Unsupervised Error Bounds for Multimodal Perception

Denver Dash (Intel), Tom Mitchell (CMU), Takeo Kanade (CMU), Martial Hebert (CMU)

The "Deeply Embedded" Classification Problem

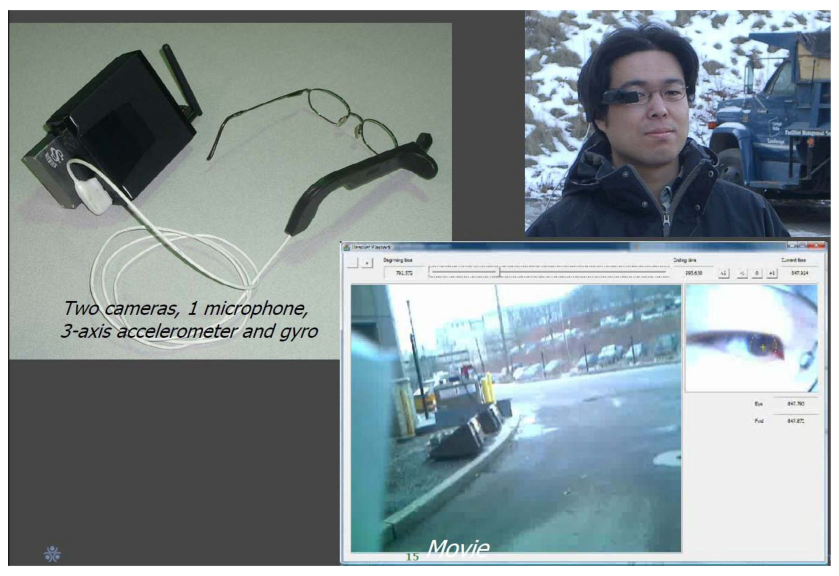
Synopsis: As we become engulfed by smart embedded devices which work together to sense and infer their environment, we require a better understanding of how to quantify the classification error made when combining a set of conditionally independent classifiers, each of which has an unknown error rate.

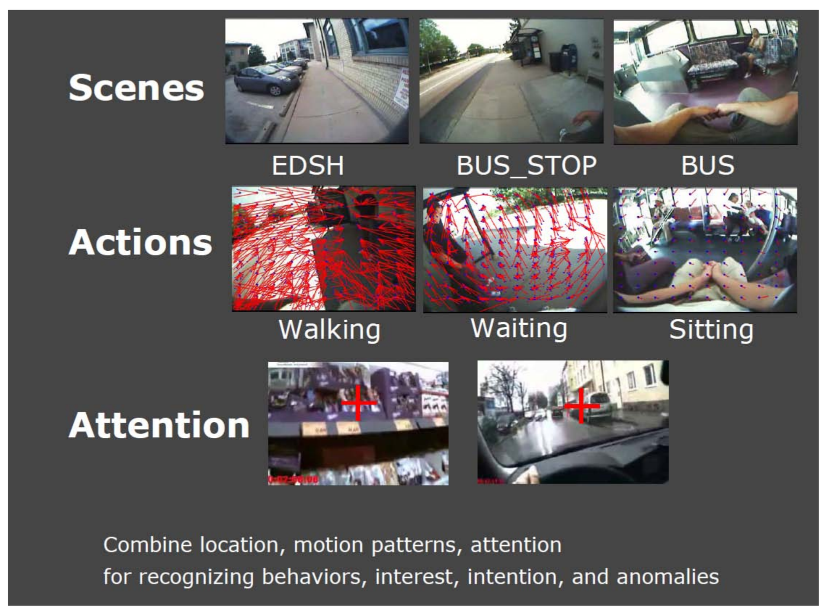
Why quantify the error?
- Allows us to combine multiple devices in a principled way
- Allows us to calculate expected utility to make useful actions.

Why assume conditional independence?
- Each device is predicting the same state of the world, so they are marginally dependent.
- Each device typically operates on a distinct data stream, so conditionally independence seems reasonable.

Why assume an unknown error rate?
- Actual environment may differ from testing environment
- Some devices may be broken or gaming the system.

