Object Detection Using Nested Cascades of Boosted Classifiers:
A learning framework and its extension to the multi-class case

Rodrigo Verschae

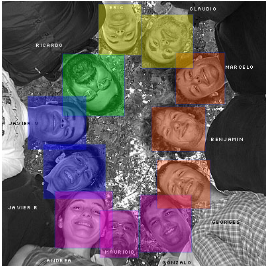
Advisor: Prof. Javier Ruiz-del-Solar

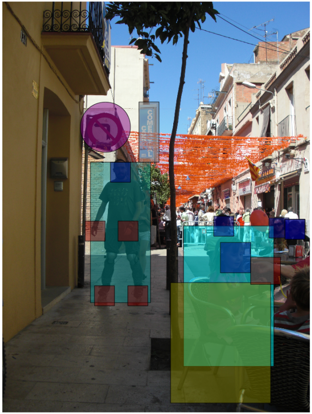
Department of Electrical Engineering, Universidad de Chile

March 31, 2010

Object detection

Goal: For a given set of object classes, to find, if any, the instances of these classes appearing on the image.

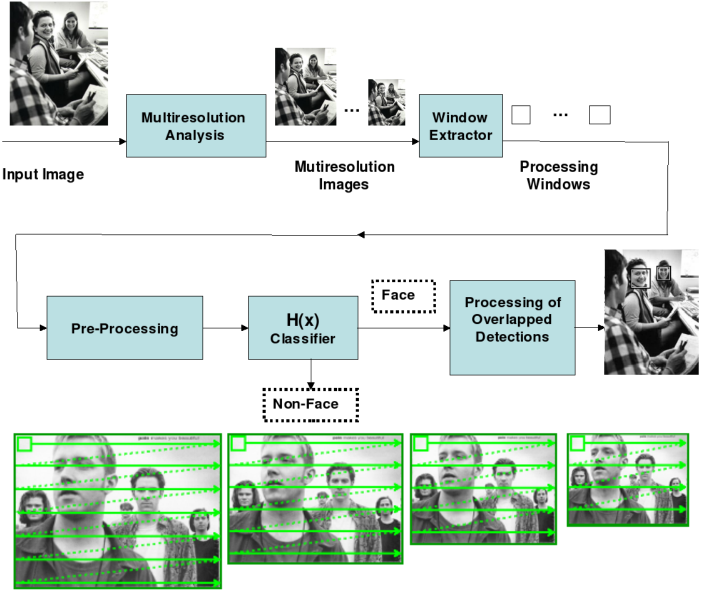




Outline

1. Introduction
   - Detection architecture
   - Cascade classifiers, Adaboost and features
   - Nested cascades
2. Training of boosted nested cascades
3. CTF multiclass nested cascades
4. TCAS: Tree of nested CAScades
5. Summary and conclusions

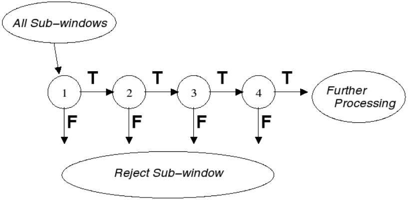
Detection Architecture



Cascade classifiers [Viola and Jones 2001]

High asymmetry

- Most windows correspond to the background
- Most non-object windows are very different from object windows
  $\longrightarrow$ average processing time depends mainly on non-object windows



Coarse-to-Fine classifier

- Reduce the (negative) hypothesis space incrementatlly
- Efficient way of organizing object detection
- First layers must be very fast

Adaboost (two-class problem)

- Training set $S = \{(x_i, y_i)\}_{i=1,\ldots,m}, y_i \in \{-1, 1\}$
- Build $H(x) : \mathbb{R}^n \to \mathbb{R}$ such that $Prob(sign(H(x)) \neq y)$ is small

Faces Non-Faces

[Images from: Ce Liu & Hueng-Yeung Shum, 2003]

Adaboost: Main ideas [Freund & Shapire]

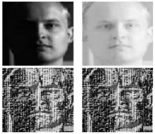
- Build additive model: $H(x) = \sum_{t=0}^{T} \hat{h}_t(x)$
- Minimize $E_{x \sim S}[\exp(-yH(x)]$ iteratively by adding terms to sum
- Drive focus to wrongly classified examples
- Each weak classifier $\hat{h}_t(x)$ does not need to be very accurate

