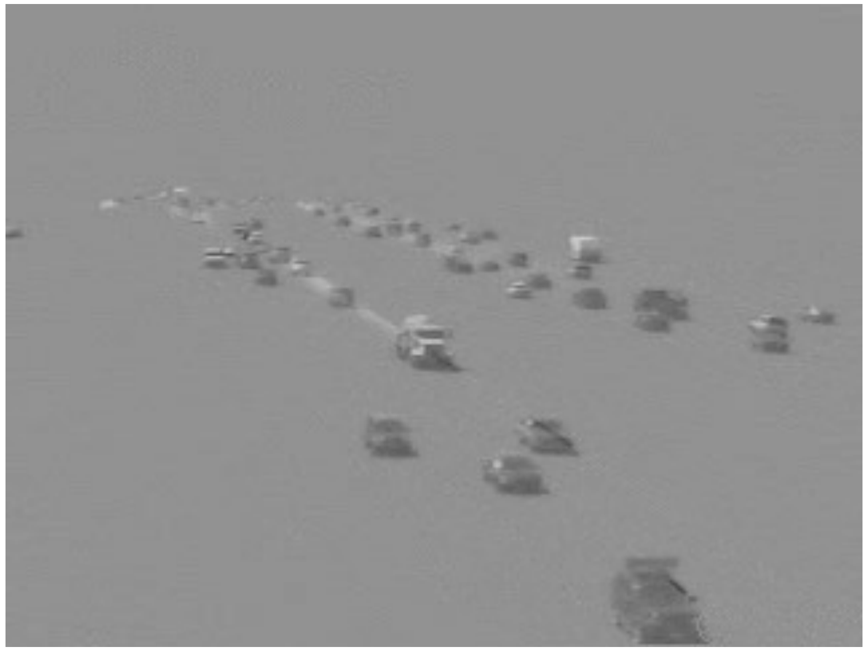
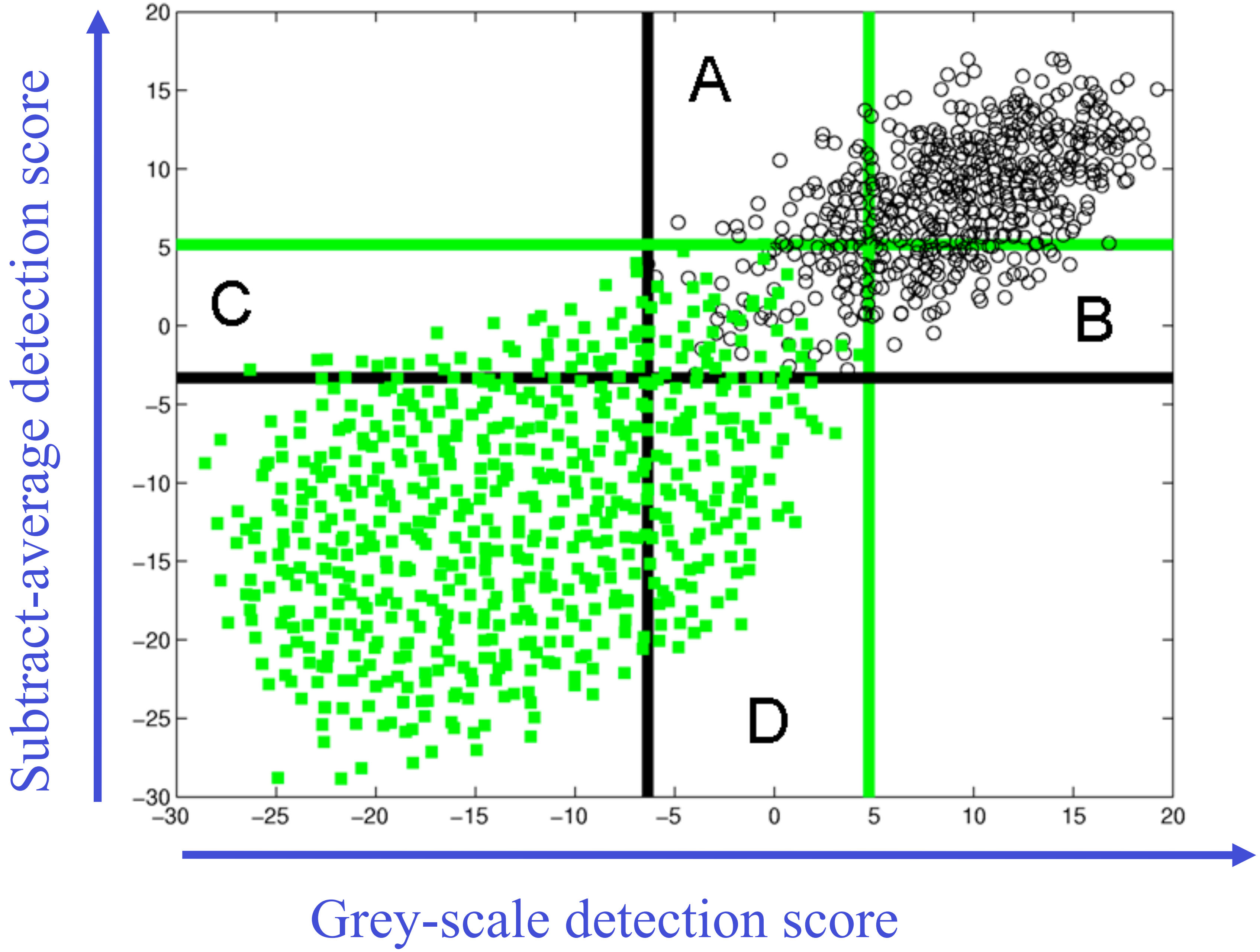
Active learning