Application: First-Person Sensing



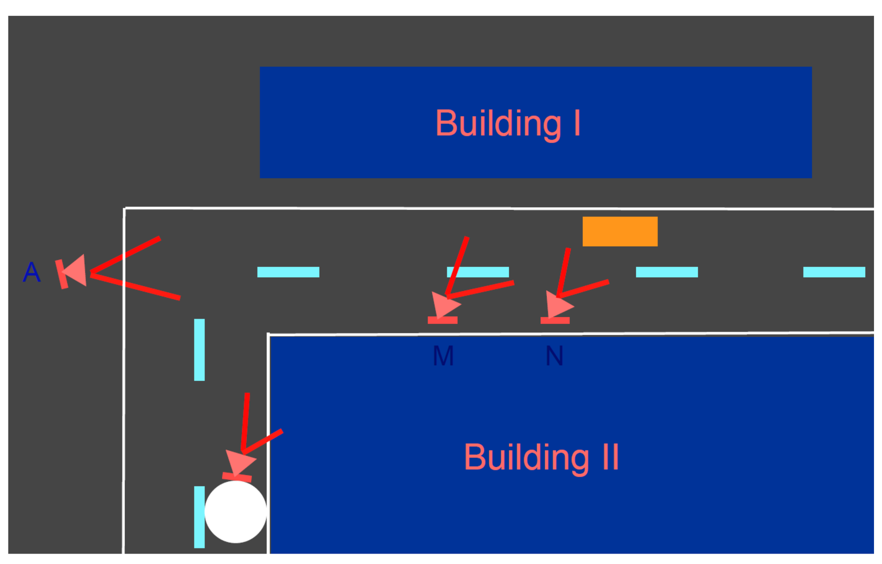


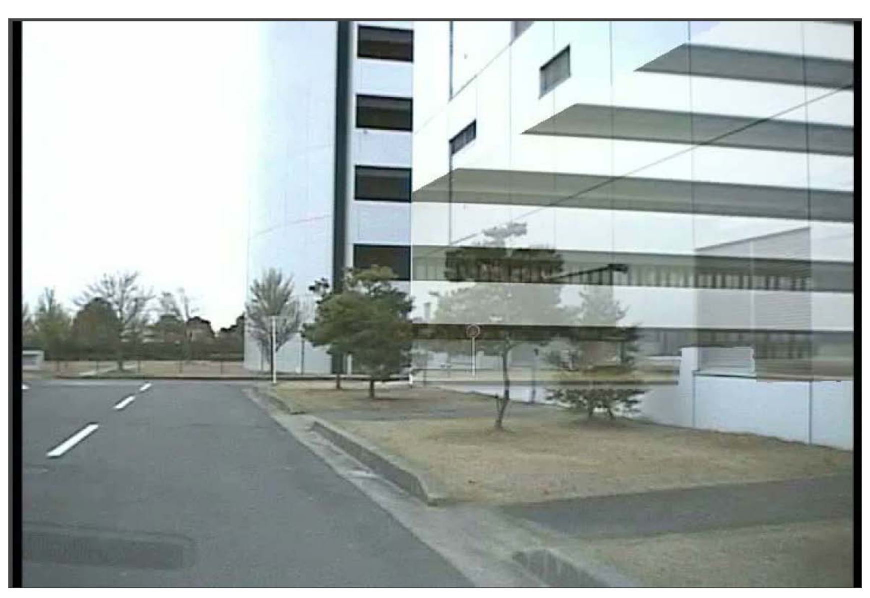
Interesting problems combining FP vision, speech + acceleration

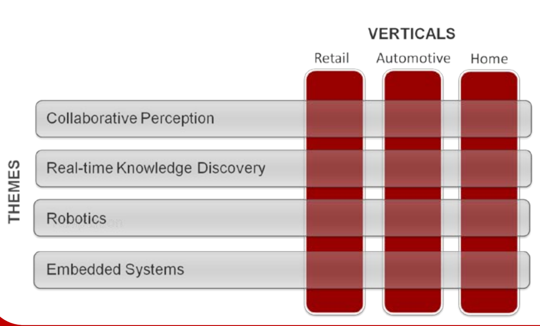
- Person identification
- Emotion detection
- Localization
- Activity Recognition
- Natural Language Command & Control
- Universal Translation
- Personal item tracking