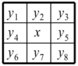
Each feature is associate to a weak classifier

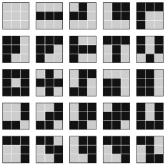
$$H(x) = \sum_{t=0}^{T} \hat{h}_t(x) = \sum_{t=0}^{T} h_t(f_t(x)) \quad (1)$$

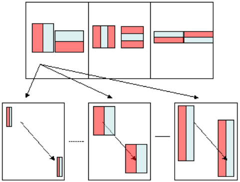
Used feature sets

- Modified Local Binary Patterns (mLBP) [Froba et al 2004]
  - Well suited for frontal faces & Invariant to linear constrast changes
- Rectangular features (Haar-like wavelets) [Viola and Jones 2001]
  - Well suited for frontal faces & Fast evaluation using the Integral Image



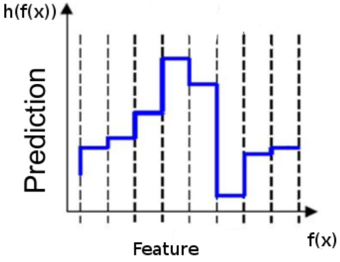






Domain partitioning weak classifiers [Shapire and Singer 1999]

- Evaluation takes $O(1)$ thanks to the use of look up tables
- Can make use non scalar features (e.g. mLBP)

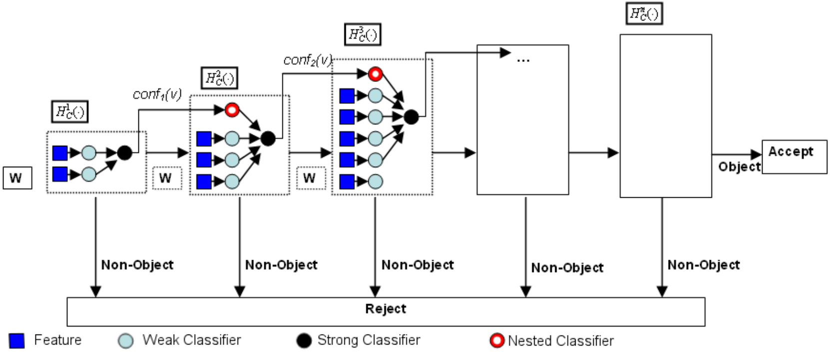


$$H(x) = \sum_{t=0}^{T} h_t(f_t(x)) \tag{2}$$

Nested Cascade [Wu 2004]

- Allows to reuse information
- Faster and more robust than the non-nested case

$$H^k(x) = \sum_{t=0}^{T_k} h_t^k(f_t(x)) + H^{k-1}(x) - b_k, k > 1 \quad (3)$$



Main problems

Training and design issues

- Structure of the classifier
  $\longrightarrow$ complexity (speed) and accuracy of each layer
- Selection of the training set
  $\longrightarrow$ which training examples to use at each layer
- Features and efficient training
  $\longrightarrow$ how to train the system in a fast way
  $\longrightarrow$ features to be used at each layer

Outline

Training Framework

$$H^k(x) = \sum_{t=0}^{T_k} h_t^k(f_t(x)) + H^{k-1}(x) - b_k, k > 1 \quad (4)$$

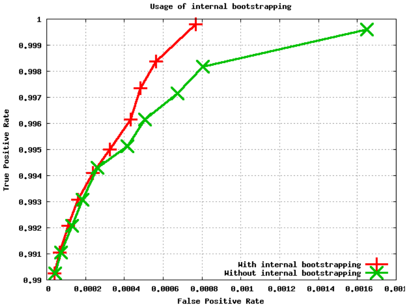
Main Contributions

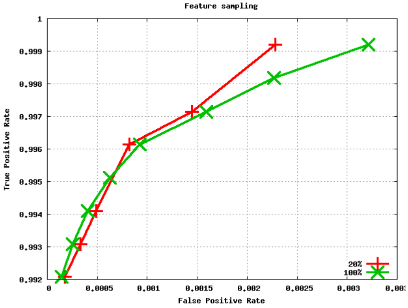
- Criterion to decide when stop the training of a layer
  $\rightarrow$ trade-off: processing speed, false positives and true positives
- Selection of negative examples: Internal and external bootstrapping
- Heuristics to speed up the training (feature sampling, feature types)

