Robust Face Detection with Multi-Class Boosting

Yen-Yu Lin and Tyng-Luh Liu

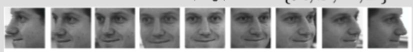
Institute of Information Science, Academia Sinica, Taipei, Taiwan

1. Summary

- We aim to establish a robust system to detect faces under various poses/circumstances, e.g., lighting, rotation, and occlusion.
- Our approach
  - Different from using direct sharing of weak learners for the multi-class classification, we instead consider sharing a good projection so that each class of data has its own decision boundary.
  - MBHboost (multi-class Bhattacharyya boost) derives the weak learners by iteratively minimizing the weighted error upper bound of all classes.
  - Our system needs only one cascade to perform the multi-class detection.

2. Face Detection using MBHboost

- Notations:
  - Set of face classes to be detected, e.g., $\Gamma = \{\mathcal{A}, \mathcal{B}, ..., \mathcal{I}\}$



$\mathcal{A}$ $\mathcal{B}$ $\mathcal{C}$ $\mathcal{D}$ $\mathcal{E}$ $\mathcal{F}$ $\mathcal{G}$ $\mathcal{H}$ $\mathcal{I}$

  - $\forall \mathcal{X} \in \Gamma$, the type-$\mathcal{X}$ training data is

$$D^{\mathcal{X}} = \{(x_1^{\mathcal{X}}, y_1^{\mathcal{X}}), ..., (x_{|D^{\mathcal{X}}|}^{\mathcal{X}}, y_{|D^{\mathcal{X}}|}^{\mathcal{X}})\} = D^{\mathcal{X}+} \cup D^{\mathcal{X}-}$$

  - Each projection ϕ in the projection set Φ is uniquely defined by a rectangle feature.

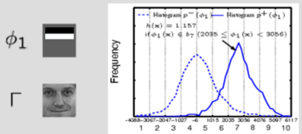
2.1 MBH Weak Learners

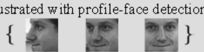
- **Single-class detection** (illustrated with frontal-face detection)
  - For $\phi \in \Phi$, the projected data $\phi(D^{\mathcal{E}})$ form two weighted histograms $p^{\mathcal{E}+}(\phi)$ and $p^{\mathcal{E}-}(\phi)$ over a bounded real line with m bins, $\{b_k\}_{k=1}^m$
  - In Lin et al. [ECCV04], the optimal projection ϕ is chosen to be the one with the minimal Bhattacharyya coefficient.
  - By minimizing the error upper bound, i.e., $\prod Z_t$, the weak learner associated with ϕ is given by

$$h^{\mathcal{E}}(\mathbf{x}) = \ln\sqrt{p_k^{\mathcal{E}+}(\phi)/p_k^{\mathcal{E}-}(\phi)}, \quad \text{if } \phi(\mathbf{x}) \in b_k.$$