Collaborative First-Person Sensing

- X-ray Vision
- Total Situation Awareness
- Traffic Detection

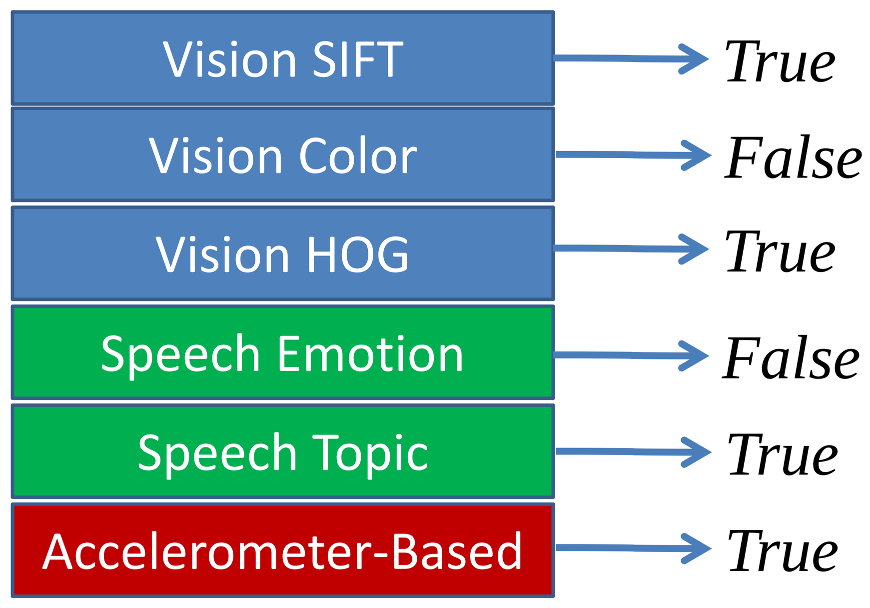






Unsupervised Error Bounds

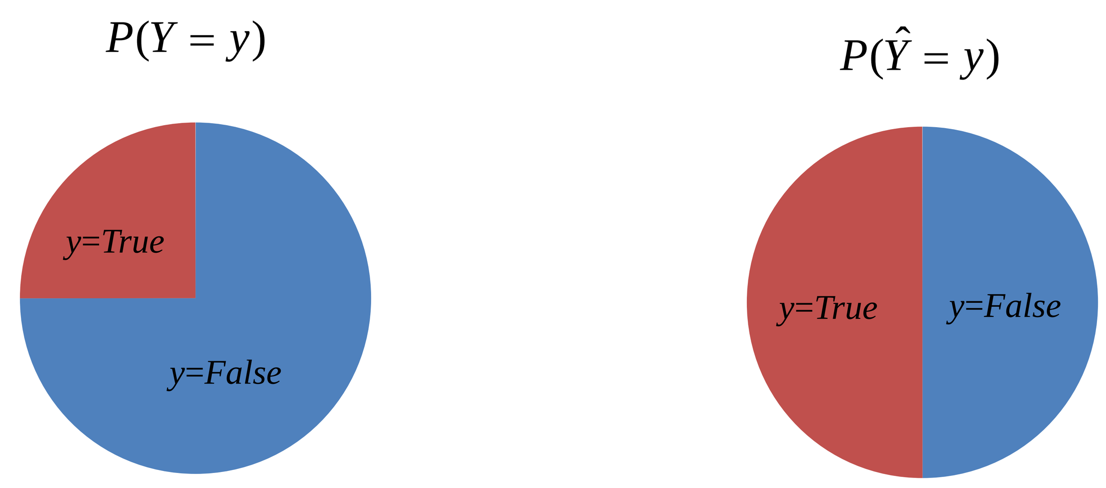
Given N black-box classifiers with unknown accuracy, estimate the error of a joint classification.