Co-training


I-405
NE 124TH ST
© WSDOT


SR 520
MIDSPAN
© WSDOT

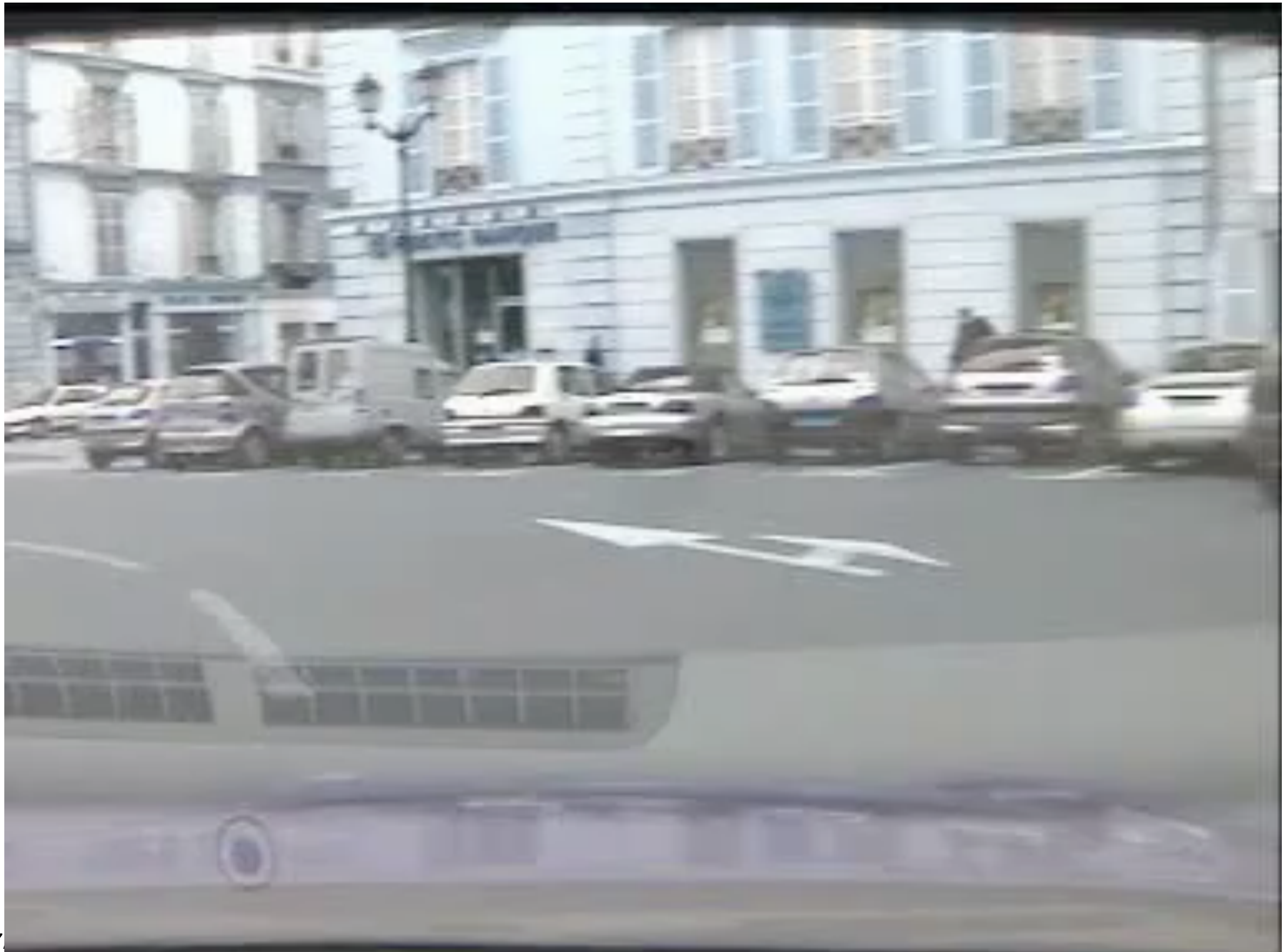



A
B
C
D
Subtract-average detection score
Grey-scale detection score

Summary

- **Boosting** is a method for learning an accurate classifiers by combining many weak classifiers.
- Boosting is **resistant to over-fitting**.
- **Margins** quantify prediction confidence.
- **High noise** is a serious problem for learning classifiers- can't be solved by minimizing convex functions.
- **Robustboost** can solve some high noise problems. Exact characterization still unclear.
- **Jboost** - an implementation of ADTrees and various boosting algorithms in java.
- **Book** on boosting coming this spring.
- Thank you, questions?

Pedestrian detection - typical segment



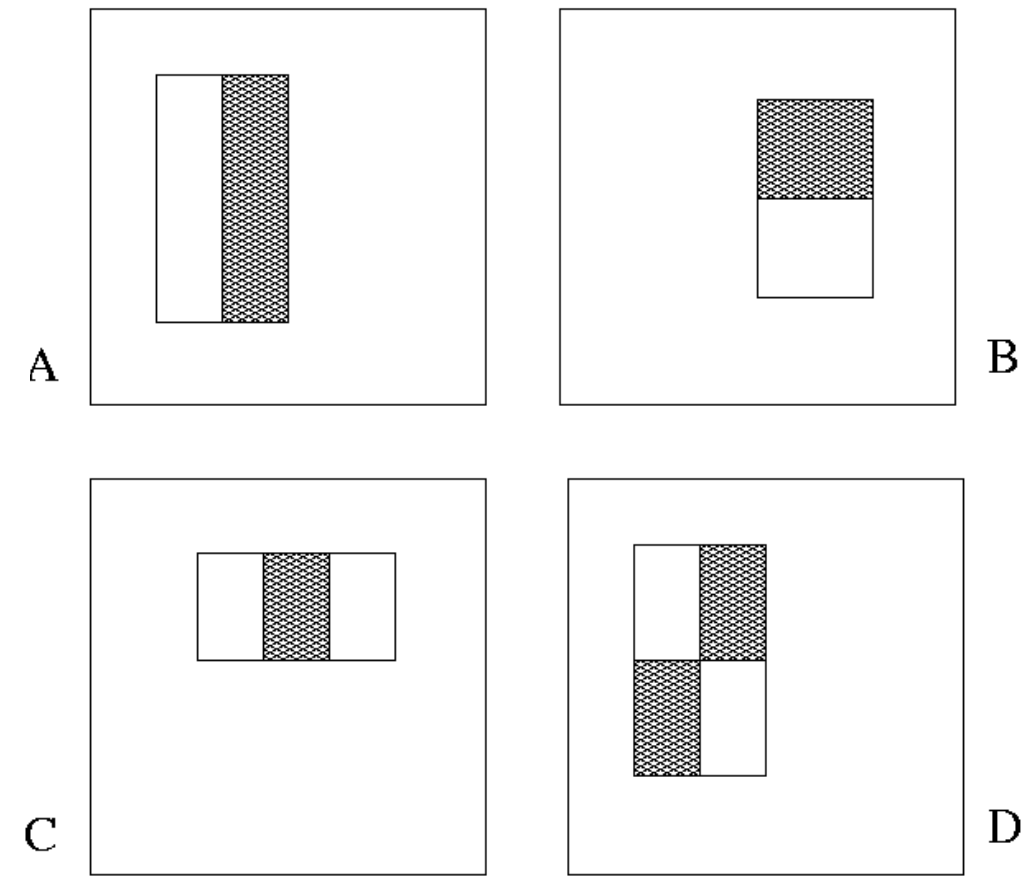
Current best results

Image Features

"Rectangle filters"

Similar to Haar wavelets

Papageorgiou, et al.



$$h_t(x_i) = \begin{cases} 1 & \text{if } f_t(x_i) > \theta_t \\ 0 & \text{otherwise} \end{cases}$$

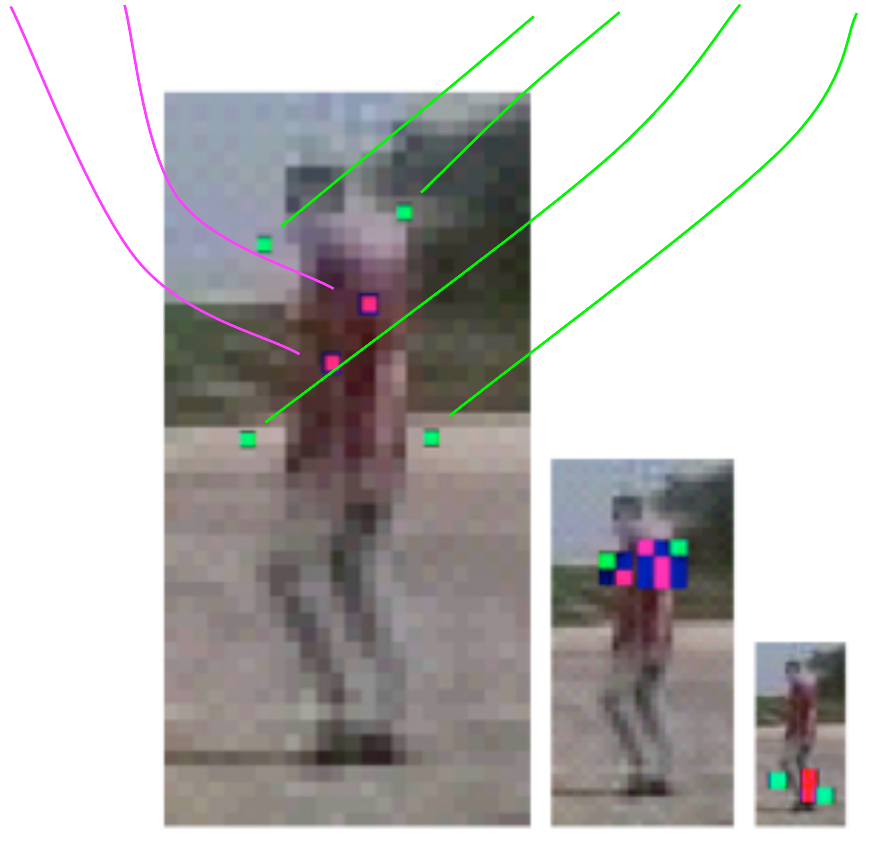
Very fast to compute using "integral image".

$60{,}000 \times 100 = 6{,}000{,}000$
Unique Binary Features

Combined using adaboost

Yotam's features

max (p1,p2) < min(q1,q2,q3,q4)