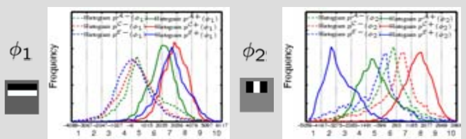
ϕ_1

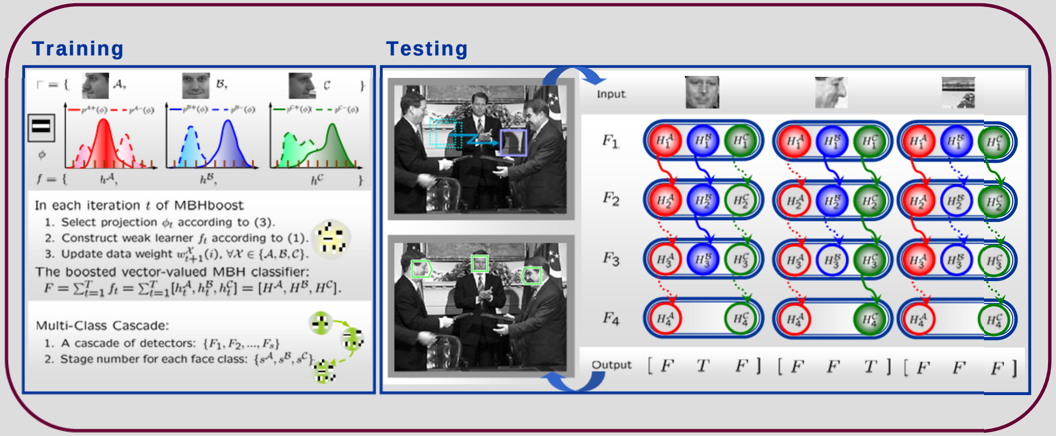
Γ



- **Multiple-classes detection** (illustrated with profile-face detection)
  - For simplicity, consider $\Gamma = \{$  $\}$ $\mathcal{A}$ $\mathcal{C}$ $\mathcal{E}$
  - What could be of concern by sharing weak learners among classes?
    1. Need to decide a subset of classes that shares a weak learner.
       - Can be pre-defined by human knowledge, Li et al. [ECCV02] and Lin et al. [ECCV04].
       - Or decided by searching, Torralba et al. [CVPR04].
    2. Need to decide the shared decision boundary.
       - Direct sharing? The histogram pairs may be far differently distributed!

ϕ_1 ϕ_2





- **The vector-valued MBH weak learners**
  - An MBH weak learner f associated with ϕ is defined by

(1)

(2)

- **Advantages:**
  - It can be shared by all classes, no human knowledge or searching.
  - Each component independently learns a decision boundary.
  - Computational efficiency: The value of k is identical in all components.

2.2 An Optimal Criterion

- *MBHboost is guaranteed to iteratively minimize the weighted error bound for multi-class classification, if the weak learner at iteration t is selected as follows.*

$$\phi_t = \arg\min_{\phi \in \Phi} \sum_{\mathcal{X} \in \Gamma} \Delta_t^{\mathcal{X}} \times \mathrm{BHC}_t^{\mathcal{X}}(\phi), \quad (3)$$

where $\Delta_t^{\mathcal{X}} = \sum_{i=1}^{|D^{\mathcal{X}}|} \exp\left(-y_i^{\mathcal{X}} H_{1:t-1}^{\mathcal{X}}(x_i^{\mathcal{X}})\right)$,

and $\mathrm{BHC}_t^{\mathcal{X}}(\phi) = \sum_{k=1}^m \sqrt{p_k^{\mathcal{X}+}(\phi) p_k^{\mathcal{X}-}(\phi)}$.

With some effort, the criterion can be illustrated as:

$$\sum_{\mathcal{X} \in \Gamma} \Delta_t^{\mathcal{X}} \times \mathrm{BHC}_t^{\mathcal{X}}(\phi) = \frac{1}{2} \sum_{\mathcal{X} \in \Gamma} |D^{\mathcal{X}}| \prod_{\tau=1}^t Z_\tau^{\mathcal{X}}, \quad (4)$$

$|D^{\mathcal{X}}|$: the weight of class $\mathcal{X}$,

$\prod_{\tau=1}^t Z_\tau^{\mathcal{X}}$: the error bound for classifying $D^{\mathcal{X}}$.

Algorithm : *MBHBoost*

Input : Face classes, Γ; $D^{\Gamma} = \bigcup_{\mathcal{X} \in \Gamma} D^{\mathcal{X}}$; Projection set, Φ; Number of iterations, T.
Output : A vector-valued MBH classifier F.

Initialize: the weight vector $w_1^{\mathcal{X}}(i) = 1/|D^{\mathcal{X}}|$, *for* $i = 1, 2, \ldots, |D^{\mathcal{X}}|$ *and* $\forall \mathcal{X} \in \Gamma$.

for $t \leftarrow 1, 2, \ldots, T$ **do**
1. Determine the optimal projection ϕ_t from Φ by solving (3).
2. Construct the MBH weak learner f_t associated with ϕ_t using (1) and (2).
3. $w_{t+1}^{\mathcal{X}}(i) \leftarrow w_t^{\mathcal{X}}(i) \exp\left(-y_i^{\mathcal{X}} h_t^{\mathcal{X}}(\mathbf{x}_i^{\mathcal{X}})\right) / Z_t^{\mathcal{X}}$, for $i = 1, 2, \ldots, |D^{\mathcal{X}}|$, and $\forall \mathcal{X} \in \Gamma$.
($Z_t^{\mathcal{X}}$ is a normalization factor such that $w_{t+1}^{\mathcal{X}}$ is a distribution.)

