Visual Tracking with Online Multiple Instance Learning

Boris Babenko, Ming-Hsuan Yang, Serge Belongie

Kelsie Zhao

Content

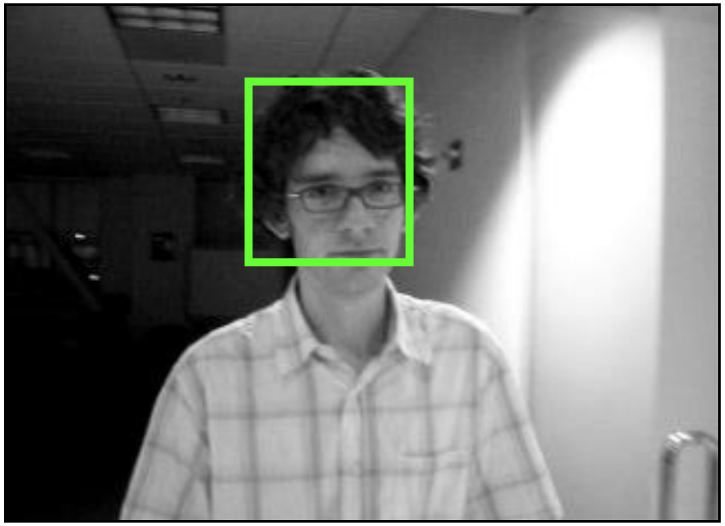
- Goal
- Background: Tracking by Detection
- Previous Work
- New Tracking Solution
  - MILTrack
  - Online MILBoost
- Experiments & Results

Goal

Track one arbitrary object in video, given its location in first frame

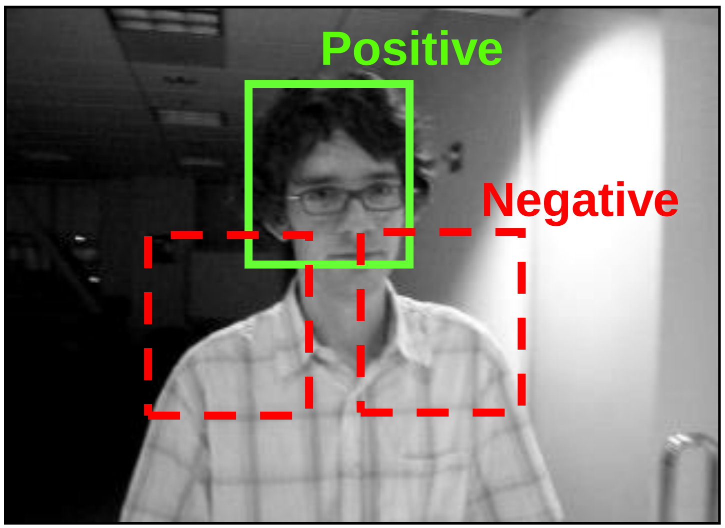
Background: Tracking by detection

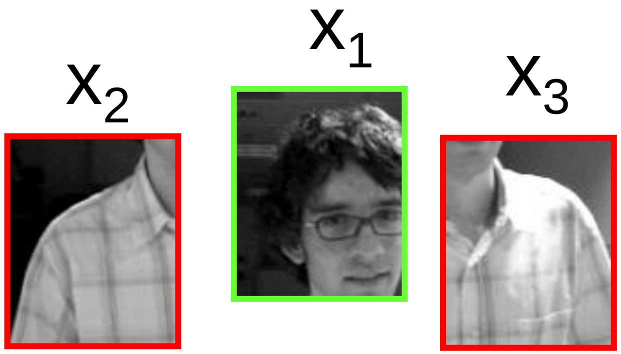
- Frame 1 is labeled, tracker location known



Background: Tracking by detection

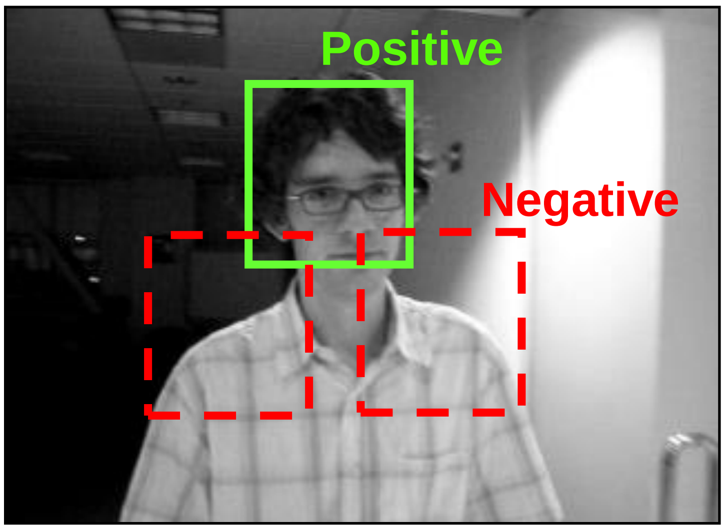
- Crop one positive and some negative patches near tracker