Faster to calculate than Viola and Jones

Search for a good feature based on genetic programming

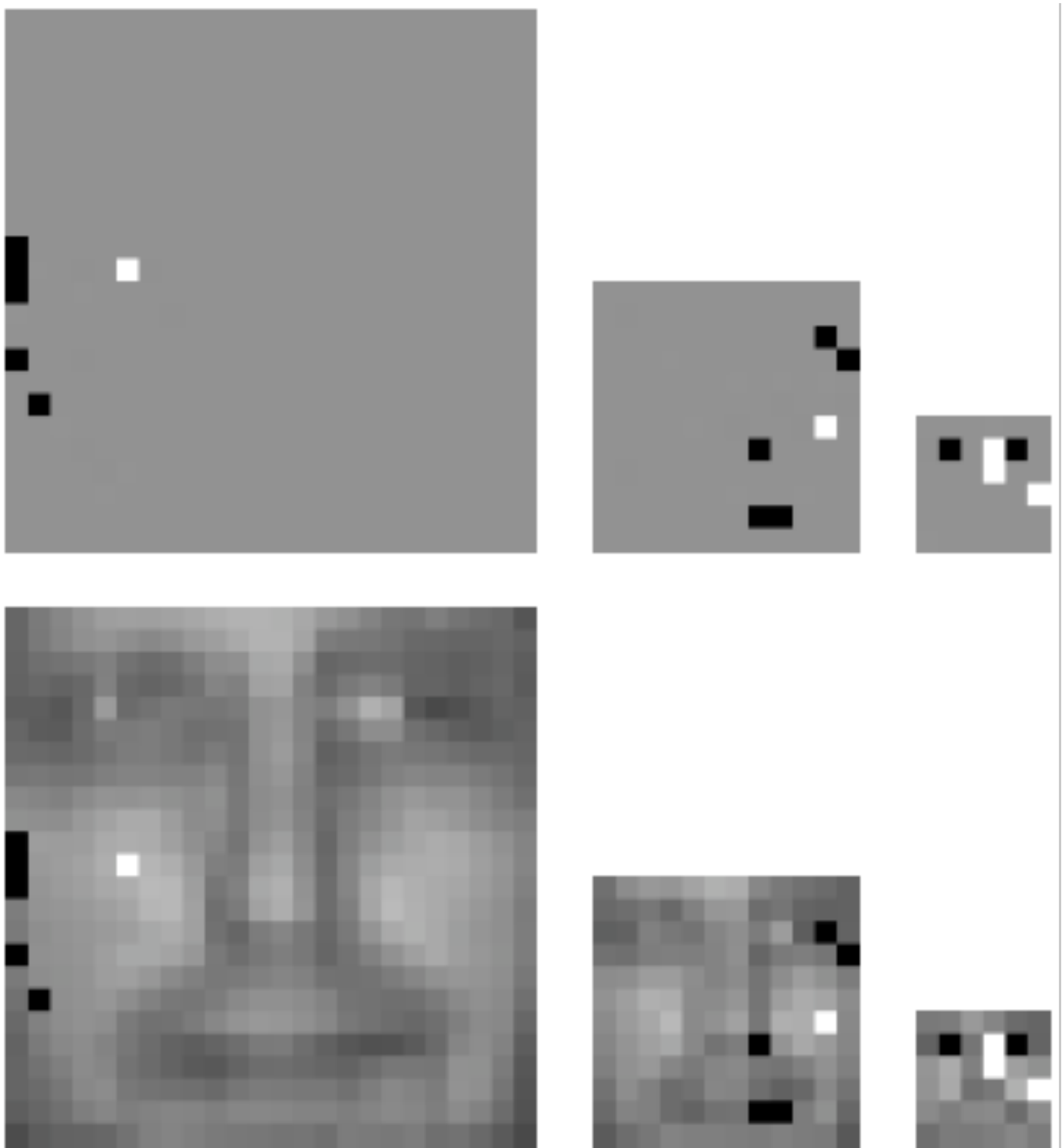
Definition

•Feature works in one of 3 resolutions: full, half, quarter

•Two sets of up to 6 points each

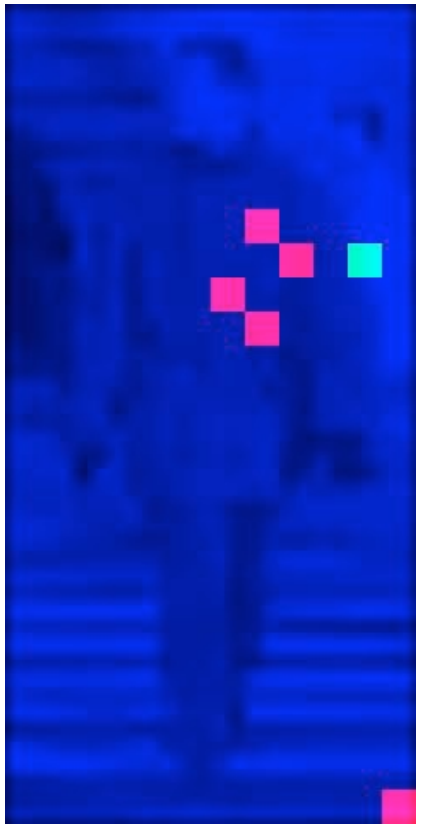
•Each point is an individual pixel

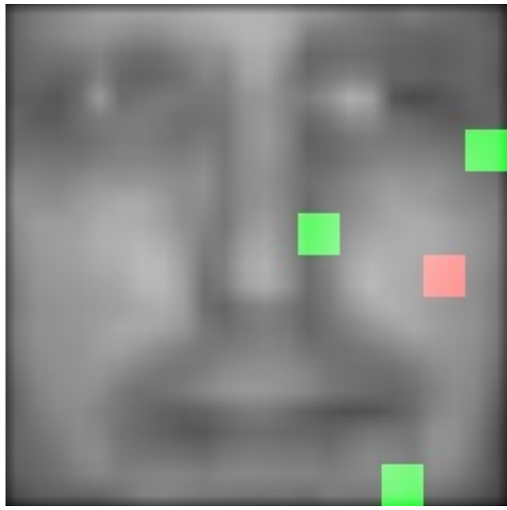
•Feature says yes if *all* white points have higher values then *all* black points, or vice versa

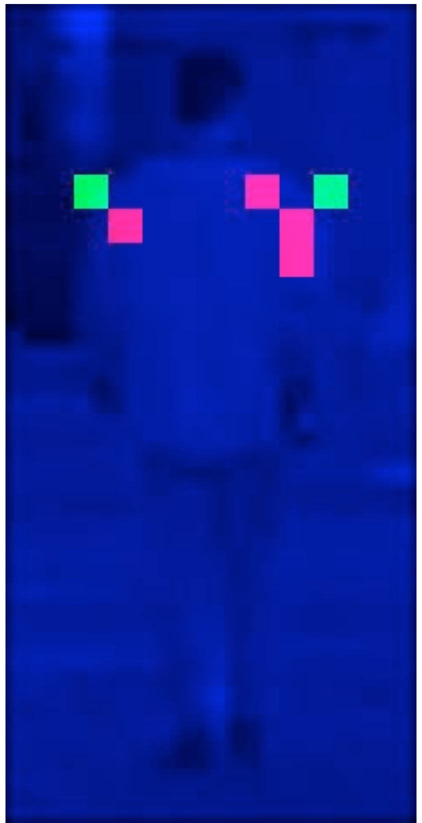


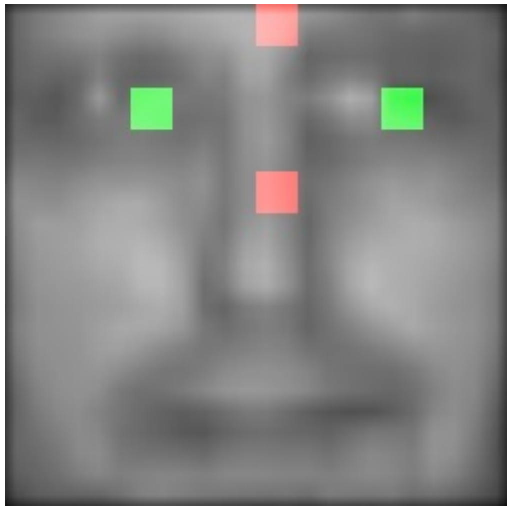
Advantages

- Deal better with the variation in illumination, no need to normalize.
- Highly efficient (3-4 image access operations). 2 times faster than Viola&Jones
- 20% of the memory