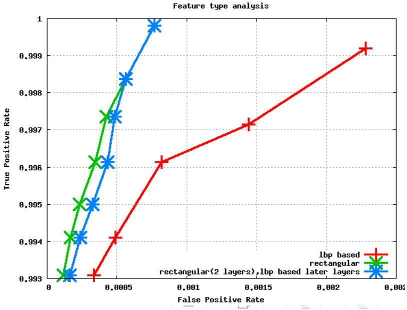
Results

- The structure of the cascade does not have to be predefined
- Short running time (~50 ms, images of 320x240, Pentium 4 GHz)
- Short training time (~15 hours vs weeks in [Viola & Jones] – Pentium 4 GHz, 10000 training samples per layer)

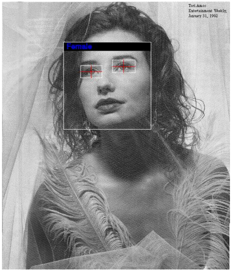
Results

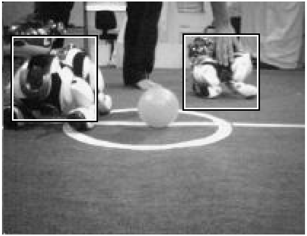

Usage of internal bootstrapping
True Positive Rate
False Positive Rate
With internal bootstrapping
Without internal bootstrapping

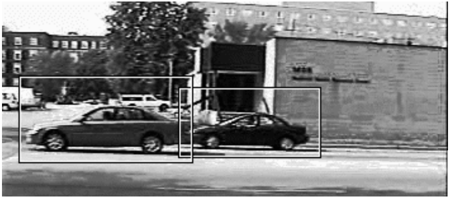

Feature sampling
True Positive Rate
False Positive Rate
20%
100%


Feature type analysis
True Positive Rate
False Positive Rate
lbp based
rectangular
rectangular(2 layers),lbp based later layers

Results: face, eye, Aibo robot, and car detection

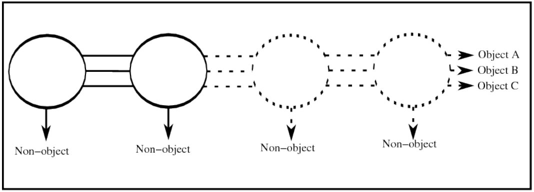






Outline

Multiclass cascade



Multiclass boosted classifier

$$\vec{H}(x) = \sum_{t=0}^{T} \vec{h}_t(f_t(x)) \tag{5}$$

Multiclass nested cascade

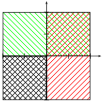
$$\vec{H}_k(x) = \vec{H}_{k-1}(x) + \sum_{t=0}^{T} \vec{h}_{t,k}(f_{t,k}(x)) - \vec{b_k} \tag{6}$$

with $\vec{H}_0(x) = 0.$

Multiclass formulation

- Multiclass case (Vectorboost, [Huang et al 2007])
  - Training sample: $(x_i, \mathbf{V}_i)$, The vector set $\mathbf{V}_i$ represents the target sub space defined as the intesection of halfspaces
  - The sample $(x, \mathbf{V}_m)$ is classified correctly

$$\iff \forall \vec{V} \in \mathbf{V}_m, \vec{V} \cdot \vec{H}(x) \geq 0 \tag{7}$$



Example: 2 overlapping objects classes and a non-object class

Class	Vectors	Objective region	Region on figure
Object 1	$(1, 0)$;	$x > 0$	Red
Object 2	$(0, 1)$;	$y > 0$	Green
Non Object	$(-1, 0); (0, -1)$	$x < 0, y < 0$	Dashed Grey

Vectorboost

- Training set $S = \{(x_i, \mathbf{V}_i)\}_{i=1,\ldots,n}$
- $\mathbf{V}_i = \{\vec{V}_i^j\}_j$: vector set representing the objective region of x_i
- Initialize Weights: $w_{i,j}(0) = 1$
- For $t = 0, \ldots, T$
  1. Normalized weigths $\{w_{i,j}(t)\}_{i,j}$ so that they add to one
  2. Select $\vec{h}_t$ and f_t, such that Z_t is minimized, with

$$Z_t = \sum_{i=1}^{n} \sum_{j:V_i^j \in \mathbf{V}_i} w_{i,j}(t) \exp\left(-\left\langle \vec{V}_i^j, \vec{h}_t(f_t(x_i))\right\rangle\right) \tag{8}$$