Assuming conditional independence:

$$P(\hat{Y} = True \mid \{\hat{Y}\}_i) = \prod_i P(\hat{Y}_i = True \mid Y = True) P(Y = True)$$

One Idea: Use marginal probability of class state (Donmez, et al., 2010):



$$\underbrace{P(\hat{Y} = y)}_{\text{known}} = \sum_i \underbrace{P(\hat{Y} = y \mid Y = y_i)}_{\text{constrained}} \underbrace{P(Y = y_i)}_{\text{known}}$$

Another Idea: Use classifier agreement rate

E.g., consider 2 conditionally independent binary classifiers

Assume: $P(\hat{Y}_1 \neq \hat{Y}_2) \leq \delta$ $\qquad 0 \leq \delta \leq 1$

Let $\theta_{iy} = P(\hat{y}_i = y \mid Y = y)$ and $\theta_{i\bar{y}} = P(\hat{y}_i = \bar{y} \mid Y = \bar{y})$

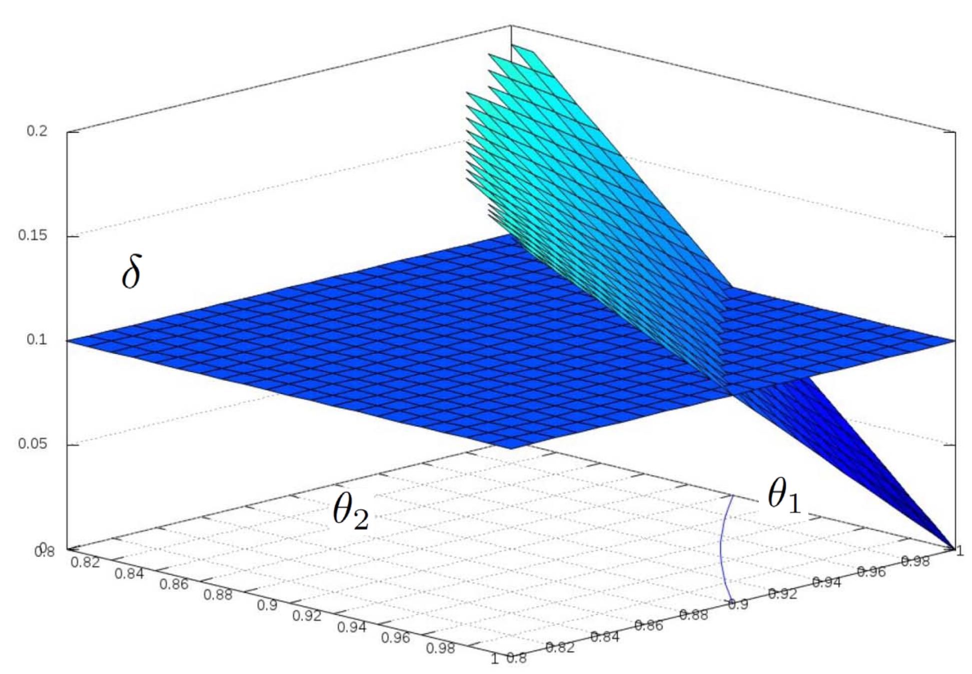
$$\begin{aligned} & P(Y = y) \cdot \{\theta_{1y}(1 - \theta_{2y}) + (1 - \theta_{1y})\theta_{2y}\} + \\ & P(Y = \bar{y}) \cdot \{(1 - \theta_{1\bar{y}})\theta_{2\bar{y}} + \theta_{1\bar{y}}(1 - \theta_{2\bar{y}})\} \leq \delta \end{aligned}$$

If known $\Rightarrow$ constrained

Assuming k classifiers with *balanced error*: $\theta_{iy} = \theta_{i\bar{y}} \equiv \theta_i$

One can derive the loose bound: $\theta_i \geq 1 - \dfrac{\delta}{k - 1}$

The loose bound can be tightened further by integrating over the constrained area with uniform priors:



The region where the θ_1-θ_2 surface lies below the $\delta = 0.1$ plane.

Carnegie Mellon University

intel