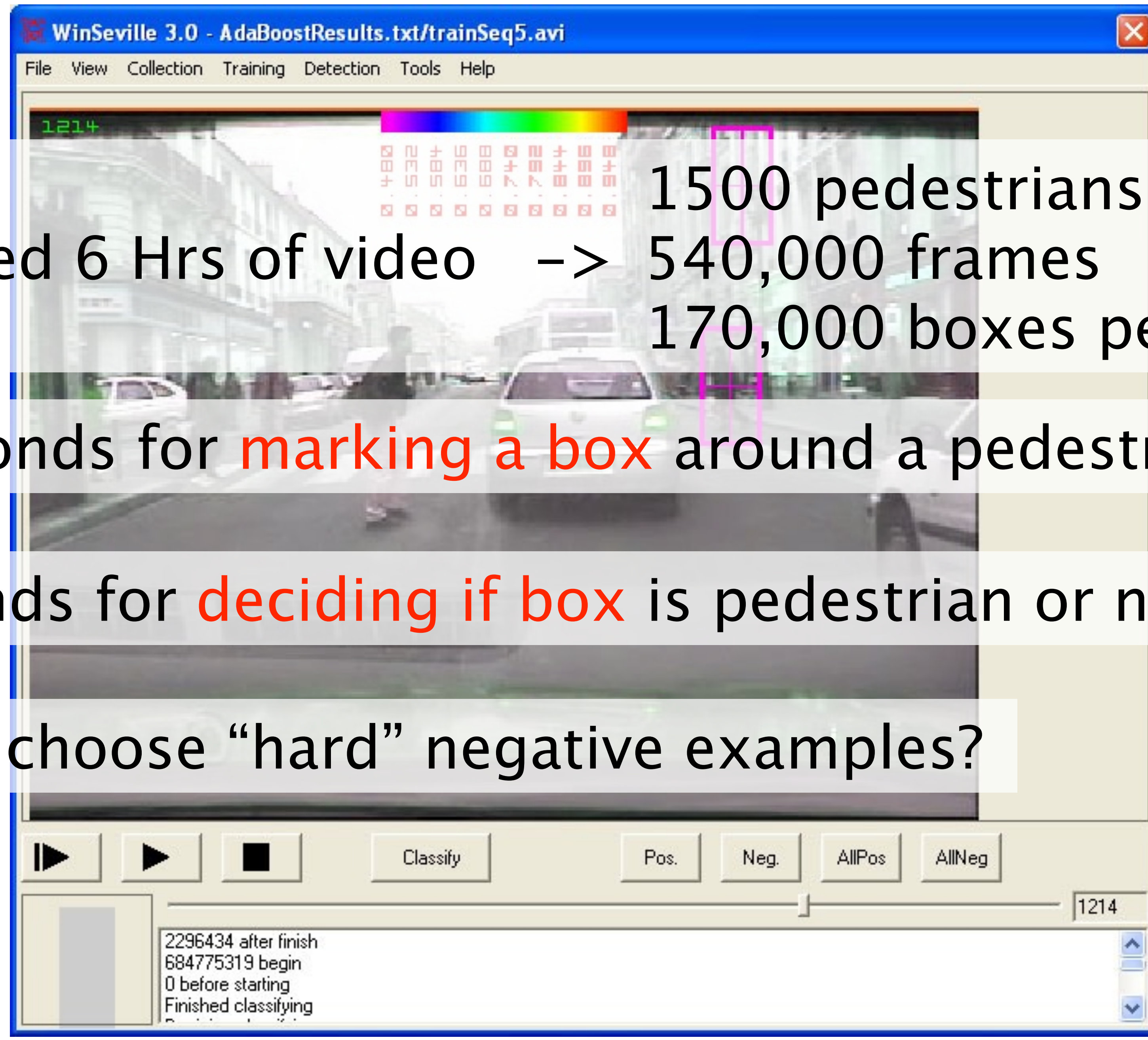




Steps of batch learning

- Collect labeled examples
- Run learning algorithm to generate classification rule
- Test classification rule on new data.

Labeling process



Collected 6 Hrs of video -> 1500 pedestrians
540,000 frames
170,000 boxes per frame

20 seconds for marking a box around a pedestrian.

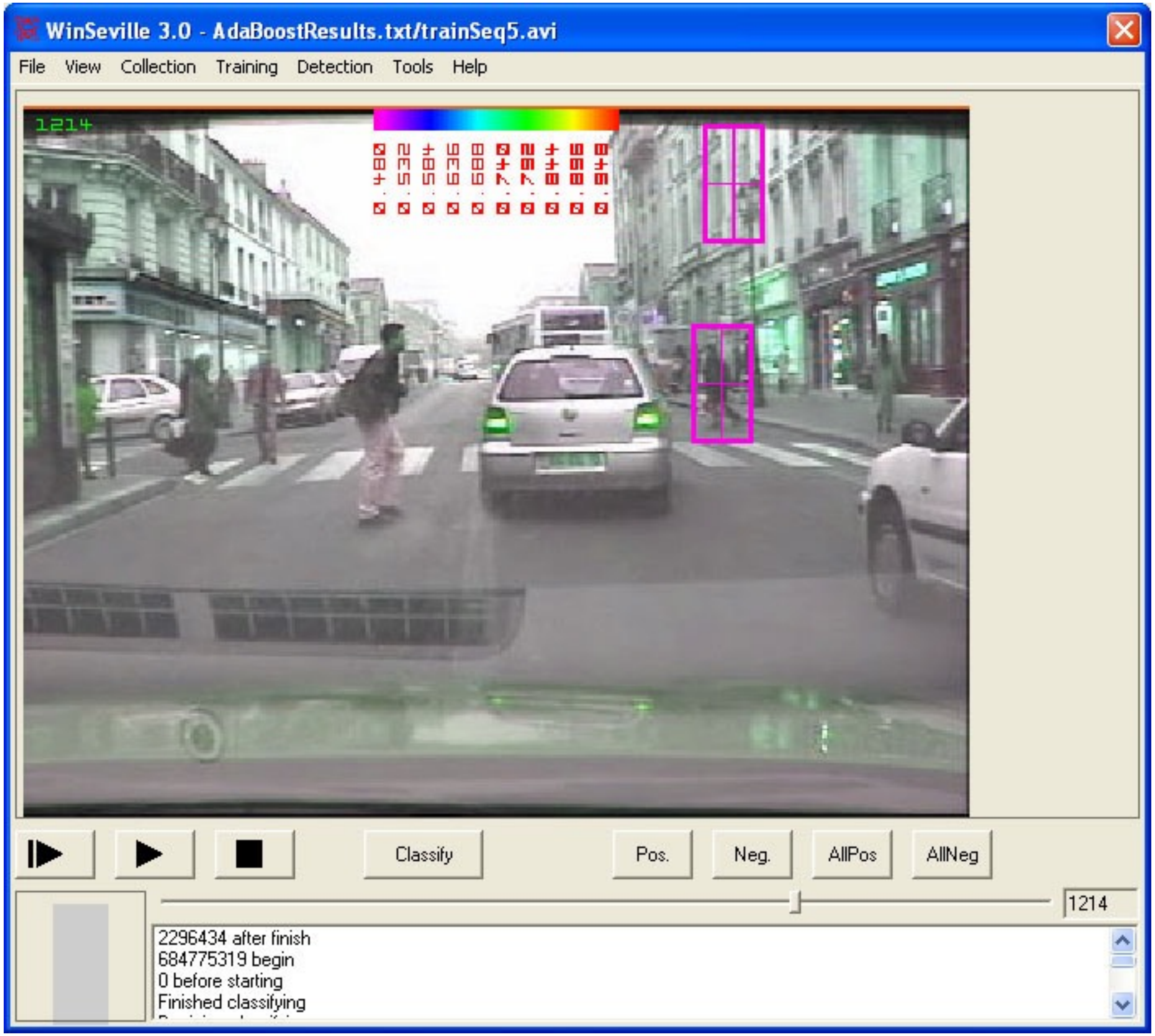
3 seconds for deciding if box is pedestrian or not.

How to choose "hard" negative examples?

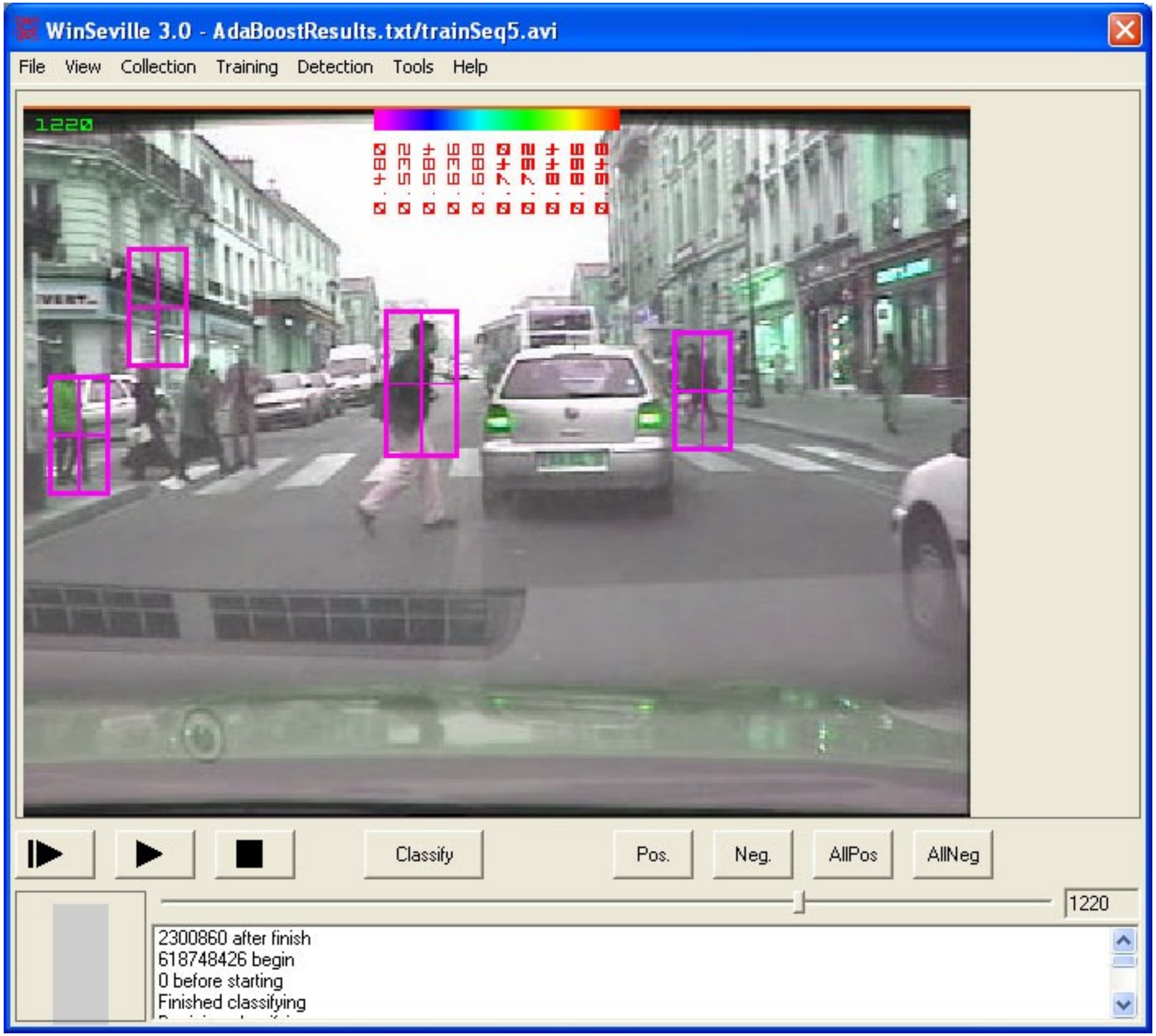
Steps of active learning

- Collect labeled examples
- Run learning algorithm to generate classification rule
- Apply classifier on new data.
  and
  label informative examples.