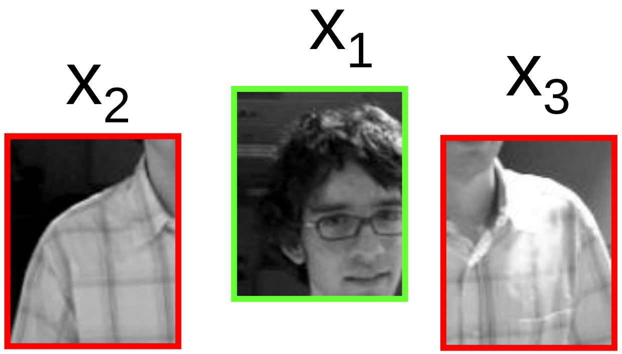




Background: Tracking by detection

- Use patches to train the classifier





$$\{(x_1, 1), (x_2, 0), (x_3, 0)\}$$

$$\downarrow$$

Classifier

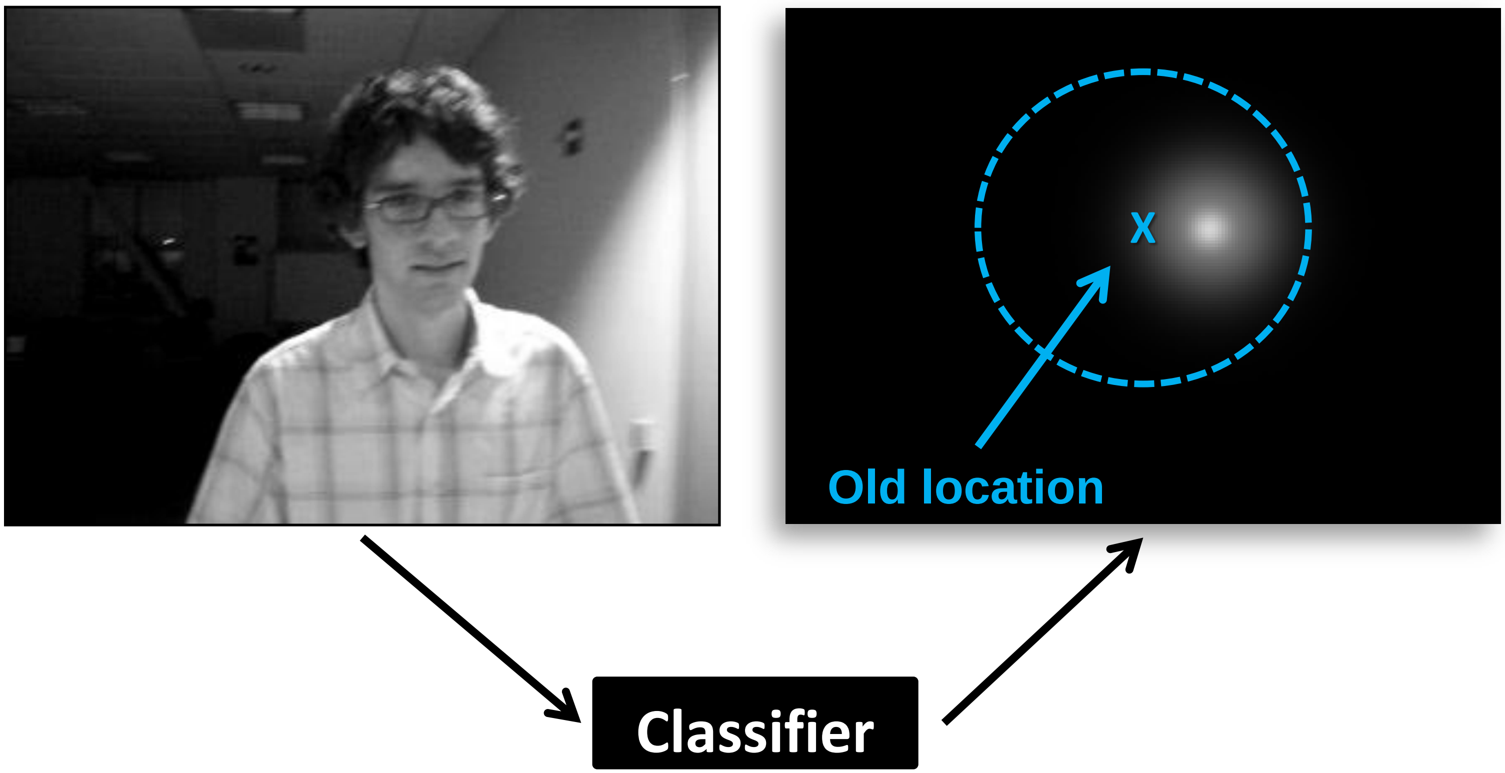
Background: Tracking by detection

- Frame 2 comes

Classifier

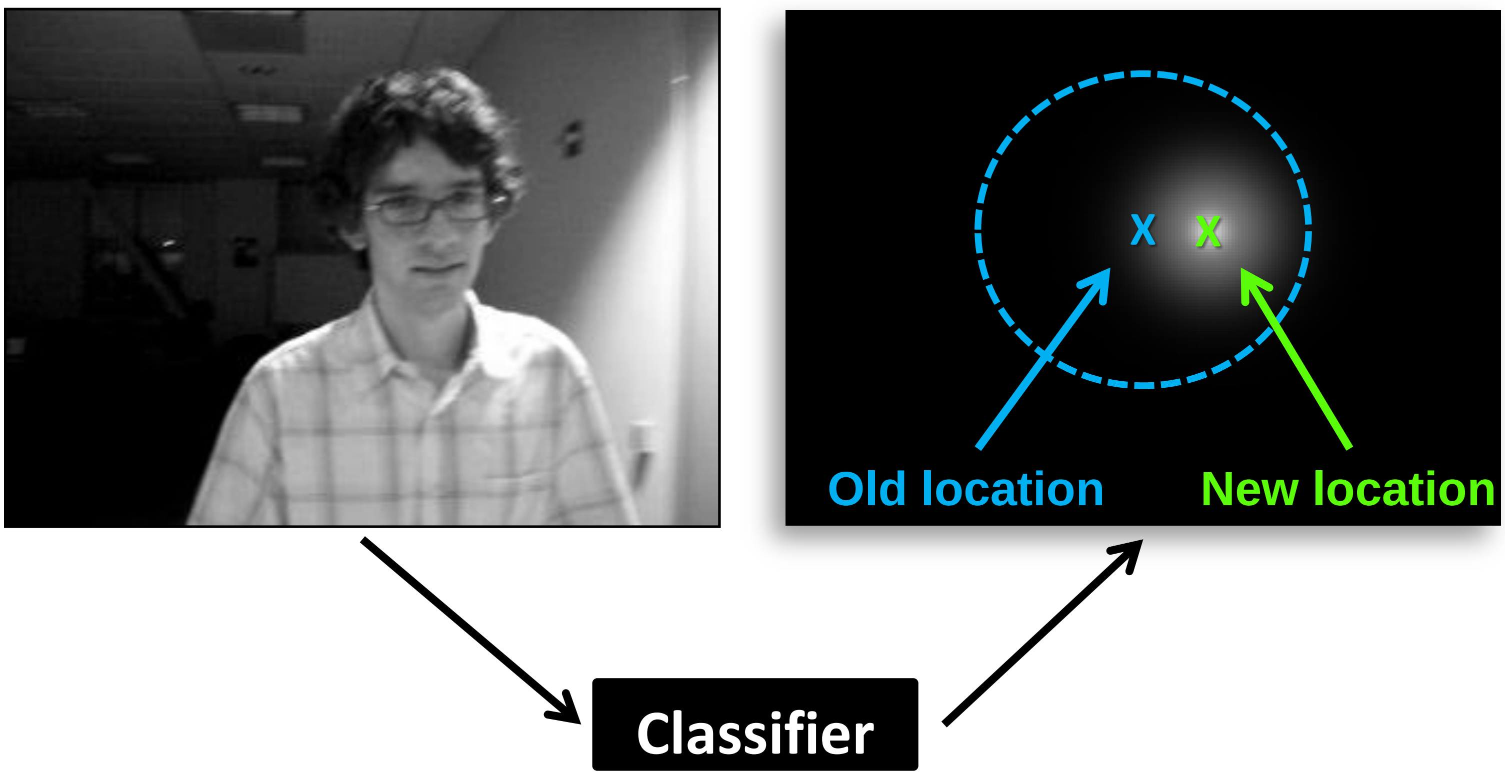
Background: Tracking by detection

- Calculate classifier response within a range of the old tracker location



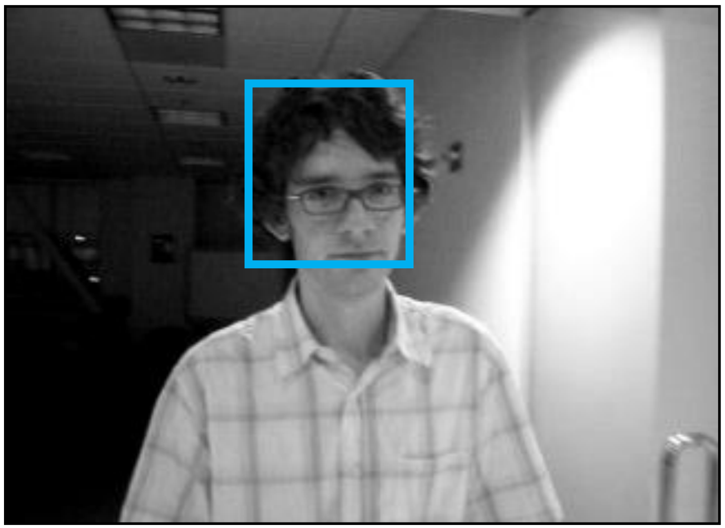
Background: Tracking by detection

- Find the maximum response location

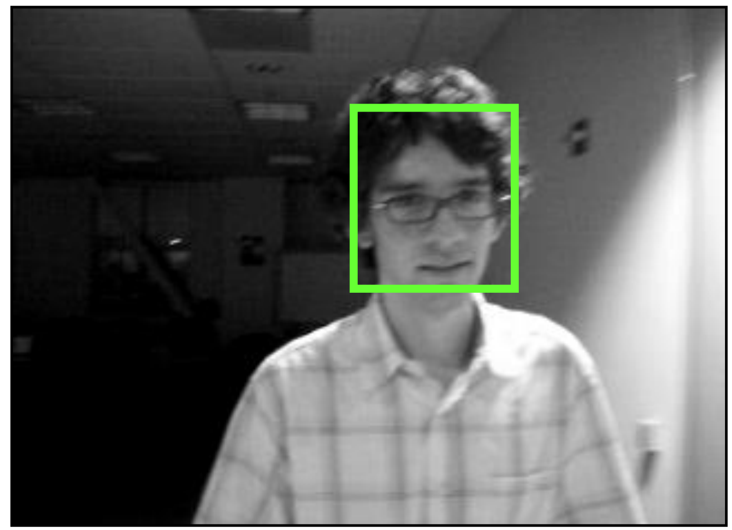


Background: Tracking by detection

- Move tracker



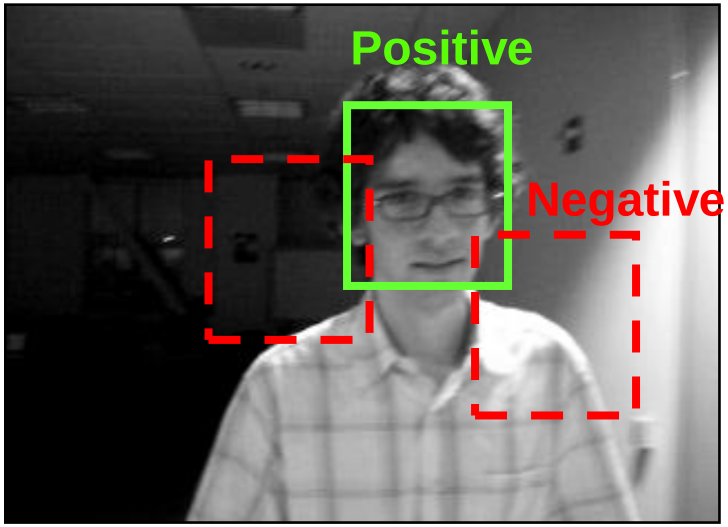
Frame 1



Frame 2

Background: Tracking by detection

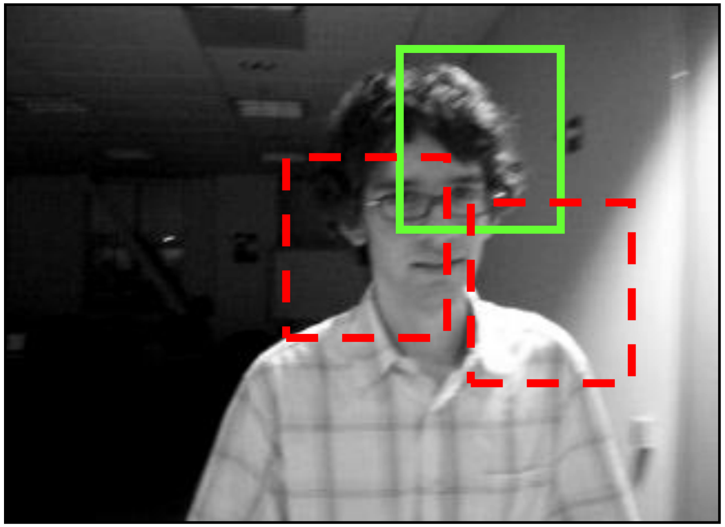
- Repeat



Frame 2

Background: Tracking by detection

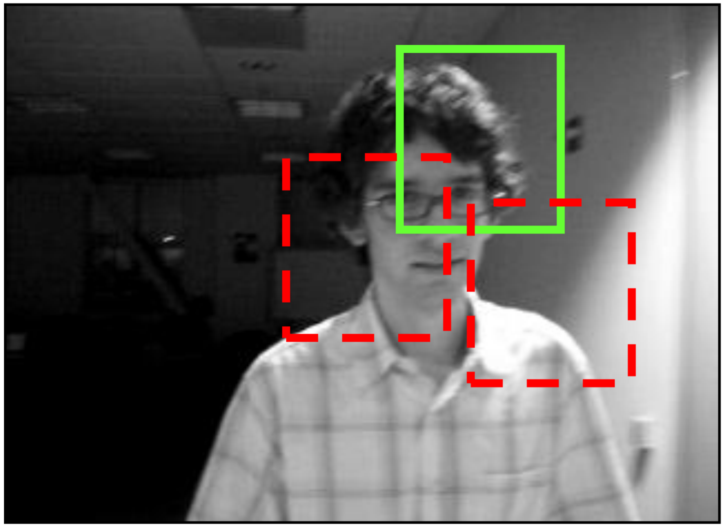
- Problem: If tracker location is not precise, might select bad training examples



Background: Tracking by detection

- Problem: If tracker location is not precise, might select bad training examples

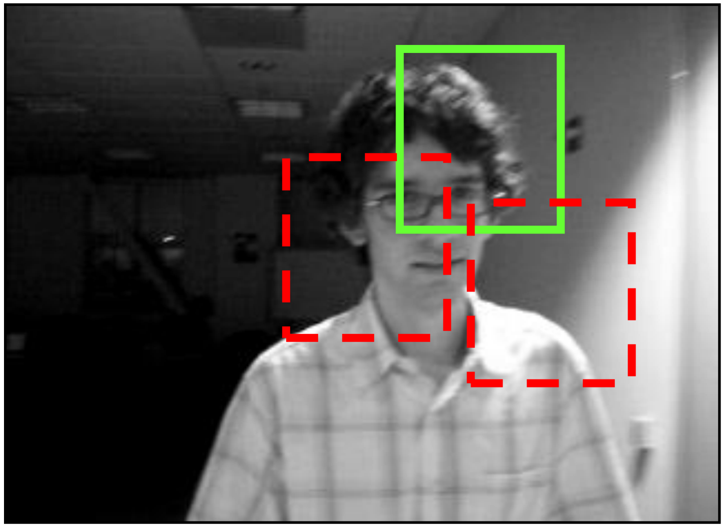
Model start to degrade!