  3. Update weigths: $w_{i,j}(t+1) = w_{i,j}(t) \exp\left(-\left\langle \vec{V}_i^j, \vec{h}_t(f_t(x_i))\right\rangle\right)$
- return $\vec{H}(x) = \sum_t \vec{h}_t(f_t(x))$

Multiclass weak classifiers

- Independent [Huang et al 2007] $H(x,m) = \sum_{t=1}^{T} h_t(f_t(x), m)$
- Joint ([Torralba et al 2007]) $H(x,m) = \sum_{t=1}^{T} \beta_t^m h_t(f_t(x)), \beta_t^m \in \{0,1\}$
- Coupled (proposed) $H(x,m) = \sum_{t=1}^{T} \gamma_t^m h_t(f_t(x)), \gamma_t^m \in \mathbb{R}$

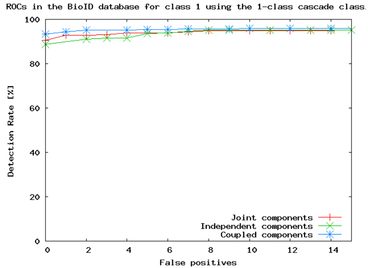
Method	Optimization Problem	Solution	N. of Prob.	N. of Variables	Order $O()$
Adaboost	$\min_{c_j} \sum_j w_+^j e^{-c_j} + w_-^j e^{+c_j}$	Analitic	J	J	$N+J$
Joint	$\min_{c_j, b_m \in \{0,1\}} \sum_{j,m} w_+^{j,m} e^{-\beta_m c_j} + w_-^{j,m} e^{\beta_m c_j}$	Analytic	2^M	$J(+M)$	$N + J2^M$
Independent	$\min_{c_{j,m}} \sum_m \sum_j w_+^{j,m} e^{-c_{j,m}} + w_-^{j,m} e^{c_{j,m}}$	Analytic	J	JM	$N+JM$
Coupled	$\min_{c_j, \gamma_m \in \mathbb{R}} \sum_{j,m} w_+^{j,m} e^{-\gamma_m c_j} + w_-^{j,m} e^{\gamma_m c_j}$	quasi-Newton	1	$J+M$	$N+(J+M)^2$

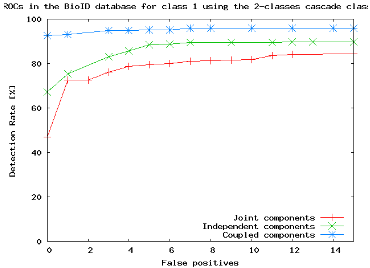
Weak classifier: domain partitioning (J bins); M classes; N training examples

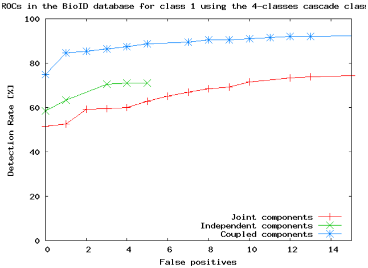
Results: Training time RIP face detection

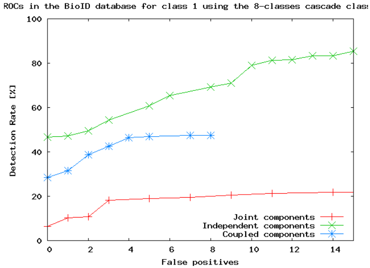
- 1-class: $\sim$2 hours all variants
- 8-classes: Independent 51 hours, Coupled 56 hours, Joint 109 hours

Results

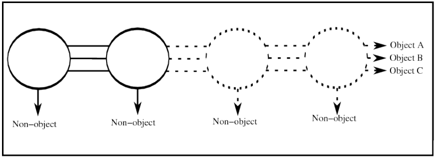

ROCs in the BioID database for class 1 using the 1-class cascade class.
Detection Rate [%]
False positives
Joint components
Independent components
Coupled components
0
2
4
6
8
10
12
14
0
20
40
60
80
100

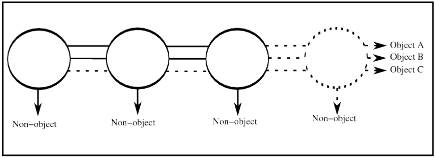

ROCs in the BioID database for class 1 using the 2-classes cascade clas
Detection Rate [%]
False positives
Joint components
Independent components
Coupled components
0
2
4
6
8
10
12
14
0
20
40
60
80
100


