Multimodal Understanding

Contextual Understanding in Intelligent vehicles

- Understanding Driver's and Passengers' actions and intentions serves a wide range of purposes starting from aiding the driver for safe navigation to more interactive in-car entertainment systems.
- A combination of various modalities is being investigated for robust Automatic Speech Recognition (ASR) and to attain contextual understanding within vehicles.
- Different modalities are important for understanding the surrounding environment (inside and outside) for intelligent vehicles:
  - Visual information
  - Audio
  - GPS
  - Inertial Measurement Unit
  - CAN bus data
  - Driver head gaze
  - Speaker information
  - Audio from stereo etc.

Sensors used:
- Softkinetic depth cameras
- Microphone arrays
- GPS+IMU
- Front facing camera
- CAN Bus

Platform:
- ROS based framework
- Long term data collection capabilities
- Easy to add new sensors

Performance and On-going Work

Performance

- An accuracy of 81% was obtained in differentiating speech and silence using the VAD component.
- Different acoustic models are being evaluated for in-car speech recognition and a Word Error Rate of 59% was obtained in very noisy environments.
- All the computations are done on board in the car on a Jetson Tegra K1 GPU boards i.e., no data needs to be streamed. This avoids any privacy issues to continuously listen to conversations in cars.

Challenges and On-going work

- Acoustic scene is dependant on many factors in a car like the position of the windows, type of road, air-conditioner status, car stereo etc.
- Continuously listen to the conversations in the car and interrupt at the right time.
- Differentiating between conversation between individuals and commands to the system.
- More efforts are being focused on fusing CAN bus information with the data from audio and visual sensors to improve the performance of the speech recognition and contextual awareness.

This work is done in collaboration with the Integrated Socially Intelligent Sidekick (ISIS) project at Electronics Research Laboratory, Volkswagen of America.