Background: Tracking by detection

- Problem: If tracker location is not precise, might select bad training examples

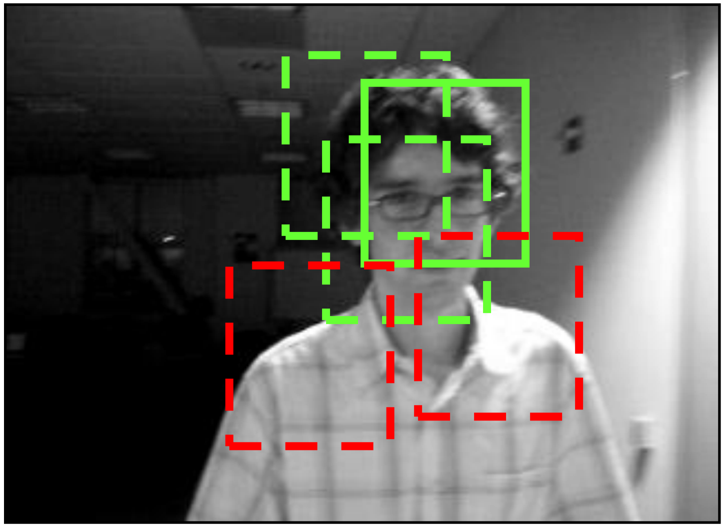
Model start to degrade!

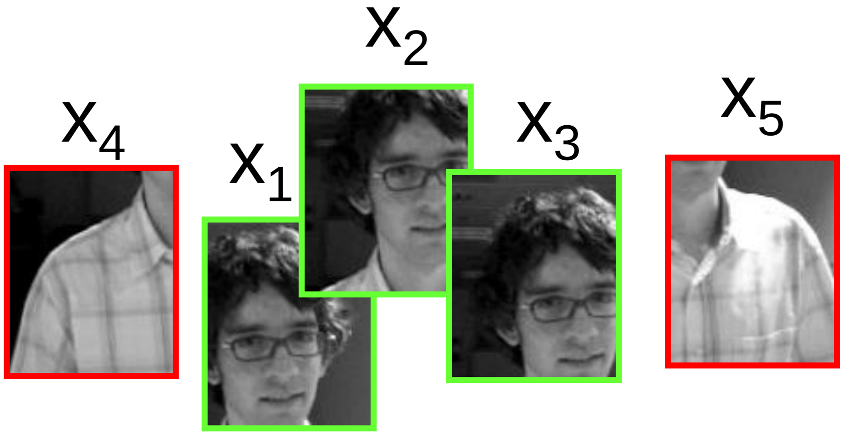


- How to select good training examples?

Previous Work

- Solution 1: multiple positive examples around tracker location



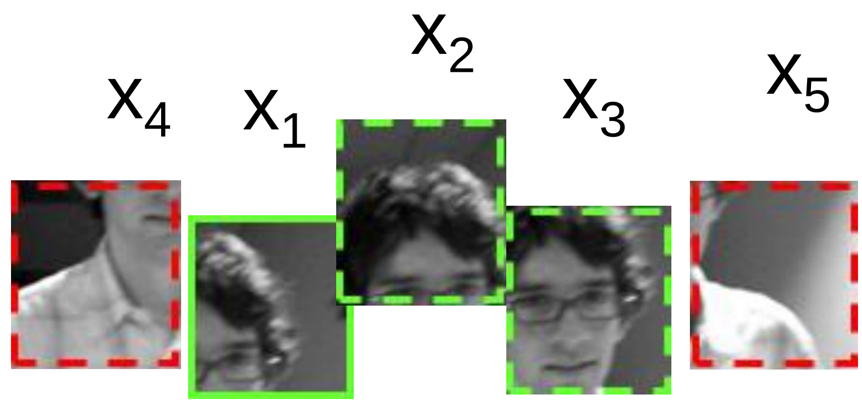


$\{(x_1, 1), (x_2, 1), (x_3, 1), (x_4, 0), (x_5, 0)\}$

Classifier

Previous Work

- Solution 1
  Might confuse classifier!



$\{(x_1, 1), (x_2, 1), (x_3, 1), (x_4, 0), (x_5, 0)\}$

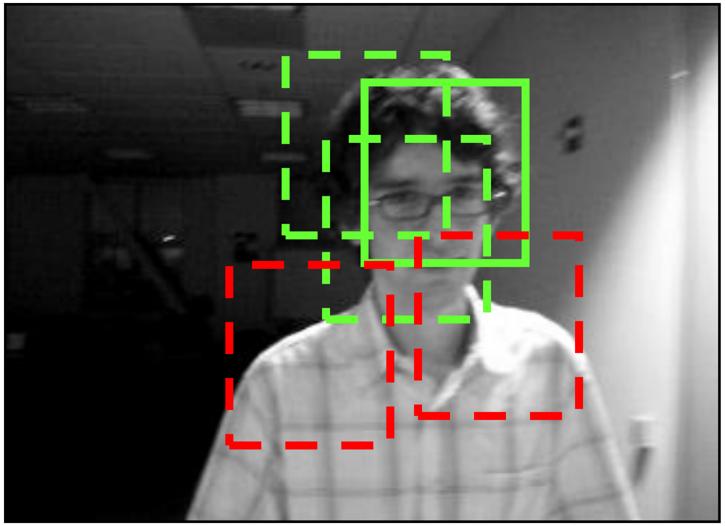
Classifier

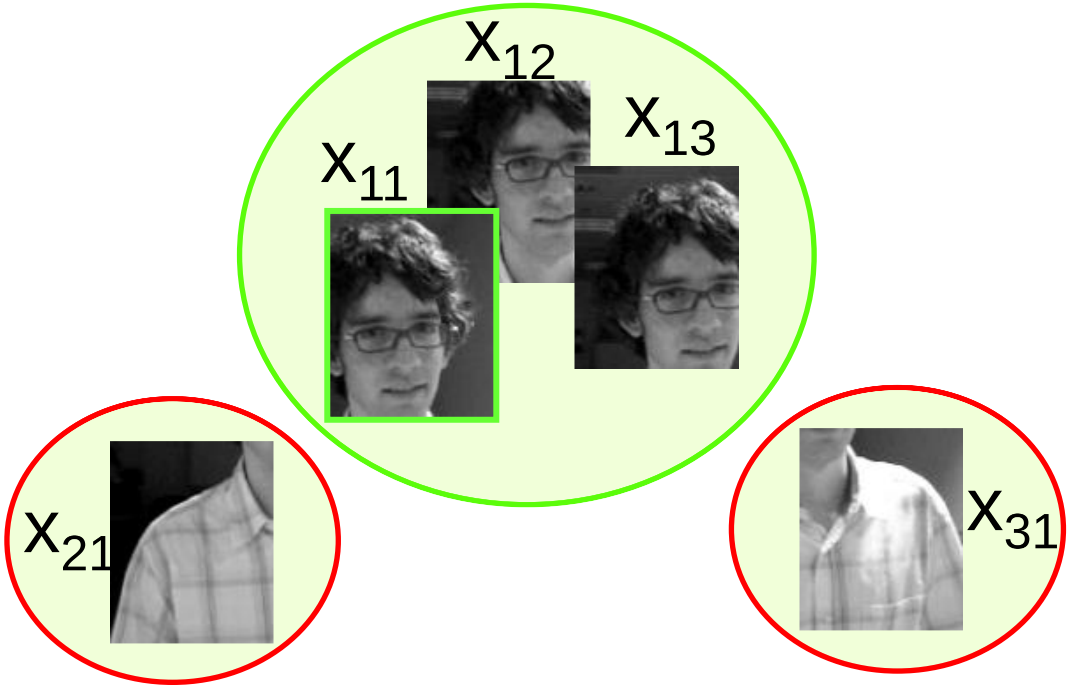
Previous Work

- Solution 2: Multiple Instance Learning (MIL)

[Keeler ‘90, Dietterich et al. ‘97]