SEVILLE screen shot 1



SEVILLE screen shot 2



Margins

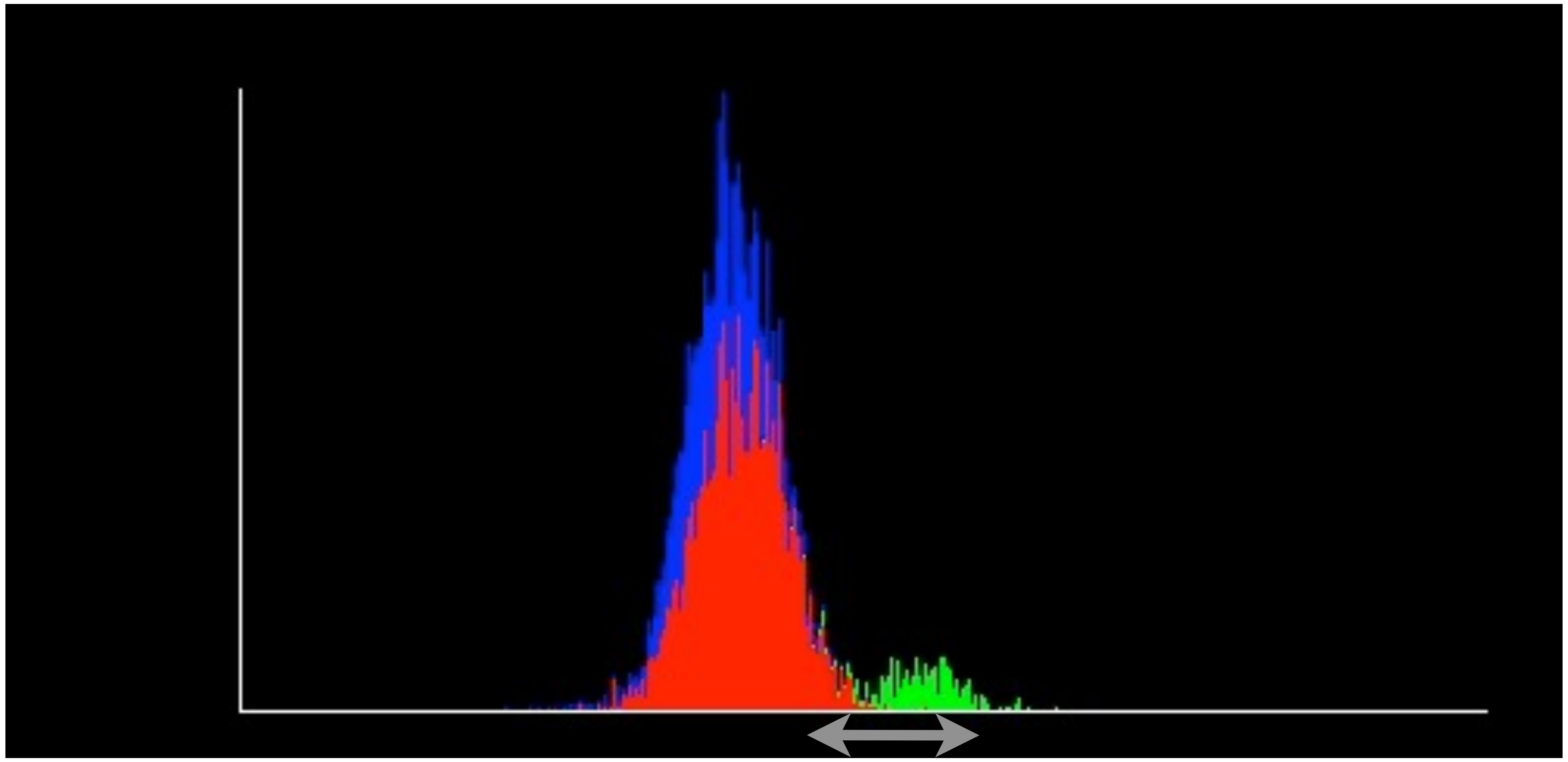
Consider the following:

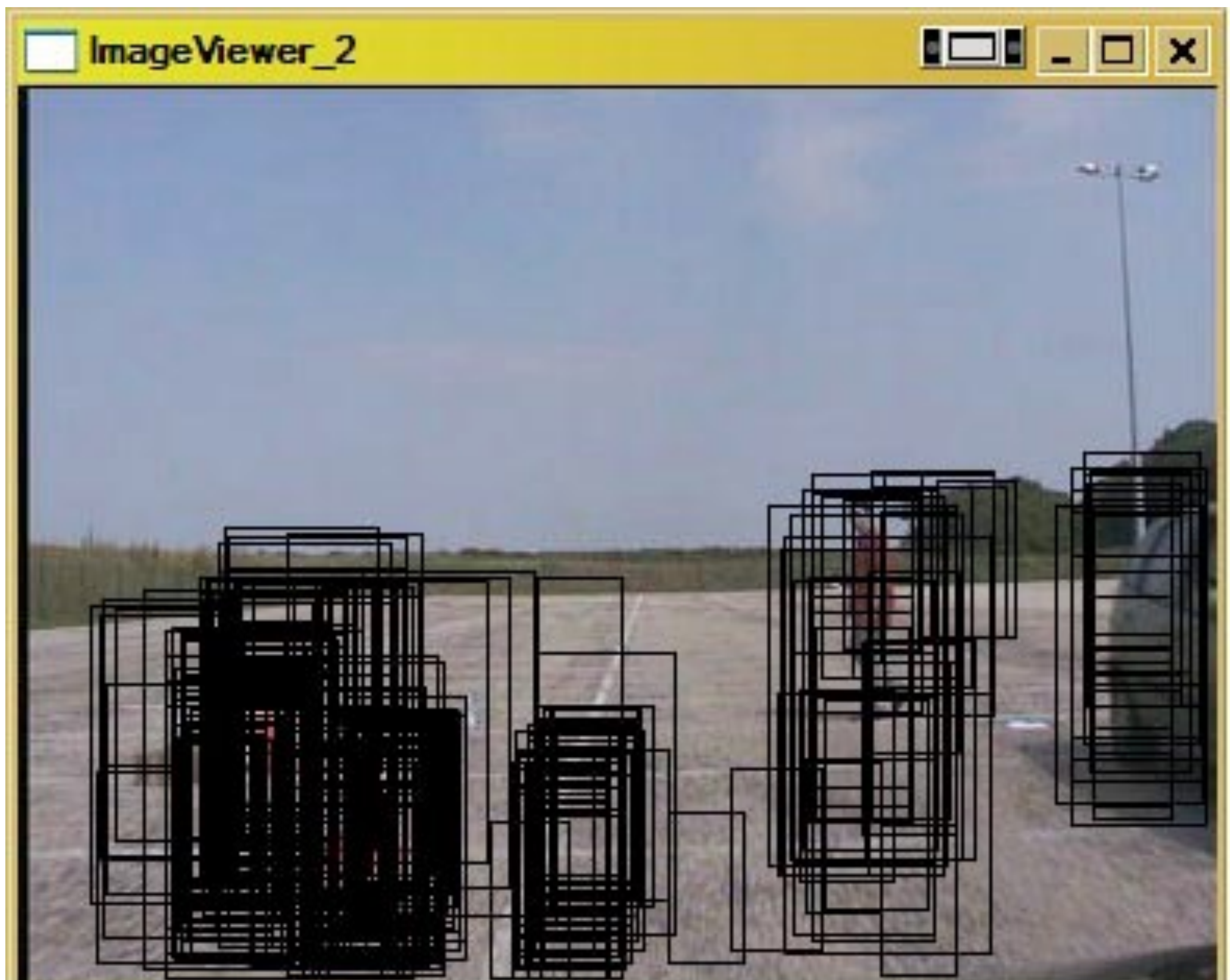
An example: <x,y> e.g. <  ,+1>

Normalized score: $$-1 \le \frac{\sum_{t=1}^{T} \alpha_t h_t(x)}{\sum_{t=1}^{T} |\alpha_t|} \le 1$$

