Probabilistic Estimation of the Gaze Region of the Driver using Dense Classification

Sumit Jha* and Carlos Busso
Drivers’ Visual Attention

- Primary driving related task
  - Mirror checking actions [Li and Busso, 2016]
  - Lane change
  - Turns and cross sections

- Secondary tasks
  - Mobile phones and in-vehicle entertainment unit
  - Co-passengers in the car
  - Billboards and other distractions from the environment

Gaze detection a challenging problem in car environment

Often approximated by head pose

Related Work

- Studying eyes-off-the-road [Liang and Lee, 2010]
- Predicting discrete gaze zones from the head pose [Vora et al., 2017]
- Relating driving actions to head pose
  - Mirror checking actions [Li and Busso, 2016]
  - Lane change [Doshi and Trivedi, 2012]

Motivations

- **Head pose – Gaze relation not deterministic [Jha and Busso, 2016]**
  - The variability depends on the location of gaze

- **Probabilistic prediction of driver’s visual attention from head pose**
  - Region of gaze provides important information about visual attention

Previous Work

- Predicting probability based gaze region based on the head pose on the head pose of the driver
  - Example model using GPR [Jha_2017]
    \[
    Y = h(\vec{x})^T \beta + f(\vec{x})
    \]
    Deterministic component
    Probabilistic component
    
    \[
    f(\vec{x}) \sim GP(0, K(\vec{x}, \vec{x}'))
    \]

- Aim to design a more flexible model
  - Non-parametric estimation of probability
  - Adaptable model with more control over the parameters

Regression as Classification

- **Non-parametric probability estimation using softmax**
  - Softmax learns a probability distribution giving confidence value for each label
  - Better way of learning probability than GPR [VandenOord_2016]

- **Solving regression as a classification problem**
  - Class labels need to be ordered (error 1 < error 2)

- **Implicit multitask learning with multidimensional features**
  - Classification in the grid of 2 variables
  - Problem becomes dense with $N^2$ classes for high resolution

---

- **Input 6 degrees of head pose**
  - Head position \((x,y,z)\)
  - Orientation \((\alpha,\beta,\gamma)\)

- **Fully connected layer followed by CNN**
  - Learn gaze representation in 4 x 2 discretized level
- Upsample followed by CNN
  - Learn the gaze representation at 8x4 discretization
- Repeat to get incrementally higher resolution
  - Train at each resolution
- Softmax activation at the output layers to obtain probability maps that sum to 1
Data for the study

- Camera-1 → Face
- Camera-2 → Road
- Markers on the windshield
- Use Apriltags for tracking head movement

**Ask subjects to look at each point multiple times at random**
Ground truth gaze data during Naturalistic Driving

- Collected when the subject is driving the car
- Subject asked to look at points
- Data collected in a straight road with minimum maneuvering task
- Data collected with 16 subjects (10 males 6 females)
AprilTags for Head Pose Estimation

- Head pose estimation challenging in driving environment
- Avoid the error in head pose estimation to affect the performance of the model
- AprilTags [Olson, 2011]
  - 2D barcodes that can be robustly detected in an image
- Headband designed with 17 AprilTags
- Useful for robust detection of head pose across conditions

Implementation of the Proposed Model

- Keras on top of tensorflow to learn the model
- Final output is obtained at 256 x 128 (7 stages)
- Entire network trained at each stage
- Learning rate lowered at later stages with more number of epochs
  - $10^{-2}$ for first 5 stages, 200 epochs
  - $10^{-3}$ for last 2 stages, 500 epochs
- Driver independent partition
  - 14 subjects for training
  - 1 subject for validation
  - 1 subject for test
Results of Experimental Evaluation

- **Accuracy versus resolution**
  - Area represented as a portion of the hemisphere in front of the driver
  - Study the performance at different stages
  - As we increase resolution the precision increases
Prediction of visual attention

50% confidence

95% confidence
Prediction of visual attention
Comparison with GPR

- Performance of basic architecture slightly worse compared to GPR
- Possible improvements
  - Deeper architecture in each upsampling
  - Cost sensitive loss functions
  - Continuous and more exhaustive gaze data (as opposed to limited discrete points in the space)
Conclusions and future work

- Deep learning framework to learn the probability distribution of gaze from head pose
- Incrementally learn higher resolution
- Incorporate information from the eye to increase accuracy
Thank You