People Identification and Secondary Action Recognition Methods in Vehicles Using Depth Sensors

Feburary 2025

Weirong Shao

別表６
（３）

Keio University

Thesis Abstract

No.

Registration Number	☑ "KOU"　☐ "OTSU" No.　*Office use only	Name	Shao Weirong

Thesis Title

People Identification and Secondary Action Recognition Methods in Vehicles Using Depth Sensors

Thesis Summary

Our study focuses on exploring methods for identifying people and recognizing secondary actions using depth sensors in vehicles. With the increasing implementation of advanced driver-assistance systems (ADAS) and autonomous vehicles, accurately detecting and interpreting human presence and behavior inside and around the vehicle is crucial for ensuring safety and enhancing user experience. Depth sensors, which capture 3D spatial data, offer significant advantages in tracking movement and recognizing activities and receive little impact from changes in the external environment. This research aims to develop effective methods for people identification and secondary action recognition based on depth sensors in vehicles, contributing to safer and smarter vehicle environments.

In Chapter 1, we introduce the background of depth sensor applications, as well as people identification and action recognition technologies to date.

In Chapter 2, we address the limitations of traditional RGB-based people identification in vehicles by employing 3D Light Detection and Ranging (LiDAR) technology. 3D LiDAR uses infrared lasers to generate a detailed spatial representation without revealing personal details. We focus on enhancing identification accuracy by overcoming occlusion challenges in vehicle interiors, particularly for backseat passengers. Our approach includes converting LiDAR data into depth images using Dynamic Non-linear Mapping (DNLM) and employing generative image inpainting to reconstruct occluded areas. We validate the effectiveness of the YOLOv5 model in recognizing individuals from processed 3D LiDAR images, demonstrating improved accuracy and privacy protection.

In Chapter 3, we address the limitations of traditional action reco2gnition methods that rely on RGB data by utilizing a Time-of-Flight (ToF) depth sensor. ToF technology captures depth information based on light travel time, maintaining high accuracy across diverse environments, including low-light and high-glare situations. We introduce a novel deep-learning model for SA recognition. Our model leverages Convolutional Neural Networks (CNN) for feature extraction, enhanced by a Spatial Enhance Attention Mechanism (SEAM) that focuses on critical frame areas to enhance spatial feature analysis for accurate action recognition. A Bidirectional Long Short-Term Memory (Bi-LSTM) network is integrated to capture bidirectional temporal dependencies, further refining action interpretation. Evaluations on the public benchmark Drive&Act dataset demonstrate that our model outperforms existing state-of-the-art models for SA recognition. Ablation experiments confirm the superiority of depth video over conventional RGB video for SA recognition under varying lighting conditions. Our contributions showcase the potential of integrating CNN, SEAM, and Bi-LSTM to advance real-world action recognition applications through enhanced spatial and temporal modeling.

In Chapter 4, we improve the time sequence component with Feed-Forward Attention (FFA) based on the deep learning framework proposed in Chapter 3, by incorporating a specialized time-sequence module that enables the model to extract more relevant information within the time-allocated window for secondary actions. This enhancement has led to a notable improvement in action recognition accuracy. Additionally, we introduced advanced data augmentation and super-resolution techniques to diversify our dataset, enhancing the model's generalization and adaptability across varied scenarios

In Chapter 5, we conclude this dissertation and discuss future work.