Previous Work: Multiple Instance Learning

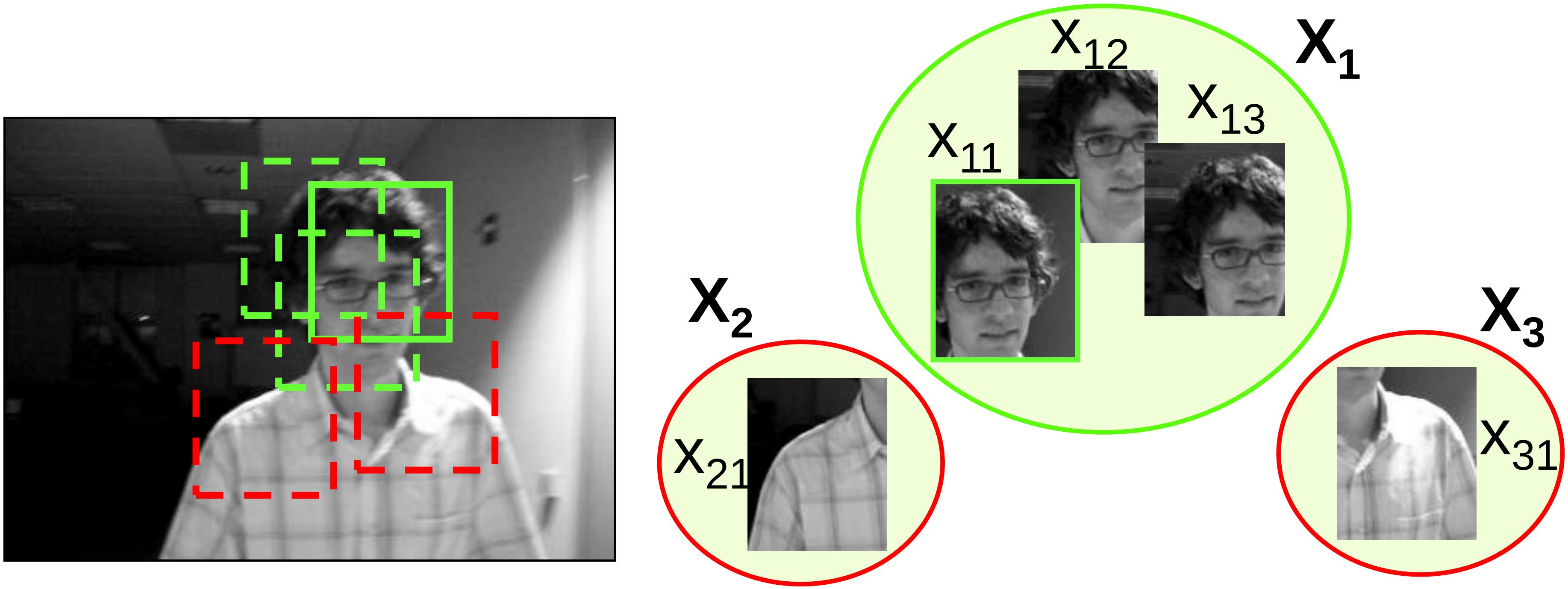




- Multiple examples in one bag

[Keeler '90, Dietterich et al. '97]

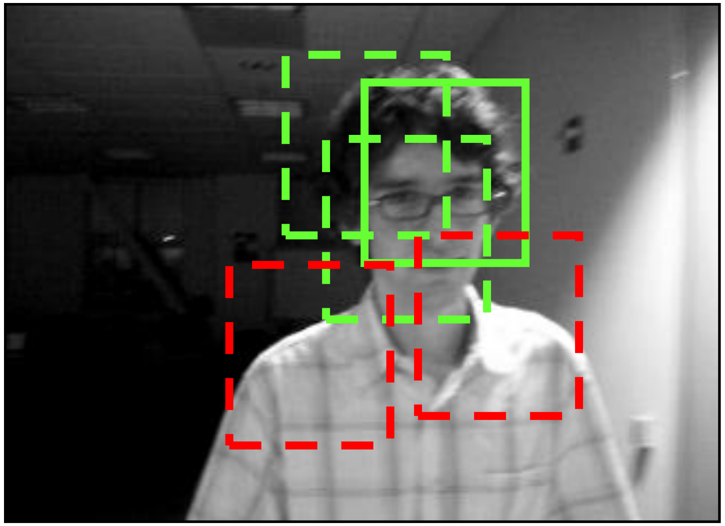
Previous Work: Multiple Instance Learning

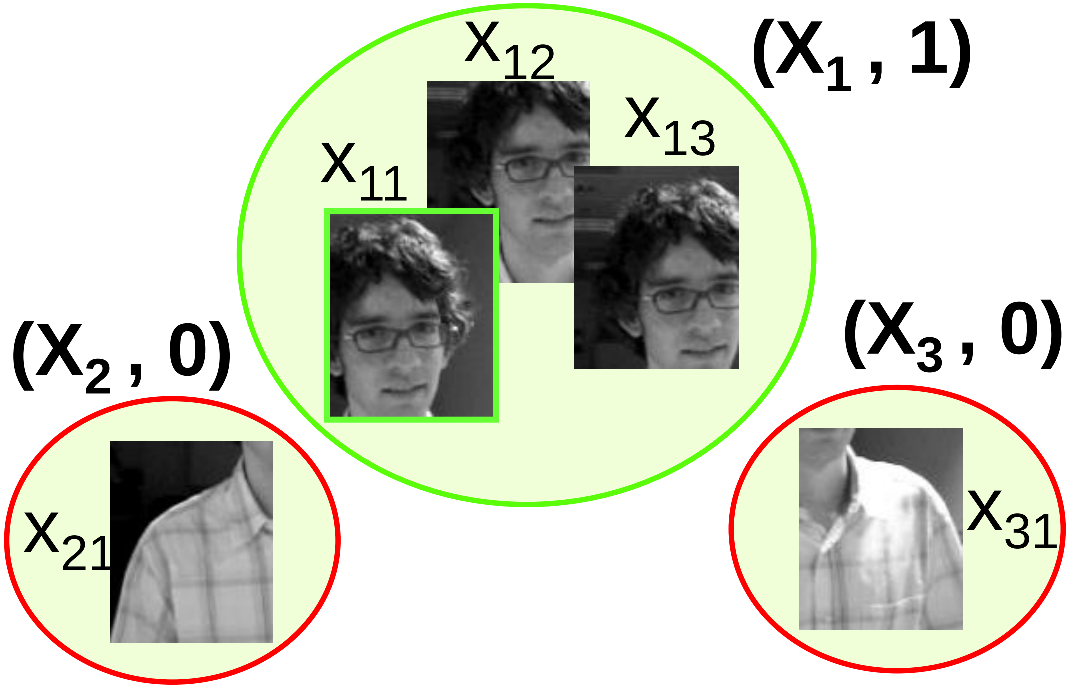


- Multiple examples in one bag

[Keeler '90, Dietterich et al. '97]

Previous Work: Multiple Instance Learning

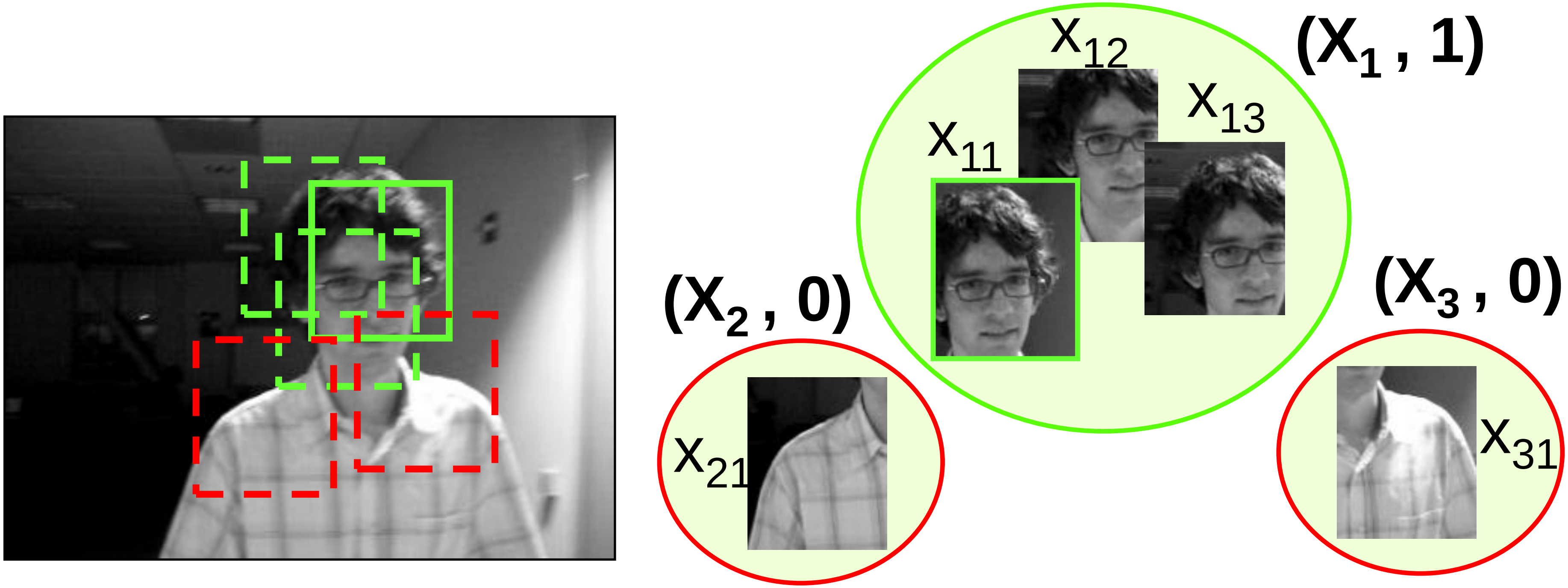




- Multiple examples in one bag
- One bag one label

[Keeler '90, Dietterich et al. '97]