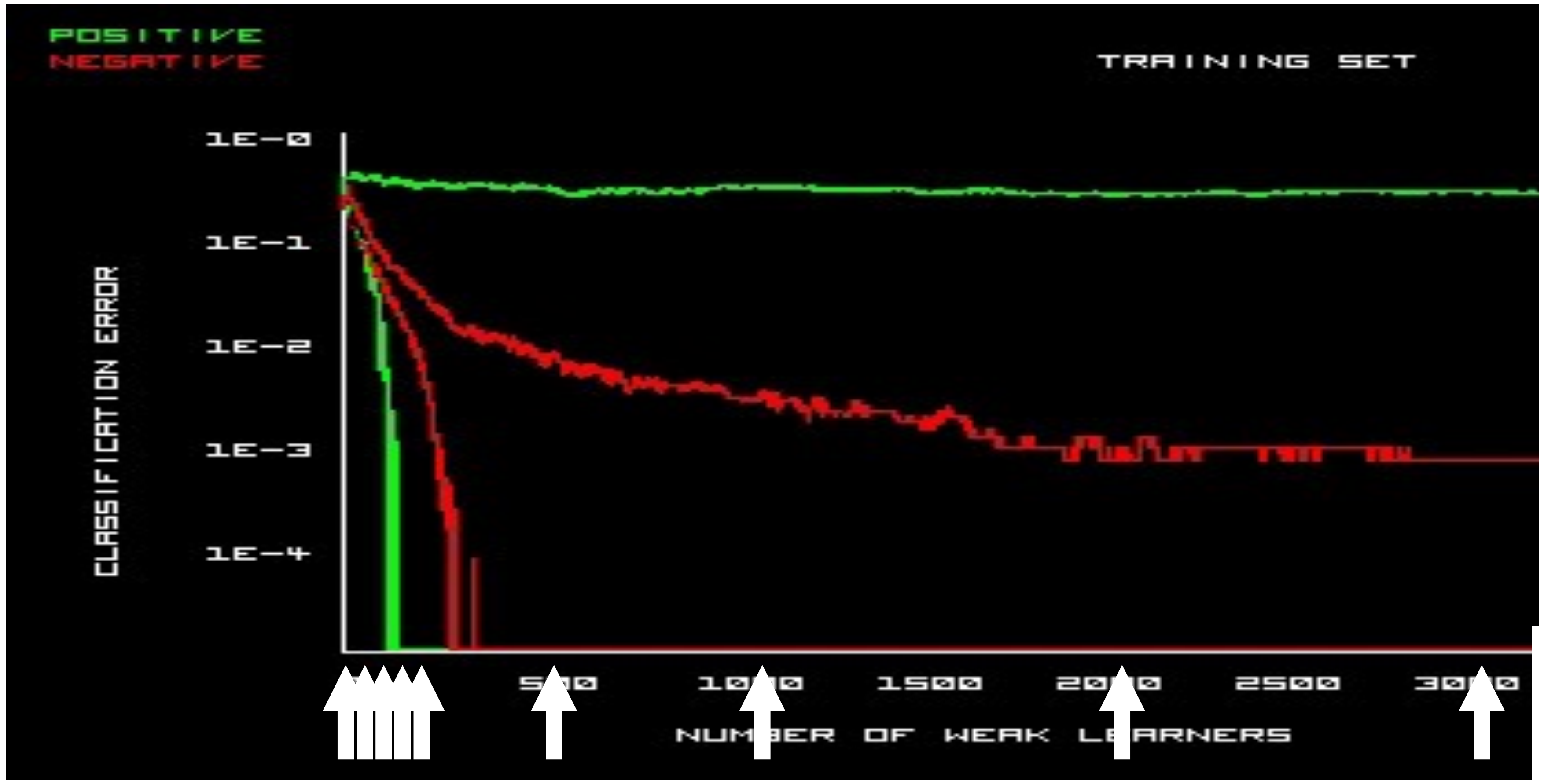
The margin is: $$y \frac{\sum_{t=1}^{T} \alpha_t h_t(x)}{\sum_{t=1}^{T} |\alpha_t|}$$

