Multiple Component Learning (MCL)

- **Derivation**
  - No assumptions about input space
  - Only needs weak classifiers for arbitrary $g(x) = \langle h, x \rangle$ (e.g., $h \in \mathbb{R}^d$)
  - For example, can use $\mathcal{G}$ (sets) in place of $\mathcal{H}$
  - Can therefore use MIL to train weak classifier
  - MIL learns a function $h^* = \arg\min_h \mathbb{E}(\text{Error}(\mathcal{H}, h, \mathcal{D}))$
  - Only need to adjust MIL to take weights

- **Algorithm**
  - Training a feature $1$. Large amount of data needed 2. Evaluating all components slow (currently working on improving speed)
  - Standard $F_i(x) = \arg\min_h \mathbb{E}(\text{Error}(\mathcal{H}, h, \mathcal{D}))$
  - Set $\alpha_i = \frac{1}{m} \log(1 - \text{Error}(F_i))$
  - Set $A_{ij}(t) = \alpha_t \log(1 - \text{Error}(F_i))$
  - Output the final classifier $\mathcal{F}(X) = \arg\min_{\mathcal{F}} \sum_i \alpha_i F_i(X)$

- **Other Applications**
  - Speaker Identification
    - Text independent handwriting identification
    - 2 people, 2 pages of text each
    - Hair features on 5000 patches
  - Writer Identification
    - Content independent speaker identification
    - 2 people, 2 pages of text each
    - Standard MCL features

- **Set Learning**
  - Standard MIL, MCL

- **Goals**
  1. Learn part-based classifier with weak supervision
  2. Part models are classifiers from rich hypothesis class (rather than Gaussian distributions, templates, etc.)
  3. No complex inference since model is discriminative

- **Overview**
  - **(1) Learning a single part**
    - Weakly supervised learning
    - Object location in positive images unknown
    - Developed for learning objects, use for parts
    - We use Multiple Instance Learning (MIL)
  - **(2) Learning diverse parts**
    - What prevents learning same part repeatedly?
    - Different weighting of data
    - Not all parts expressed in all images
  - **(3) Combining part detectors**
    - Boosting offers way of combining multiple diverse classifiers
    - Train one weak (part) classifier using MIL
    - No weight samples according to current error
    - Repeat until training error sufficiently low

- **Pedestrian Detection**
  - Low-level features
    - Features for component classifiers
    - Haars over multiple channels: gray (1), grad (1), quantized grad (6)
  - Incorporating Spatial Relations
    - Features for overall classifier
    - Densely compute component responses $C_i$
    - Final classifier retrained with Haars over $C_i$

- **Multiple Component Learning (MCL)**
  - Boosting
  - Input: $N$ training examples $x_i$, with $y_i \in \{-1, 1\}$ and $x_i = (x_{i1}, \ldots, x_{id})$, and initial data, of weights $w_i(1) = 1$ over the examples.
  - Iteration $t = 1, \ldots, T$:
    - Train a set classifier $F_t: \mathcal{X} \rightarrow \{-1, 1\}$ using distribution $D_t$. Let $P_t(x_i) = \mathbb{E} (\mathcal{F}(x_i)) = (2 \times \mathbb{E}(\mathcal{F}(x_i)), 0 - 1)$, where $D_t$ is the distribution over the examples.
    - Calculate error of $F_t$: $\text{Error}(F_t) = \frac{1}{N} \sum_i \mathbb{1}(y_i \neq \mathcal{F}(x_i))$
    - Set $\alpha_t = \frac{1}{2} \log(1 - \text{Error}(F_t))$
    - Set $A_{ij}(t) = \alpha_t \log(1 - \text{Error}(F_t))$
    - Output the final classifier $\mathcal{F}(x) = \arg\min_{\mathcal{F}} \sum_i \alpha_i F_i(x)$

- **Results on INRIA Data**
  - We achieve state of the art results.

- **Summary**
  - Advantages:
    1. General notion of parts (components)
    2. Component learning weakly supervised
    3. Principled, general algorithm
    4. State of the art results with simple features
  - Disadvantages:
    1. Large amount of data needed
    2. Evaluating all components slow (currently working on improving speed)