Robust Driver Head Pose Estimation with Temporal Modeling

Tiancheng Hu, Carlos Busso

Multimodal Signal Processing Lab (MSP)

Erik Jonsson School of Engineering & Computer Science
University of Texas at Dallas, Richardson, Texas - 75080, USA

Motivation

Background:

- Head pose is important for modelling the driver's visual attention
- Challenges:
  - Occlusions
  - Extreme head poses
  - Dynamic changes in illumination

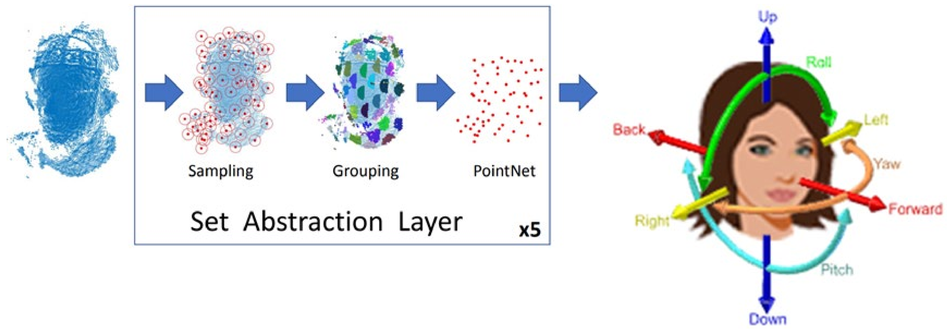
Our Work:

- Build a head pose estimation model from point cloud data
- Take into account the temporal information from continuous data

Prior Work

Multimodal Driver Monitoring (MDM) Dataset:

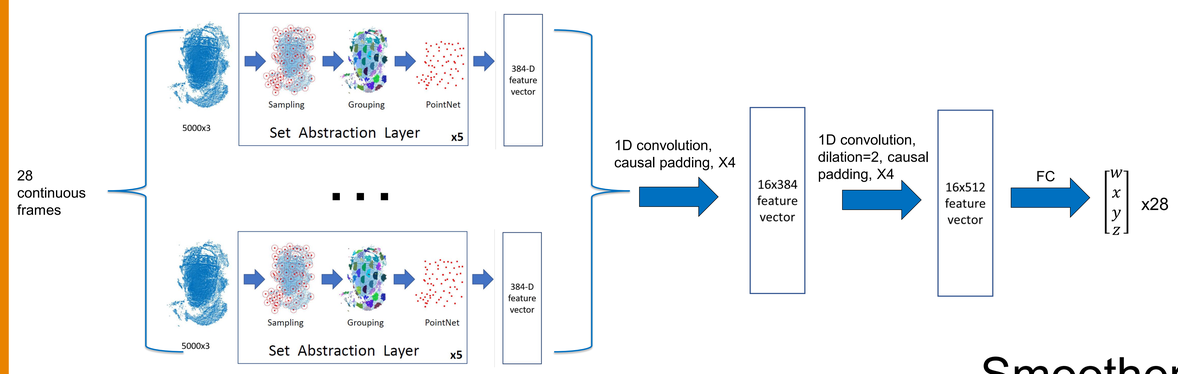
- Multimodal dataset with 4 RGB GoPro cameras, 1 Pico Flexx time-of-flight depth camera, microphone array, and controller area network (CAN) bus
- Diverse scenes including naturalistic driving conditions
- 52 subjects with balanced gender ratio
- Head pose label provided by Fi-Cap [1]



Driver Head Pose Estimation from Point-Cloud Data [2]

- Treat head pose estimation as a regression problem
- Point-cloud based model with sampling, grouping and PointNet layers
- Extract features at different scales, which are aggregated
- Multiple sets of layers to extract high-level features
- Better performance than Openface 2.0, which is a state-of-the-art automatic head pose detection algorithm [3]

Temporal Point Cloud Deep Learning Model



Model Details

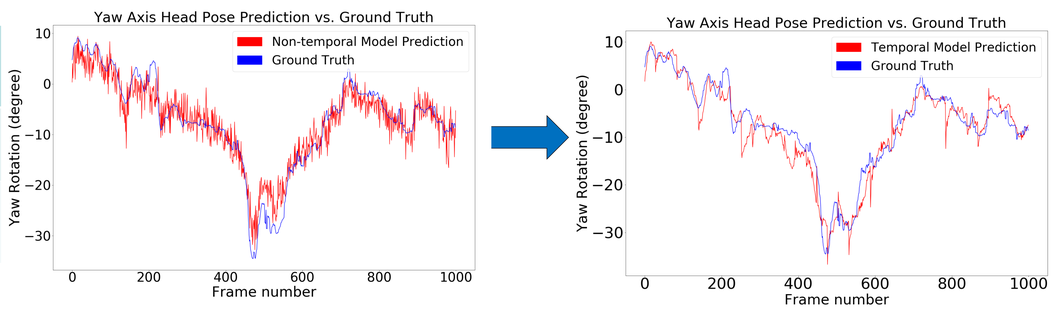
- Loss function: MSE

$$\sum_{i=1}^{n}\left(Y_{true} - Y_{predicted}\right)^2$$

- Input: 3D point cloud of driving data
- Output: 4D vector (Quaternions)
- Train: 32 subject
- validation 10 subjects, test 10 subject

Smoother Prediction

Root Mean Squared Error	Roll(°)	Yaw(°)	Pitch(°)
Non-temporal Model	10.55	11.87	7.65
Proposed Temporal Model	11.07	9.23	7.61



Conclusions

- Build a deep learning model utilizing temporal information in driver head pose estimation
- More smoothness and consistency in predictions compared to non-temporal model
- **Future Work**
  - Continue to collect more training data
  - Evaluate more effective ways to utilize temporal information
  - Build fusion model that combines information from RGB data and point cloud data
- **References:**

[1] S. Jha and C. Busso, "FI-CAP: Robust Framework to Benchmark Head Pose Estimation in Challenging Environments," *2018 IEEE International Conference on Multimedia and Expo (ICME)*, San Diego, CA, 2018, pp. 1-6.

[2] T. Hu, S. Jha and C. Busso, "Robust Driver Head Pose Estimation in Naturalistic Conditions from Point-Cloud Data," *2020 Intelligent Vehicles Symposium (IV)*, Las Vegas, NV, 2020, to appear.

[3] T. Baltrusaitis, A. Zadeh, Y. C. Lim and L. Morency, "OpenFace 2.0: Facial Behavior Analysis Toolkit," *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, 2018, pp. 59-66.

This work was supported by Semiconductor Research Corporation (SRC) / Texas Analog Center of Excellence (TxACE), under task 2810.014