unconstrained conditions.

The first part of this study explores heartbeat-based person identification using continuous-wave (CW) Doppler radar. Two signal representations are compared: direct time-series classification and spectrogram-based image classification. Models such as InceptionTime, ResNet, and a custom CNN are applied to evaluate performance under realistic conditions. To assess robustness, the radar signals are corrupted with various noise types—including Gaussian, impulse, speckle, and salt-and-pepper noise. To mitigate the effects of noise, four time-series denoising techniques (Savitzky–Golay filter, wavelet denoising, median filtering, and Butterworth low-pass filtering) and three image-based methods (autoencoder, Deep Image Prior (DIP), and Dilated-Residual U-Net (DRUNet)) are implemented. Results show that spectrogram-based models can achieve up to 100% accuracy in noise-free settings, and that DIP and wavelet-based methods significantly improve performance under high-noise conditions. These findings demonstrate the feasibility of radar-based physiological identification and highlight the importance of denoising strategies for real-world deployment.

The second part focuses on spatial imaging modalities, using IR array sensors and LiDAR to collect

RGB, thermal, and depth images across resolutions from 16×12 to 640×480. Vision Transformer (ViT) and ResNet34 are evaluated across these modalities. While RGB images perform best at high resolution, thermal imaging shows consistent performance even at low resolution due to stable model attention. Depth data, while useful for shape, suffers from scattered focus at lower resolutions.

In the final part, we investigated facial image-based identification, emphasizing robustness under disguise, occlusion, and domain shifts. Experiments on the Pontificia Universidad Católica de Valparaíso-Drunk Thermal Face database (PUCV-DTF), IRIS (Imaging, Robotics, and Intelligent Systems) dataset, and a custom-collected multi-modal dataset show that ViT achieves high accuracy in clean conditions but struggles with generalization. ResNet-based models offer improved robustness under low resolution and occlusion. Cross-subject and cross-dataset tests reveal limited generalizability, while integrated gradients highlight modality-specific attention behaviors.