margin > 0 means correct classification



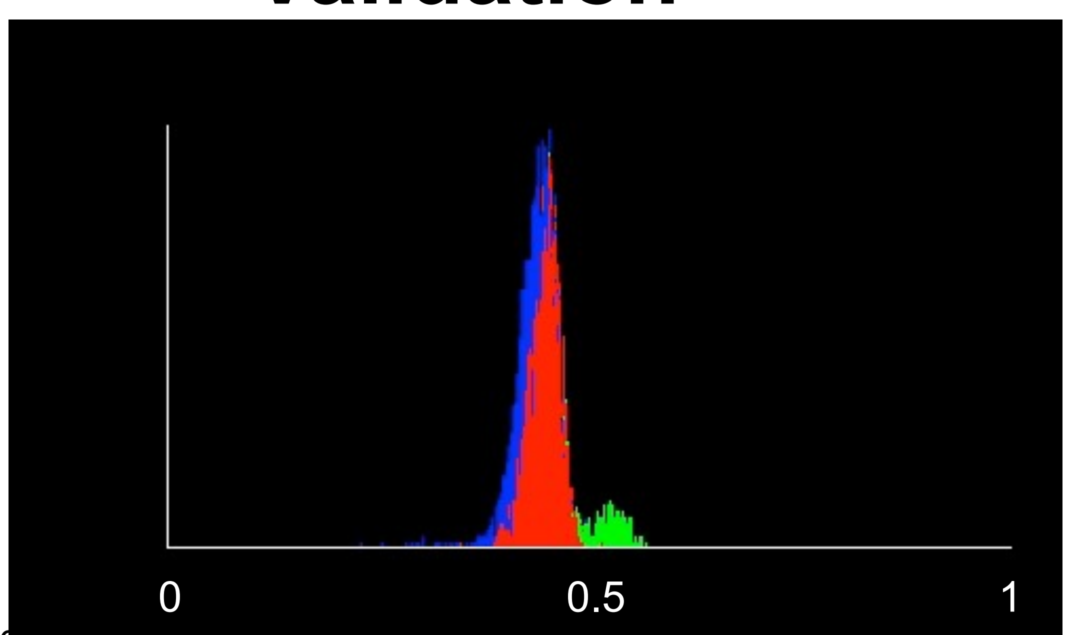

ImageViewer_2

large margins => reliable predictions

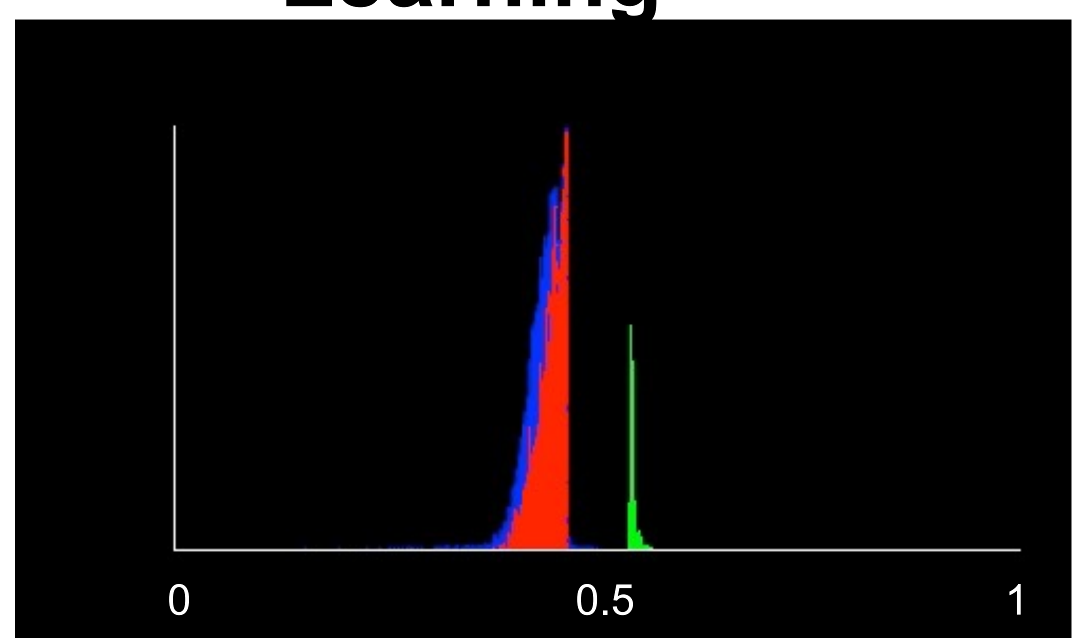


3000

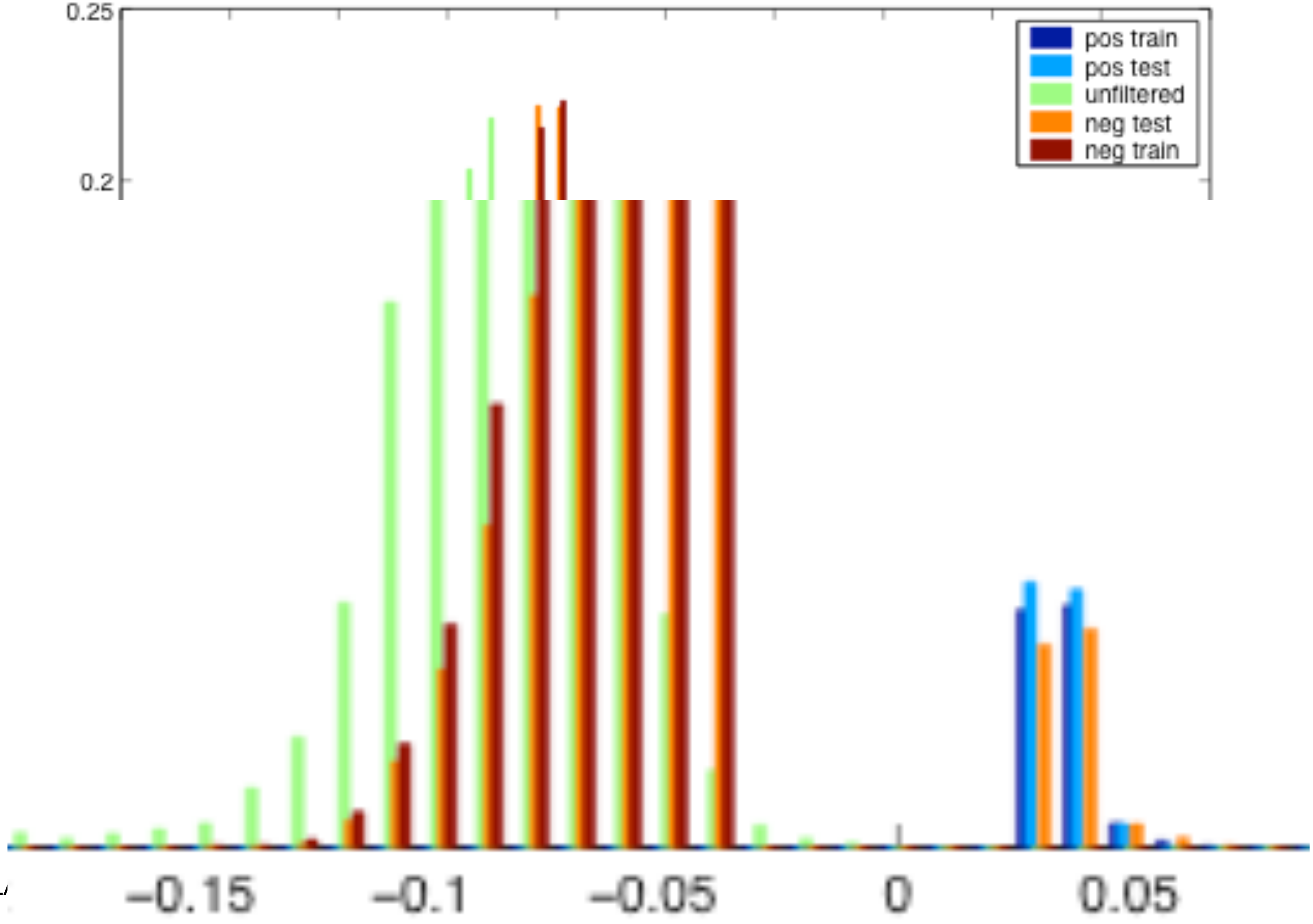
Validation



Learning



Margin Distributions



Summary of Training effort

Step	total candidates			presented	labeled	human labor	positive	negative	training time	Weak rules
1	510 K			-	16	3m	6	10	2s	1
2	680 K			364	403	3m	36	374	6s	3
3	3,400 K			153	156	4m	46	520	22s	7
4	66,470 K			805	852	10m	86	1332	1m30s	30
5	37,910 K			1350	1439	10m	182	2675	8m	59
6	116,960 K			5150	5364	1h30m	417	7804	1h10m	270
7	24,140 K			1320	863	3h	848	8236	7h30m	893
8	189,550 K			8690	8707	3h	1178	16613	17h	1500
9	209,610 K			2933	2933	3h	1486	19238	30h	2034
10	274,210 K			3861	3861	4h	2046	22533	30h	3150

Summary of Training

Only examples whose score is in this range are hand - labeled

Step	total candidates	μ^-	μ^+	presented	labeled	human labor	positive	negative	training time	Weak rules
1	510 K	-	-	-	16	3m	6	10	2s	1
2	680 K	0	1	364	403	3m	36	374	6s	3
3	3,400 K	0.6	1	153	156	4m	46	520	22s	7
4	66,470 K	0.4	1	805	852	10m	86	1332	1m30s	30
5	37,910 K	0.1	0.8	1350	1439	10m	182	2675	8m	59
6	116,960 K	0	0.6	5150	5364	1h30m	417	7804	1h10m	270
7	24,140 K	-0.02	0.5	1320	863	3h	848	8236	7h30m	893
8	189,550 K	-0.02	0.5	8690	8707	3h	1178	16613	17h	1500
9	209,610 K	-0.02	0.5	2933	2933	3h	1486	19238	30h	2034
10	274,210 K	-0.02	0.5	3861	3861	4h	2046	22533	30h	3150