ROCs in the BioID database for class 1 using the 4-classes cascade clas
Detection Rate [%]
False positives
Joint components
Independent components
Coupled components
0
2
4
6
8
10
12
14
0
20
40
60
80
100


ROCs in the BioID database for class 1 using the 8-classes cascade clas
Detection Rate [%]
False positives
Joint components
Independent components
Coupled components
0
2
4
6
8
10
12
14
0
20
40
60
80
100

Coarse-to-fine (CTF) Multiclass Cascade





the inequalities $\forall \vec{V} \in \mathbf{V}_m, \vec{V} \cdot \vec{H}(x) \geq 0$ are now verified separately through the cascade.

$$\vec{H}_k(x) = \left[\vec{H}_{k-1}(x) + \sum_{t=1}^{T_k} \vec{h}_{k,t}(f_{k,t}(x)) - \vec{b}_k \right] \odot \vec{\mathbf{A}}_{k-1}(x) \tag{9}$$

$$\text{with } \vec{H}_0(x) = \vec{0}, \text{ and } \vec{\mathbf{A}}_0(x) = \vec{1} \ ,$$

$\odot$: point-wise product

Coarse-to-fine (CTF) Multiclass Cascade

Main Contributions

- Introduction of multiclass nested cascades
- Propose couple components in multiclass weak classifiers
- Extension of procedures and algorithms to the multiclass case
- Propose a CFT search in multiclass nested cascades

Results

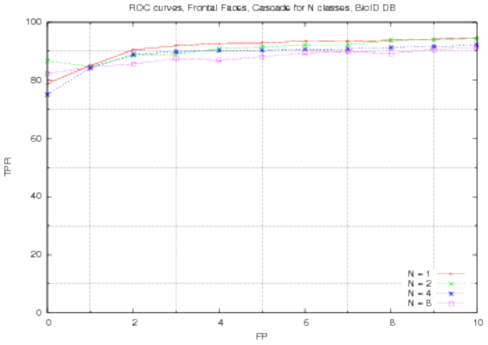
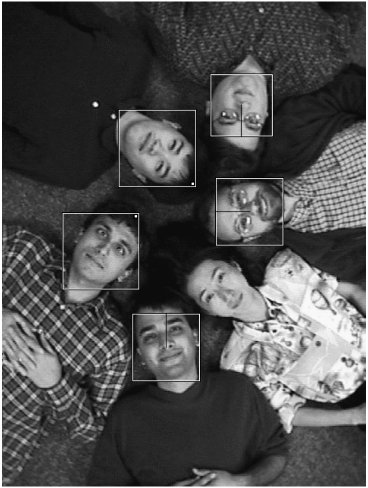


Table: Average processing time [sec] of a CTF multiclass cascade on images of 320x240 pixel size.

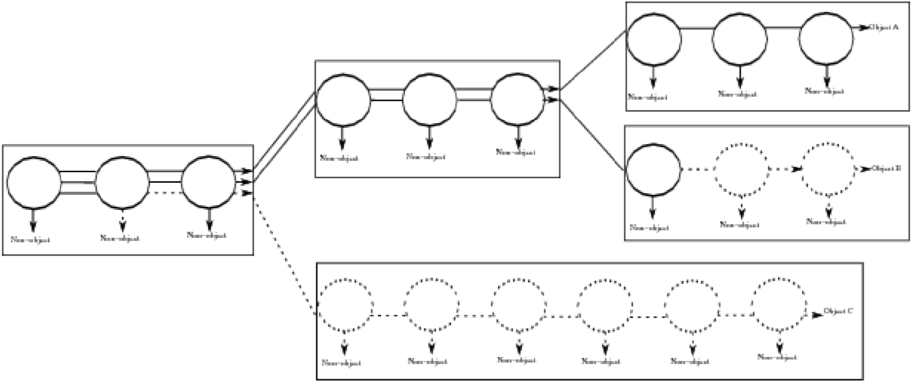
Weak classifier's training method	Number of classes			
	1	2	4	8
Processing time	0.22	0.33	0.60	0.67
Ratio of processing time: n-classes to 1-class	-	1.5	2.8	3.04

Results CTF multiclass cascade



Outline

TCAS: Tree of nested CAScades



TCAS: Tree of nested CAScades

A node $\mathbb{N}$ of $\mathbb{T}$ has $n_{\mathbb{N}}$ siblings, $\{\mathbb{N}_s\}_{s=\{1,\dots n_{\mathbb{N}}\}}$, and consist of:

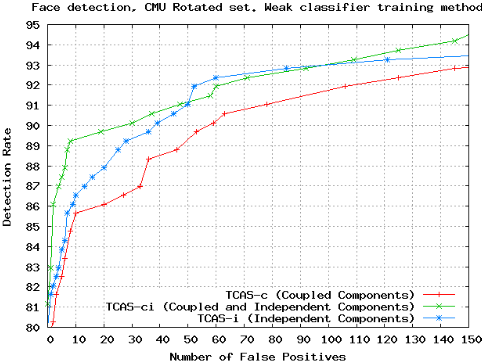
- a multiclass classifier $\vec{H}_{\mathbb{N}}$ (in our case a multiclass nested cascade),
- a mask $\vec{\mathbf{A}}_{\mathbb{N}} \in \{0,1\}^M$,
- and a "pointer", $p_{\mathbb{N}}$, to its direct ancestor.

The output of every node is defined as:

$$\vec{H}_{\mathbb{N}}(x) = \vec{H}_{p_{\mathbb{N}}}(x) \odot \vec{\mathbf{A}}_{\mathbb{N}} + \vec{H}_{\mathbb{N}}(x), \tag{10}$$

with $\vec{H}_{p_{\mathbb{N}}}(x)$ the output of the ancestor of $\mathbb{N}$, $p_{\mathbb{N}}$.

Results, RIP face detection: weak classifier

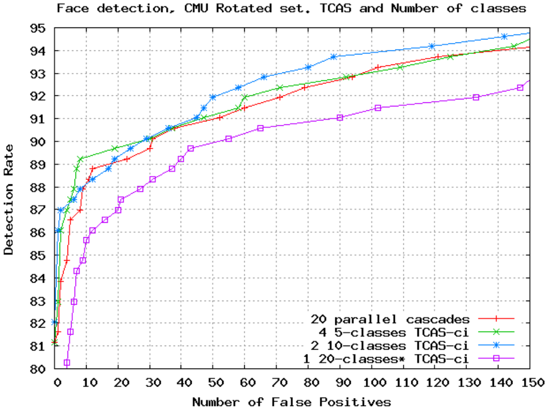


Running time of tcas-c detector:

- is ∼1.17 times faster than tcas-i
- is ∼1.22 times faster than tcas-ci

Training times: tcas-i (187 min.), tcas-ci (176 min.), tcas-c (161 min.) (Intel Core i7 at 2.67GHz with Linux and using 8 virtual cores)

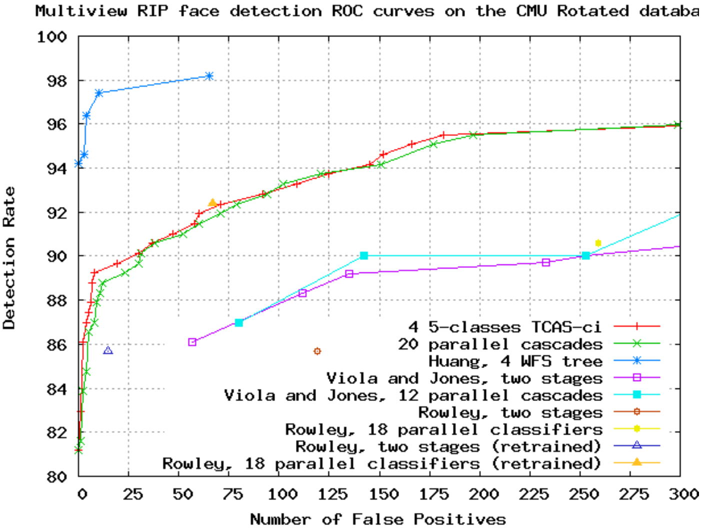
Results, RIP face detection: number of classes



Compared to 20 parallel cascades

- 5-classes tcas-ci is ∼1.7 times faster (applied 4 times)
- 10-classes tcas-ci is ∼1.8 times faster (applied 2 times)
- 20-classes tcas-ci is ∼1.2 times faster