Output an MBH classifier F: $F(\mathbf{x}) = \sum_{t=1}^T f_t(\mathbf{x}) = [H^{\mathcal{X}}(\mathbf{x}) = \sum_{t=1}^T h_t^{\mathcal{X}}(\mathbf{x}) \mid \mathcal{X} \in \Gamma]$.

3. Detection Architecture

- Note that the task of multi-class detection is performed over a single boosted cascade.
- The detector at each stage is trained by the proposed MBHboost.

3.1 Training Phase

- To design a multi-class boosted cascade, $\{F_1, F_2, ..., F_s\}$, the key issues in training, say, the detector F_k at stage k are listed below.
  - The class-specific threshold $\theta_k^{\mathcal{X}}$ of $H_k^{\mathcal{X}}$ is set for ensuring 99.5% ~ 99.9% detection rate.
  - The number of MBH weak learners is increased till each $H_k^{\mathcal{X}}$ achieves a false positive rate below 40%.
  - The stage number $s^{\mathcal{X}}$ for class $\mathcal{X}$ is increased till no sufficient type-$\mathcal{X}$ negative data can be generated. At then we exclude class $\mathcal{X}$ from the remaining training of the multi-class cascade.
  - The training procedure is completed when there are no more classes of data left for training.

3.2 Testing Phase

- Testing a pattern $\mathbf{x}$ with the learned cascade would yield a vector of Boolean responses, of which each element indicates whether the input $\mathbf{x}$ belongs to one particular face class or not.

Algorithm : *Multi-Class Cascade: Testing*

Input : A test pattern $\mathbf{x}$; Face classes Γ; A cascade of detectors $\{F_1, ..., F_s\}$;
Number of stages, $s^{\mathcal{X}}, \forall \mathcal{X} \in \Gamma$.
Output : A vector of boolean outputs, $output(\Gamma)$.

Initialize: $k \leftarrow 1$; $\Lambda \leftarrow \Gamma$;

while $\Lambda \neq \emptyset$ **do**
- Jointly evaluate $H_k^{\mathcal{X}}(\mathbf{x}), \forall \mathcal{X} \in \Lambda$;
- **foreach** $\mathcal{X} \in \Lambda$ **do**
  - **if** $H_k^{\mathcal{X}}(\mathbf{x}) < \theta_k^{\mathcal{X}}$ **then**
    - $output(\mathcal{X}) \leftarrow$ `False`; $\Lambda \leftarrow \Lambda - \{\mathcal{X}\}$;
  - **else if** $k = s^{\mathcal{X}}$ **then**
    - $output(\mathcal{X}) \leftarrow$ `True`; $\Lambda \leftarrow \Lambda - \{\mathcal{X}\}$;
- $k \leftarrow k + 1$;

4. Other Scenarios on Face Detection

- Faces with in-plane rotations:

- Faces under various lighting conditions:

- Faces with partial occlusions:

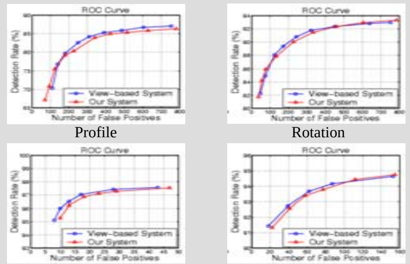
- Faces with various facial expressions:

5. Experimental Results

- Face images are collected from several databases, e.g., MIT-CBCL, AR, PIE, Yale, and IIS. Each class consists of 10,000 properly selected faces, and 10,000 nontrivial non-face samples.

5.1 Accuracy and Efficiency

- Compare our system with a view-based one, i.e., derived by separately training $|\Gamma|$ boosted cascades.
- **Accuracy:**



Profile Rotation Lighting Occlusion

- **Efficiency:**

Application	Profile	Rotation	Lighting	Occlusion
Class #, $\|\Gamma\|$	9	12	5	9
Speedup Factor	3.74	4.96	2.96	3.85