Few training examples

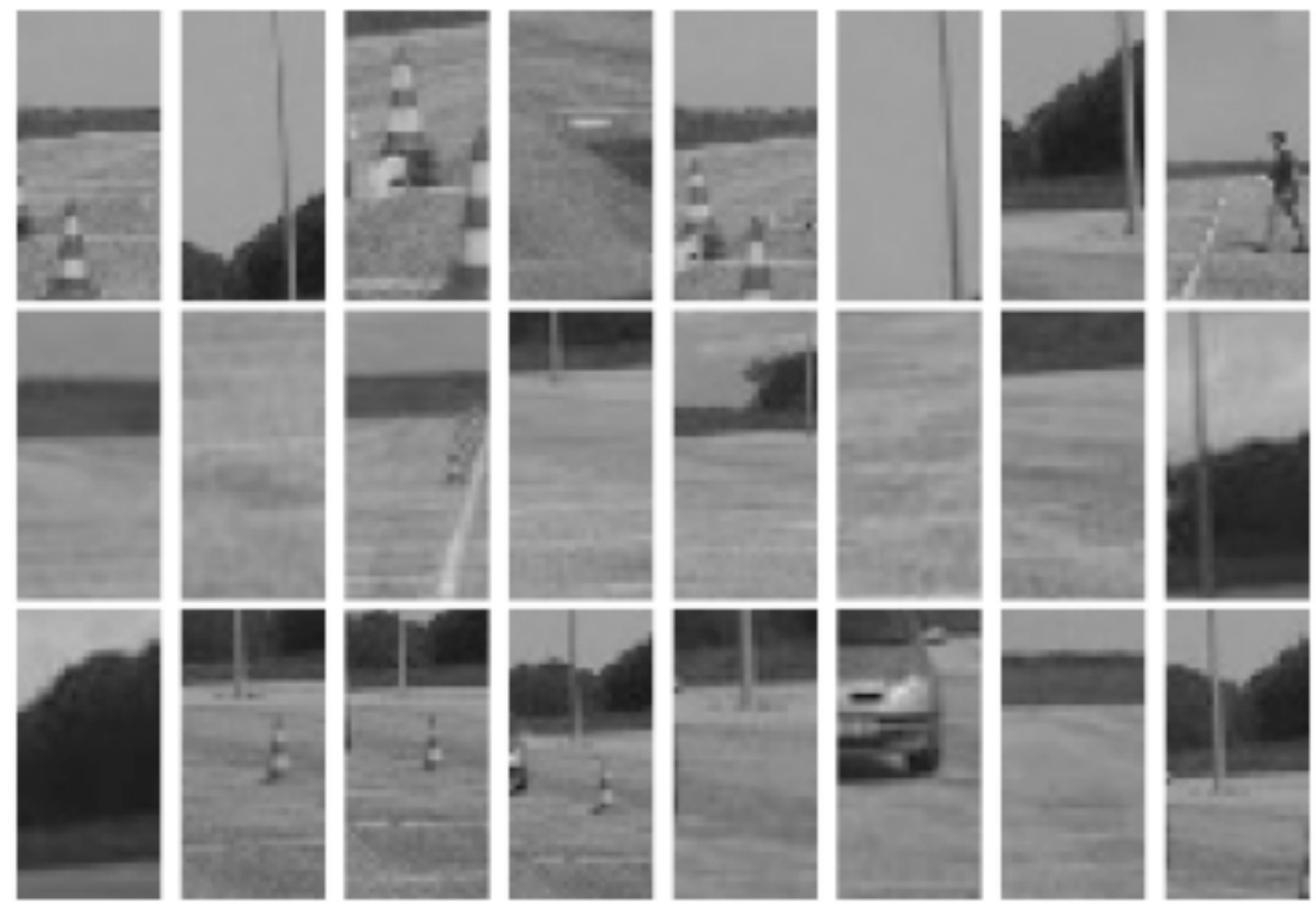
After re-labeling feedback

Final detector

Examples - easy

Positive

Negative

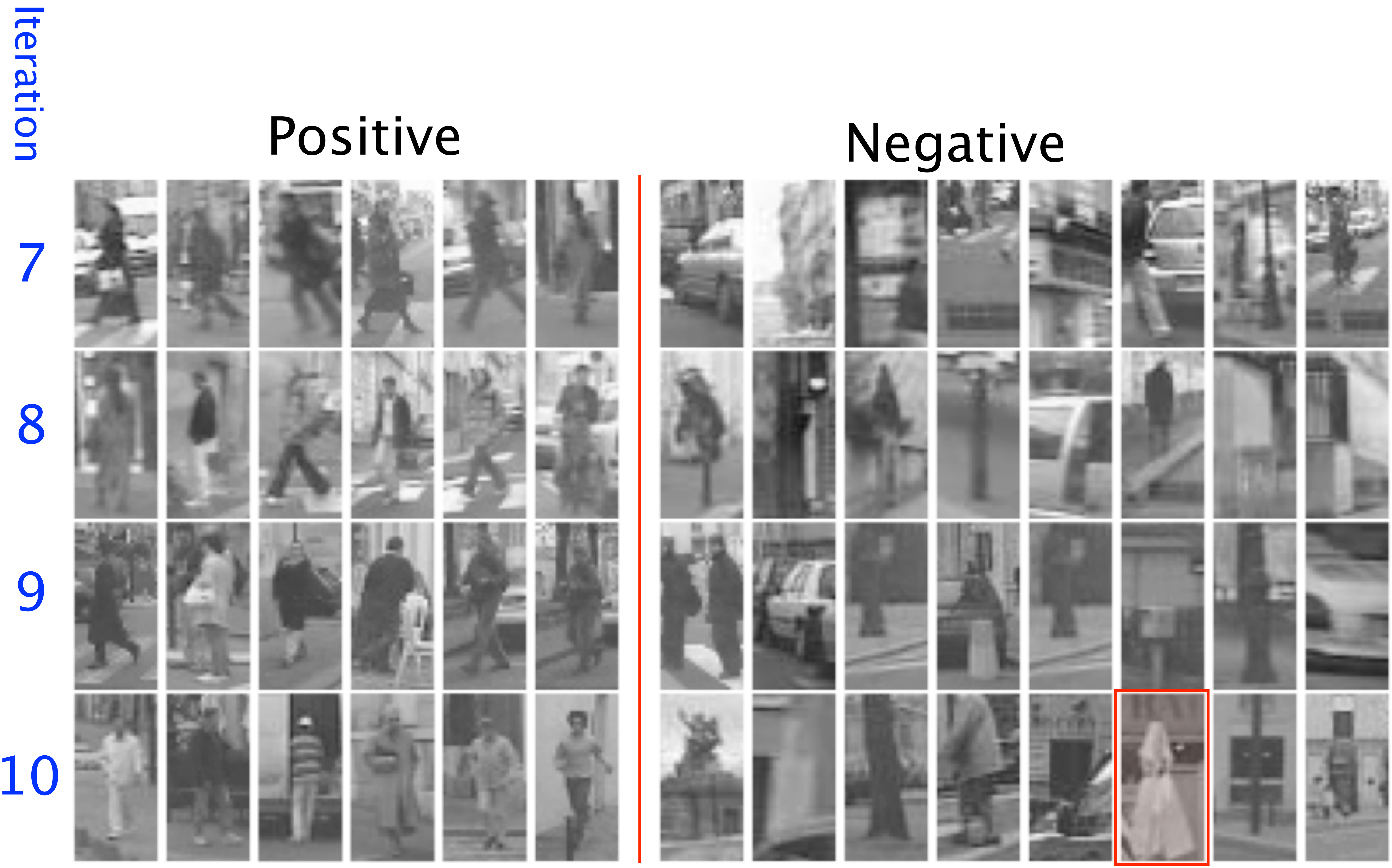


Examples - medium

Positive

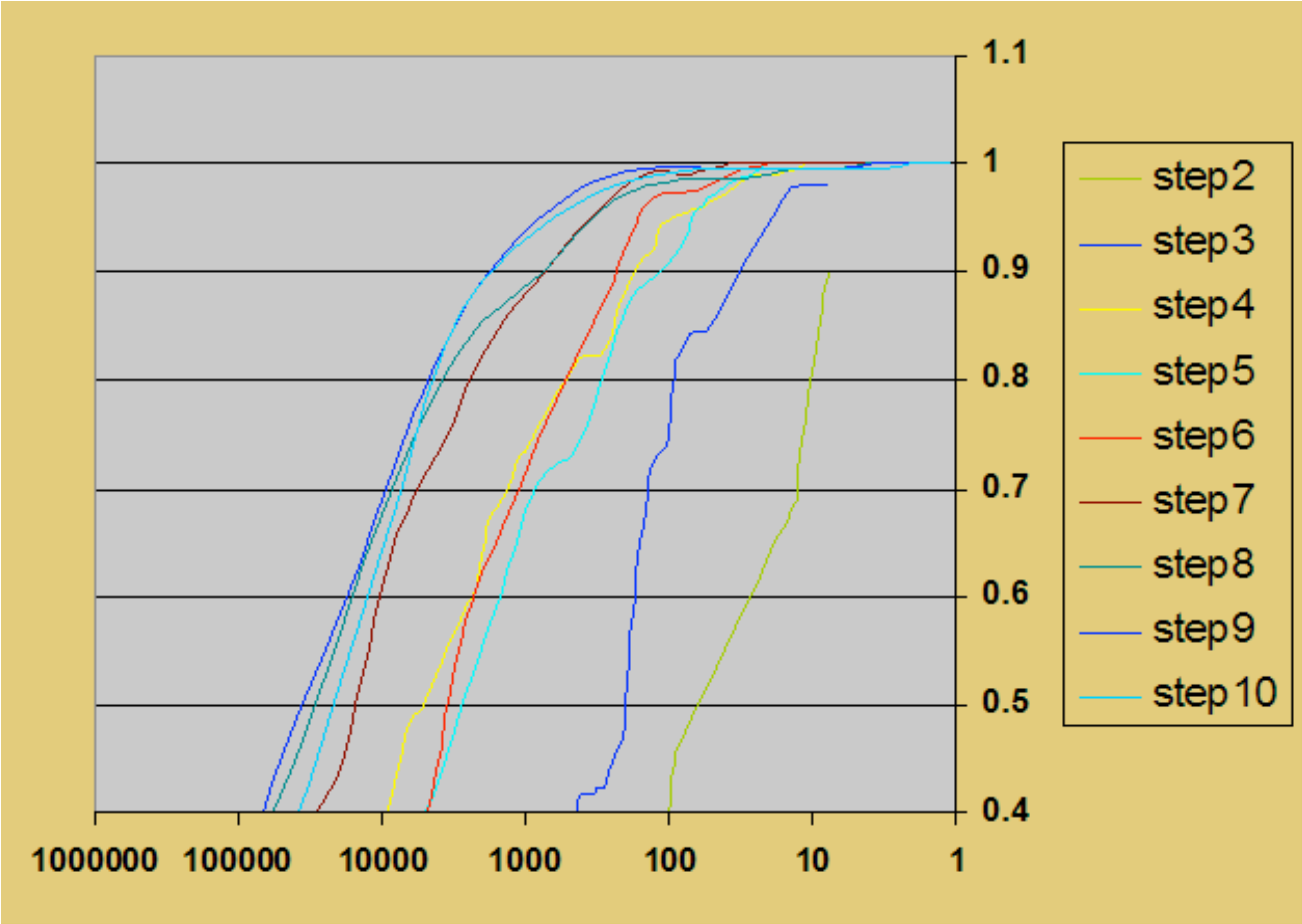
Negative

Examples - hard



And the figure in the gown is..

Seville cycles



Summary

- Boosting and SVM control over-fitting using margins.
- Margins measure the **stability** of the prediction, not conditional probability.
- Margins are useful for co-training and for active-learning.