Results



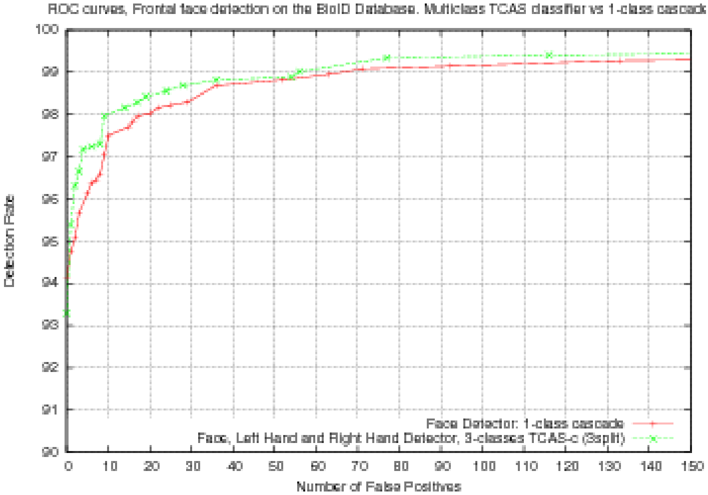
Multiview RIP face detection ROC curves on the CMU Rotated databa
Detection Rate
Number of False Positives
100
98
96
94
92
90
88
86
84
82
80
0
25
50
75
100
125
150
175
200
225
250
275
300
4 5-classes TCAS-ci
20 parallel cascades
Huang, 4 WFS tree
Viola and Jones, two stages
Viola and Jones, 12 parallel cascades
Rowley, two stages
Rowley, 18 parallel classifiers
Rowley, two stages (retrained)
Rowley, 18 parallel classifiers (retrained)

Results

Results

Results

Results: Hand and Face detection



Results: Hand and Face detection

Table: Average Processing times [sec] of frontal hand (Right and Left fist) and frontal face detector(s) on the CMU database.

Classifier	Classes	CMU-MIT
Three 1-class cascades	Face & Hands	1.70
1-class + 2-classes tcas	Face & (Left &Right Hand)	1.27
1-class + 2-classes tcas	Left & (Right Hand & Face)	1.39
3-classes cascade	Face & Hands	did not converge
3-classes tcas-i (split3)	Face & Hands	**1.01**
3-classes tcas-c (split3)	Face & Hands	1.09

Results: Hand and Face detection

Outline

Publications directly related to the thesis work

1 ISI journal article, 1 book chapter and 3 conference papers:

- **Verschae R.**, Ruiz-del-Solar J., Correa M., *A unified learning framework for object detection and classification using nested cascades of boosted classifiers*, Machine Vision and Applications 19(2): 85-103, 2008 (**ISI**)
- **Verschae, R.**, Ruiz-del-Solar, J., *State of the Art Face Detection: Cascade Boosting Approaches* , pp 406-410, in Tiwari, A.; Knowles, J.; Avineri, E.; Dahal, K.; Roy, R. (Eds.), Applications of Soft Computing: Recent Trends, 2006. (**Book Chapter**)
- **Verschae R.**, Ruiz-del-Solar J., *Multiclass Adaboost and Coupled Classifiers for Object Detection*, Lecture Notes in Computer Science 5197, pp. 560-567, Proceedings of CIARP 2008, Havana, Cuba, 2008
- **Verschae R.**, Ruiz-del-Solar J., Correa M., *Gender Classification of Faces Using Adaboost*, 11th Iberoamerican Congress in Pattern Recognition, CIARP 2006. LNCS 2445, pp 68-78, 2006 (**ISI**)

Publication summary

- Directly related to the thesis: 1 ISI journal article, 1 book chapter and 3 conference papers
- Partially related to the thesis: 4 conference papers
- Other publications at Univ. de Chile: 1 ISI journal article and 3 conference papers
- Other publications: 1 ISI journal article and 3 conference papers

Total

- 3 ISI journal papers, 1 book chapter and 13 conference papers

- Submitted (Univ. de Chile): 1 journal article and 3 conference papers
- Submitted (others): 1 journal article and 1 conference paper

Summary of contributions and results

- Framework to train cascade and tree (multiclass) object detectors
- Efficient training methods
- Fast processing (compared to parallel cascades)
- Successfully built several classifier
- Coupled weak multiclass classifiers (use at the first layers)
- CTF search in multiclass nested cascades
- CTF nested tree based classifier (tcas)

Future work and research lines

- test other kinds of features and more classes
- develop a method to group the classes
- selection of weak classifier training method at each level of the tree
- encode previous knowledge when using vectorboost

Thanks