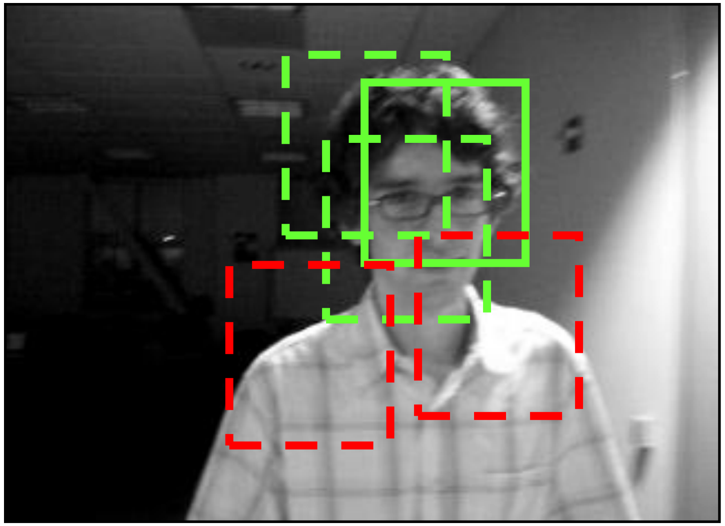
Previous Work: Multiple Instance Learning

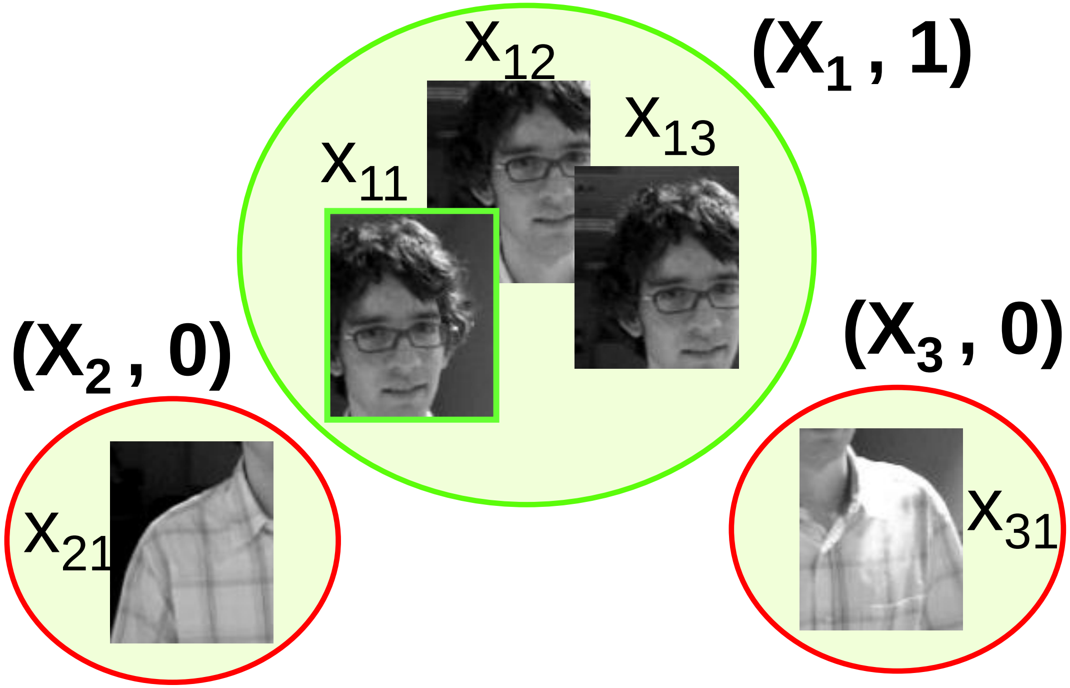


- Multiple examples in one bag
- One bag one label
- Bag **Positive** if at least one example is Positive

[Keeler '90, Dietterich et al. '97]

Previous Work: Multiple Instance Learning





$\{(X_1, 1), (X_2, 0), (X_3, 0)\}$

↓

Classifier

[Keeler ‘90, Dietterich et al. ‘97]

Previous Work: Multiple Instance Learning

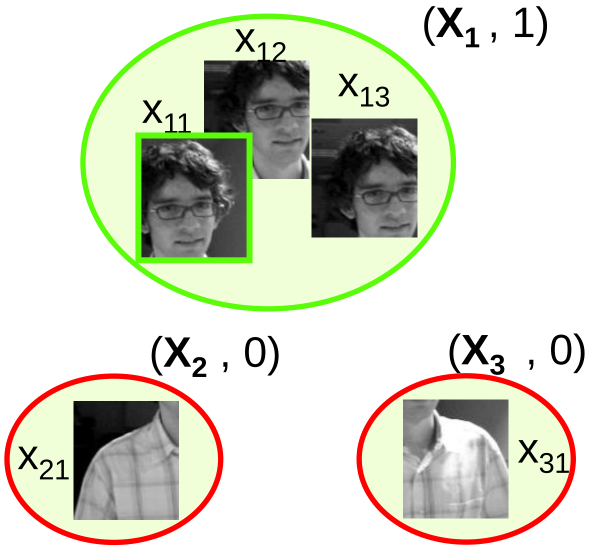
MIL training input:

$$\{(\boldsymbol{X_1}, y_1) \ldots (\boldsymbol{X_n}, y_n)\},$$

where,

$$\boldsymbol{X_i} = \{x_{i1} \ldots x_{im}\},$$

$$y_i = max_j \, y_{ij}$$



[Keeler ‘90, Dietterich et al. ‘97]

Previous Work: Multiple Instance Learning

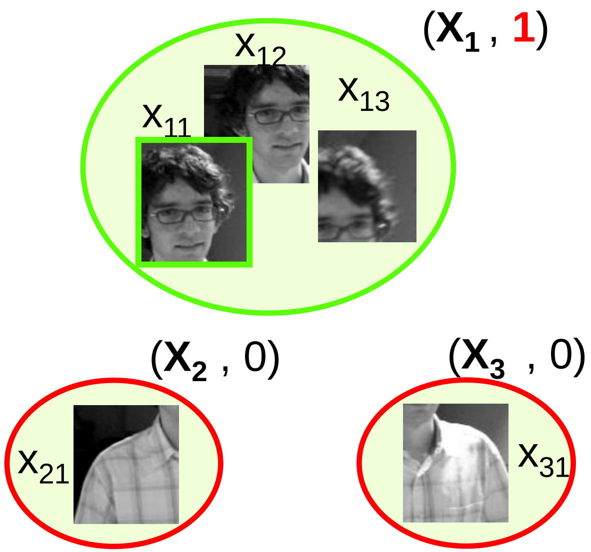
MIL training input:

$$\{(\boldsymbol{X_1}, y_1) \dots (\boldsymbol{X_n}, y_n)\},$$

where,

$$\boldsymbol{X_i} = \{x_{i1} \dots x_{im}\},$$

$$y_i = max_j \, y_{ij}$$



Bag babel is 1
if at least one instance is 1

[Keeler '90, Dietterich et al. '97]

Now we have training examples!

How to train the classifier?

Previous Work: MILBoost

- MIL + boosting

Train a boosting classifier that maximizes log likelihood of bags

$$logL = \sum_i \log(p(y_i \mid \boldsymbol{X_i}))$$

where,

$$p(y_i \mid \boldsymbol{X_i}) = 1 - \prod_j (1 - p(y_i \mid x_{ij}))$$

[Viola et al. ‘05]

Previous Work: MILBoost

- MIL + boosting

Train a boosting classifier that maximizes log likelihood of bags

$$logL = \sum_i \log(p(y_i \mid \boldsymbol{X_i}))$$

where,

$$\underset{\sim 1}{\underset{\uparrow}{p(y_i \mid \boldsymbol{X_i})}} = 1 - \prod_j (1 - \underset{1}{\underset{\uparrow}{p(y_i \mid x_{ij})}})$$

[Viola et al. ‘05]

Previous Work: MILBoost

- Problem: need all training examples

[Viola et al. ‘05]

Previous Work: MILBoost

But in tracking, only current frame available

[Viola et al. ‘05]

Previous Work: MILBoost

But in tracking, only current frame available

Need an online training algorithm for MIL

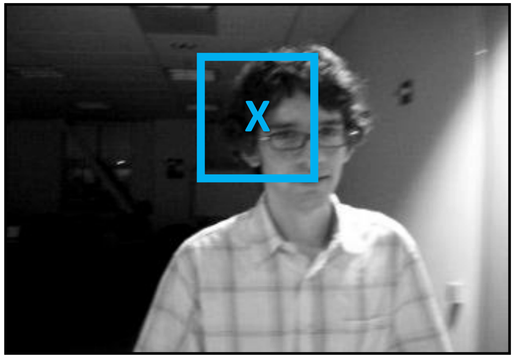
[Viola et al. '05]

Main Contribution of this paper

- Online-MILBoost:
  Online training for MIL-based classifier

- MILTrack
  New tracking solution using Online-MILBoost

MILTrack workflow



New frame comes in:

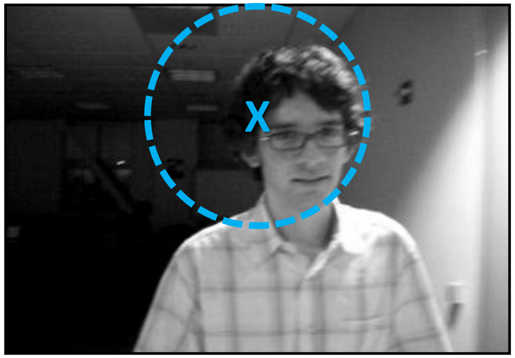
1. Crop out a set of image patches

$$X^s = \{x| < s^1 \; pixels \; from \; tracker \; location\}$$

1: s = 35 in authors' experiment

Babenko et al., 09

MILTrack workflow



New frame comes in:

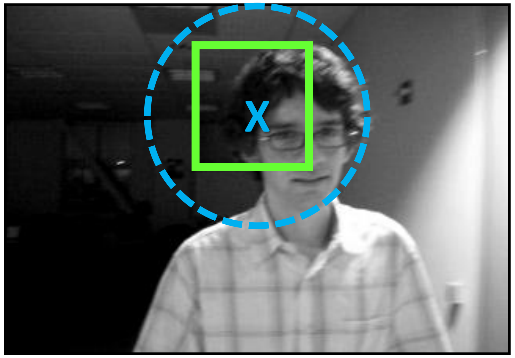
1. Crop out a set of image patches

$$X^s = \{x \mid < s^1 \; pixels \; from \; tracker \; location\}$$

1: s = 35 in authors' experiment

Babenko et al., 09

MILTrack workflow



New frame comes in:

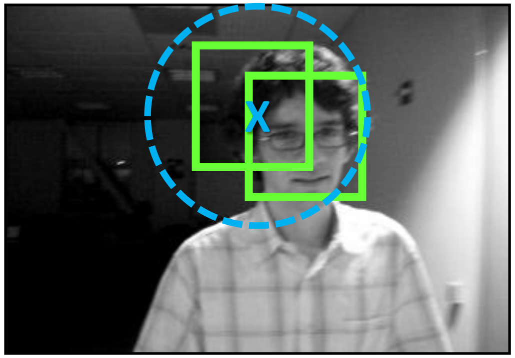
1. Crop out a set of image patches

$$X^s = \{x| < s^1 \; pixels \; from \; tracker \; location\}$$

1: s = 35 in authors' experiment

Babenko et al., 09

MILTrack workflow



New frame comes in:

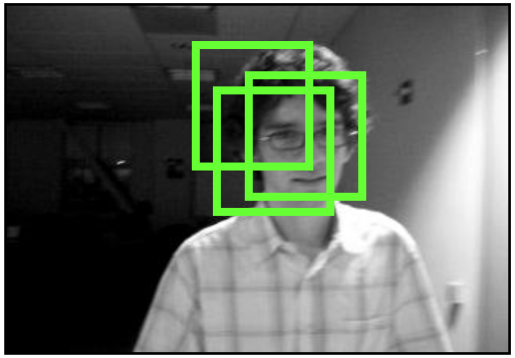
1. Crop out a set of image patches

$$X^s = \{x| < s^1 \, pixels \; from \; tracker \; location\}$$

1: s = 35 in authors' experiment

Babenko et al., 09

MILTrack workflow



New frame comes in:

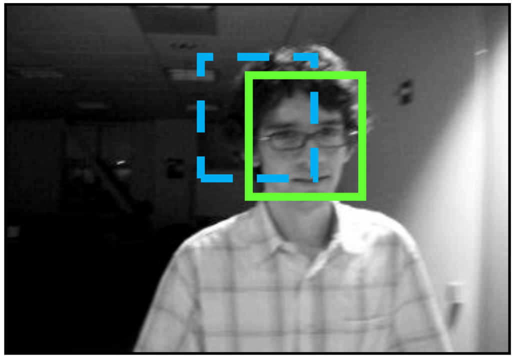
1. Crop out a set of image patches

$$X^s = \{x| < s^1\ pixels\ from\ tracker\ location\}$$

1: s = 35 in authors' experiment

Babenko et al., 09

MILTrack workflow



New frame comes in:

1. Crop out a set of image patches
$$X^s = \{x| < \ s\, pixels\ from\ tracker\ location\}$$

2. Use MIL classifier to find new tracker location
$$l_{new} = l(\underset{x \in X^s}{\operatorname{argmax}}\ p(y = 1|x))$$

Babenko et al., 09

MILTrack workflow

New frame comes in:

1. Crop out a set of image patches
$$X^s = \{x| < \ s\, pixels\ from\ tracker\ location\}$$

2. Use MIL classifier to find new tracker location
$$l_{new} = l(\underset{x \in X^s}{\operatorname{argmax}}\ p(y = 1|x))$$

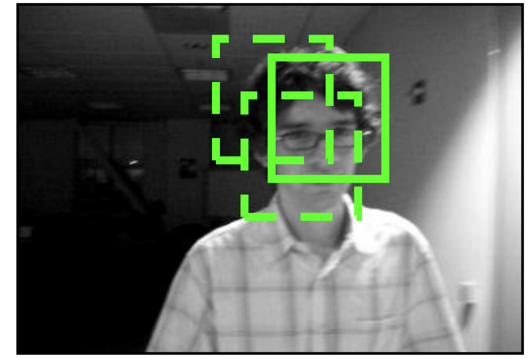
3. 1) Crop positive examples
$$X^r = \{x| < r^1\, pixels\ from\ tracker\ location\}$$

1: r = 5 in authors' experiment

Babenko et al., 09

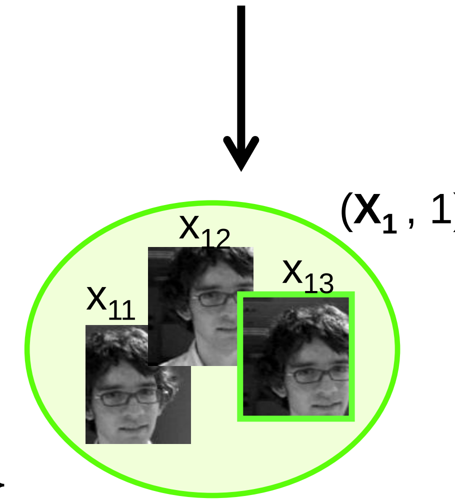
MILTrack workflow

New frame comes in:



1. Crop out a set of image patches

$$X^s = \{x | < \; s \; pixels \; from \; tracker \; location\}$$



2. Use MIL classifier to find new tracker location

$$l_{new} = l(\operatorname*{argmax}_{x \in X^s} p(y = 1 | x))$$

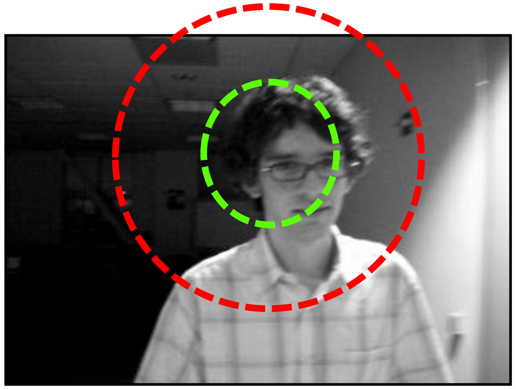
3. 1) Crop positive examples

$$X^r = \{x | < r^1 \; pixels \; from \; tracker \; location\}$$

1: r = 5 in authors' experiment

Babenko et al., 09

MILTrack workflow



New frame comes in:

1. Crop out a set of image patches

$$X^s = \{x| <\ s\ pixels\ from\ tracker\ location\}$$

2. Use MIL classifier to find new tracker location

$$l_{new} = l(\underset{x \in X^s}{\operatorname{argmax}}\ p(y = 1|x))$$

3. 1) Crop positive examples

$$X^r = \{x| < r\ pixels\ from\ tracker\ location\}$$

3. 2) Crop Negative examples

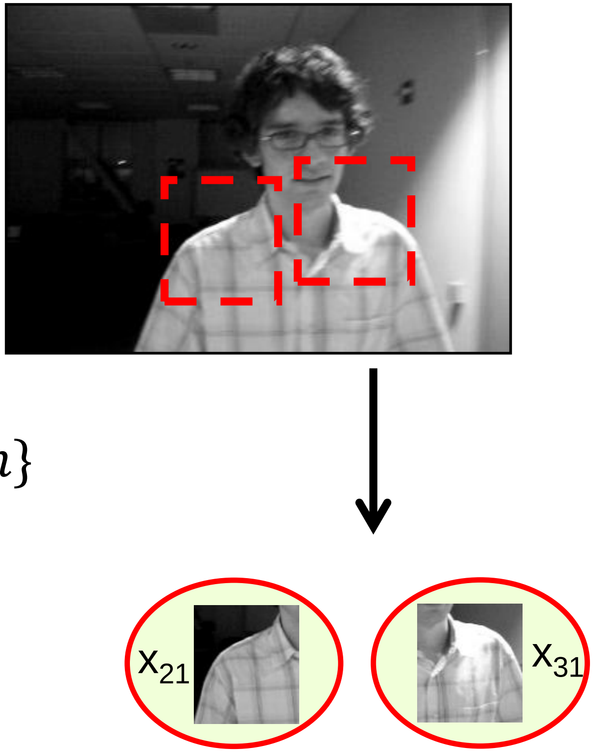
$$X^{r,\beta} = \{x|\ r\ to\ \beta^1\ pixels\ away\ from\ tracker\ location\}$$

1: β = 50 in authors' experiment

Babenko et al., 09

MILTrack workflow

New frame comes in:



$(\mathbf{X_2}, 0)$ $(\mathbf{X_3}, 0)$

1. Crop out a set of image patches

$$X^s = \{x| < \ s\ pixels\ from\ tracker\ location\}$$

2. Use MIL classifier to find new tracker location

$$l_{new} = l(\underset{x \in X^s}{\operatorname{argmax}}\ p(y = 1|x))$$

3. 1) Crop positive examples

$$X^r = \{x| < r\ pixels\ from\ tracker\ location\}$$

3. 2) Crop Negative examples

$$X^{r,\beta} = \{x|\ r\ to\ \beta^1\ pixels\ away\ from\ tracker\ location\}$$

1: β = 50 in authors' experiment

Babenko et al., 09

MILTrack workflow

New frame comes in:

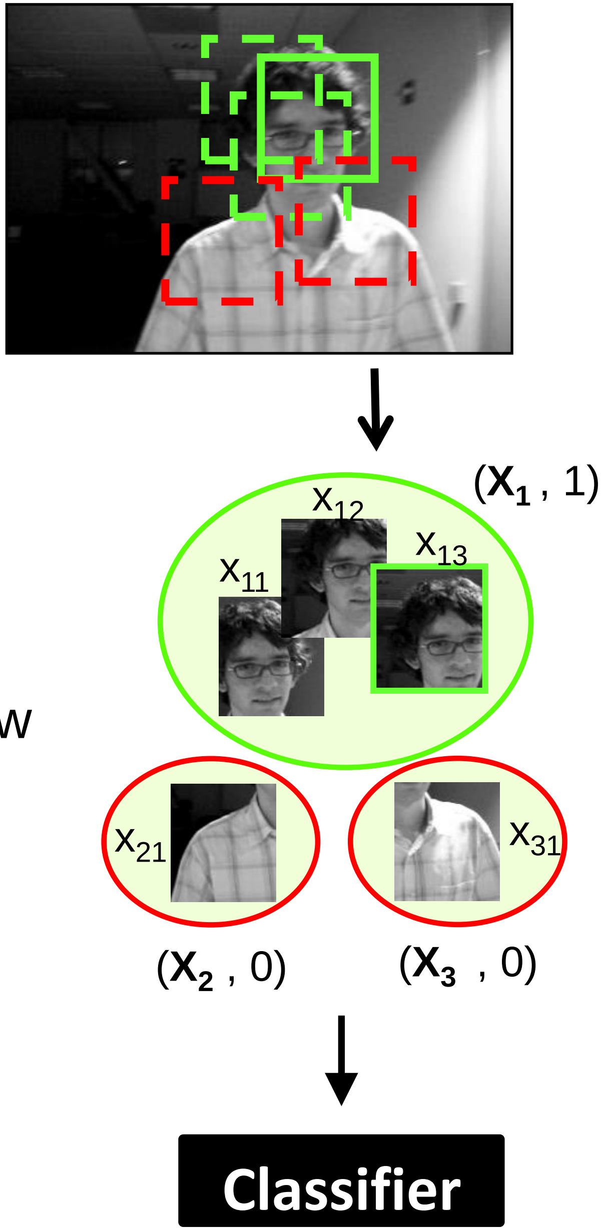
1. Crop out a set of image patches
   $$X^s = \{x| < \ s\ pixels\ from\ tracker\ location\}$$

2. Use MIL classifier to find new tracker location
   $$l_{new} = l(\underset{x \in X^s}{\text{argmax}}\ p(y = 1|x))$$

3. Crop positive and negative examples near new object location

4. Online MILBoost:

Update MIL classifier with positive and negative example bags



Babenko et al., 09

Online-MILBoost:

Image patch x

f_1
f_2
f_3
…

Babenko et al., 09

Online-MILBoost:

Image patch x

- h_k: a weak classifier using one feature

	f_1
$h_1(x)$	f_2
	f_3
	…
$h_2(x)$	f_9
	…

Babenko et al., 09

Online-MILBoost:

Image patch x

- h_k: a weak classifier using one feature

$$h_k(x) = \log\left[\frac{p(y=1|f_k(x))}{p(y=0|f_k(x))}\right]$$

	f_1
	f_2
	f_3
	…
$h_k(x)$	f_t
	…

Babenko et al., 09

Online-MILBoost:

Image patch x

- h_k: a weak classifier using one feature

$$h_k(x) = \log\left[\frac{p(y=1|f_k(x))}{p(y=0|f_k(x))}\right]$$

with,

$$p(f_k(x)|y=1) \sim \mathcal{N}(\mu_1, \sigma_1)$$
$$p(f_k(x)|y=0) \sim \mathcal{N}(\mu_0, \sigma_0)$$
$$p(y=1) = p(y=0)$$

	f_1
	f_2
	f_3
	…
$h_k(x)$	f_t
	…

Babenko et al., 09

Online-MILBoost:

Image patch x

- $\boldsymbol{H}(\boldsymbol{x})$: the MIL classifier made from weak classifiers

$$\boldsymbol{H}(\boldsymbol{x}) = \sum_{k=1}^{K} h_k(x)$$

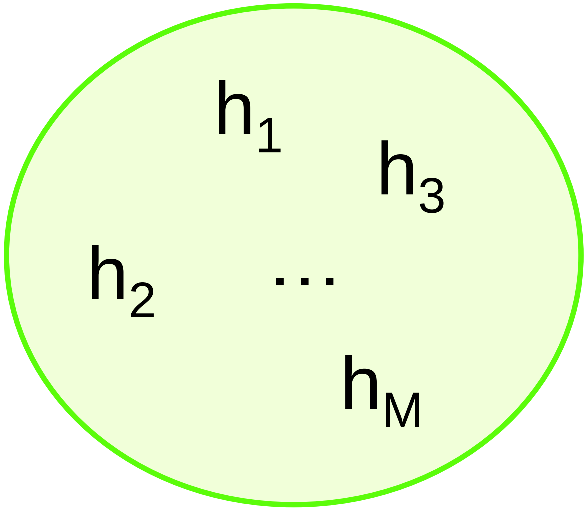
	f_1
$h_1(x)$	f_2
	f_3
...	...
$h_k(x)$	f_t
...	...

K = 50 in authors' experiment

Babenko et al., 09

Online-MILBoost:

- Always keep a pool of M >> K weak classifier candidates

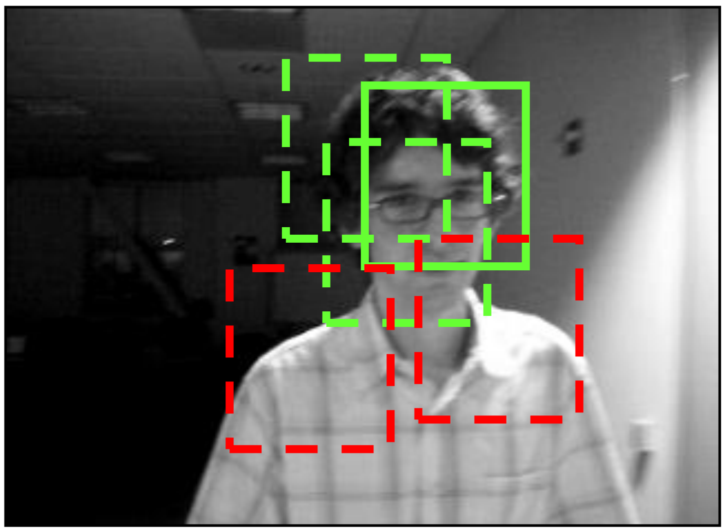


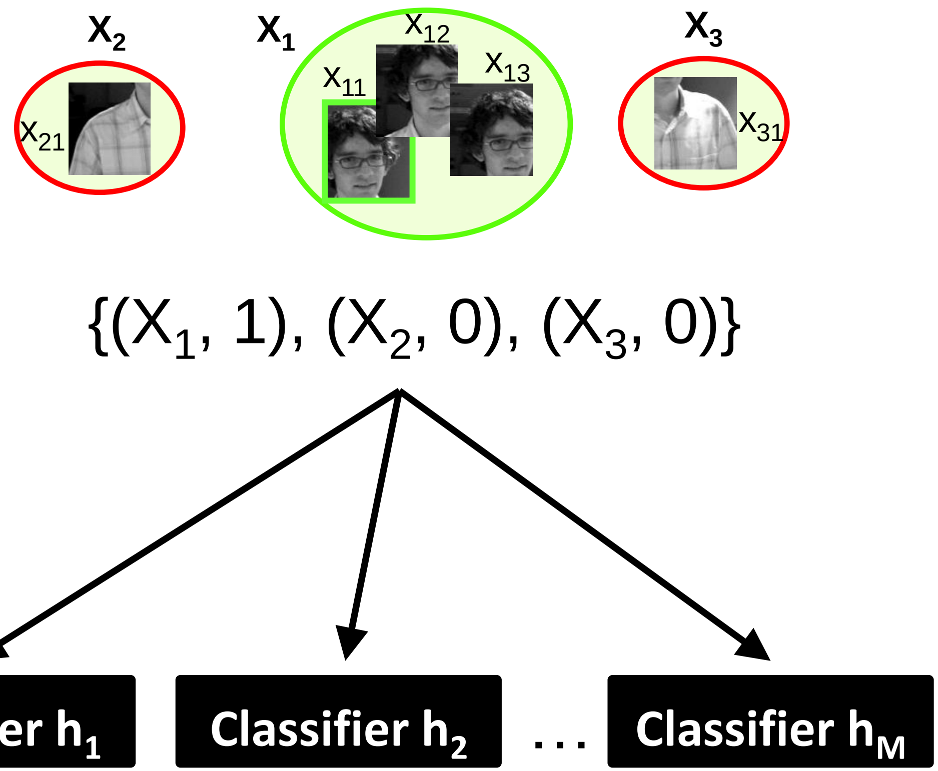
M = 250 & K = 50 in authors' experiment

Babenko et al., 09

Online-MILBoost:

- Update all M weak classifiers with positive and negative bags





Babenko et al., 09

Online-MILBoost:

- Pick best K weak classifiers to form $\boldsymbol{H}(\boldsymbol{x})$, where

$$h_k = \underset{h \in \{h_1 \dots h_M\}}{\operatorname{argmax}} \log L(H_{k-1} + h)$$

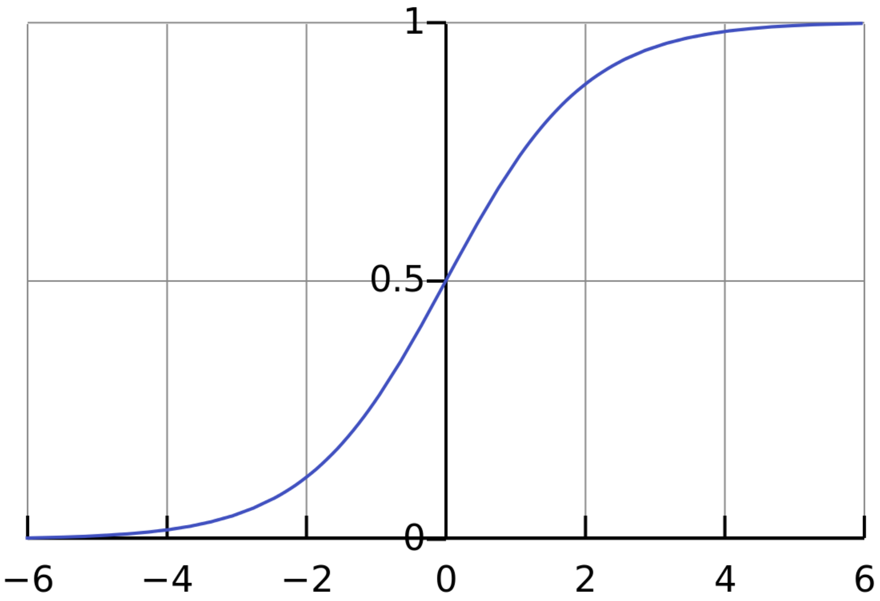
$$\boldsymbol{H}(\boldsymbol{x}) = \sum_{k=1}^{K} h_k(x)$$

where H_{k-1} is the classifier made up of the first $k-1$ weak classifiers

Babenko et al., 09

Online-MILBoost:

- Prediction : $p(y = 1|x) = \sigma(\boldsymbol{H}(\boldsymbol{x}))$

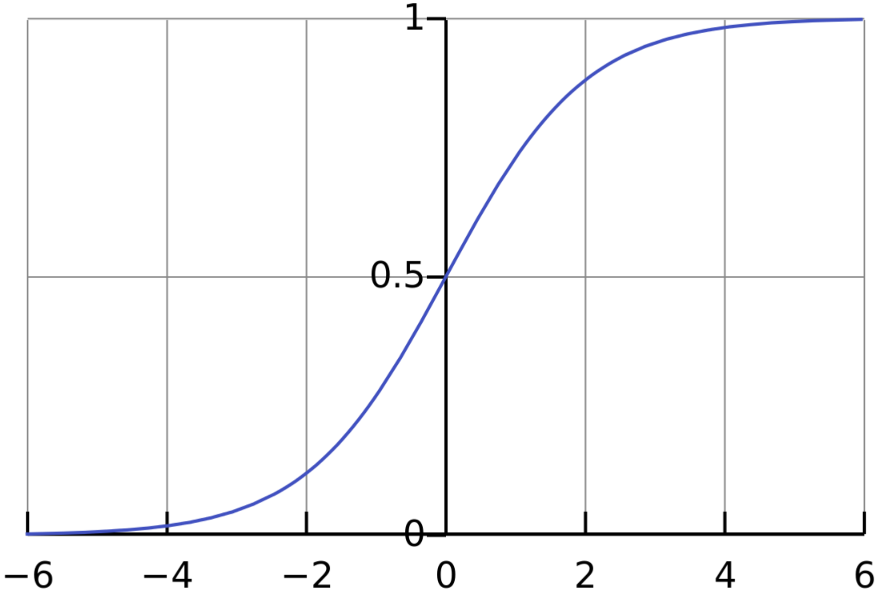


$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Babenko et al., 09

Online-MILBoost:

- Prediction : $p(y=1|x) = \sigma(\boldsymbol{H}(\boldsymbol{x}))$



$$\sigma(x) = \frac{1}{1+e^{-x}}$$

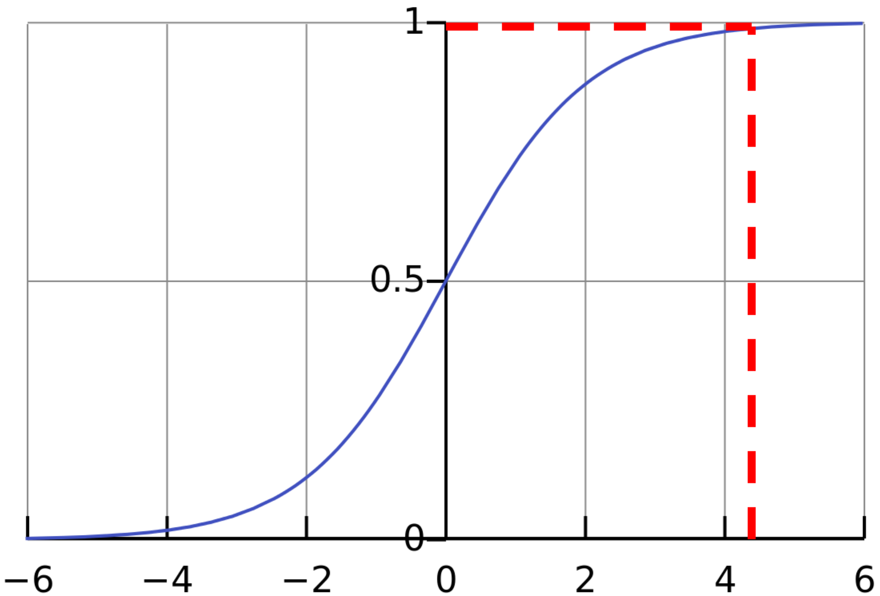
$h_1(x)$	f_1
$h_2(x)$	f_2
$h_3(x)$	f_3

$$h_1(x) = 2,\ \ h_2(x) = 1.8,\ \ h_3(x) = 0.6$$

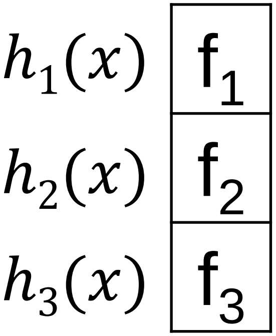
Babenko et al., 09

Online-MILBoost:

- Prediction : $p(y = 1|x) = \sigma(\boldsymbol{H}(\boldsymbol{x}))$



$$\sigma(t) = \frac{1}{1 + e^{-t}}$$



$$h_1(x) = 2,\ \ h_2(x) = 1.8,\ \ h_3(x) = 0.6$$

$$\boldsymbol{H}(\boldsymbol{x}) = \sum_{k=1}^{K} h_k(x) = 4.4$$

$$p(y = 1|x) = \sigma\big(\boldsymbol{H}(\boldsymbol{x})\big) = 0.99$$

Babenko et al., 09

MILTrack workflow

New frame comes in:

1. Crop out a set of image patches

$$X^s = \{x| < \ s\ pixels\ from\ tracker\ location\}$$

2. **Use MIL classifier to find new tracker location**

$$l_{new} = l(\underset{x \in X^s}{\operatorname{argmax}}\ p(y = 1|x))$$

3. Crop positive and negative examples near new object location

4. **Online MILBoost:**
   Update MIL classifier with positive and negative example bags

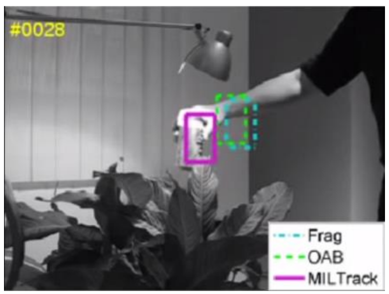
Babenko et al., 09

Experiments

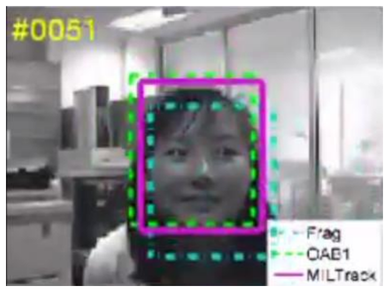
Datesets: 8 publicly available videos,

- Grayscale, 320 x 240 pixels
- Ground truth labeled every 5 frames by hand

Coke can



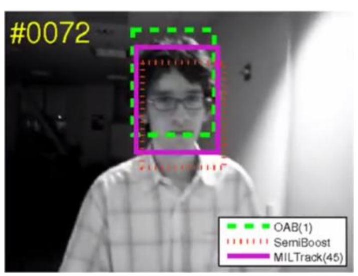
Girl



Occluded face

Occluded face 2

David



Sylvester

Tiger 1

Tiger 2

Babenko et al., 2009

Experiments

Compared with:

- **OAB1**
  Online AdaBoost w/ 1 positive example per frame

- **OAB5**
  Online AdaBoost w/ 45 positive examples per frame

- **SemiBoost**
  Label in 1st frame only.

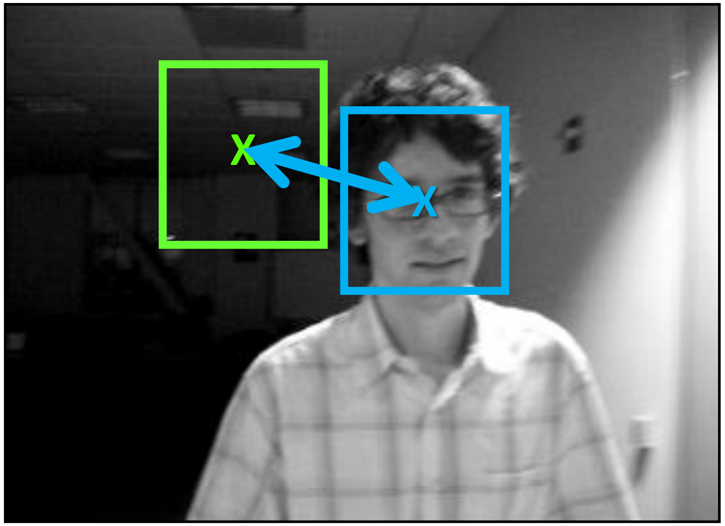
- **FragTrack**
  Static appearance model

Babenko et al., 2009

Experiments

Evaluation criterion:

Tracker position error (pixels) w.r.t. Ground truth



Babenko et al., 2009

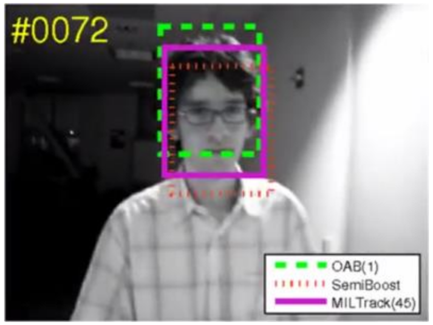
Results

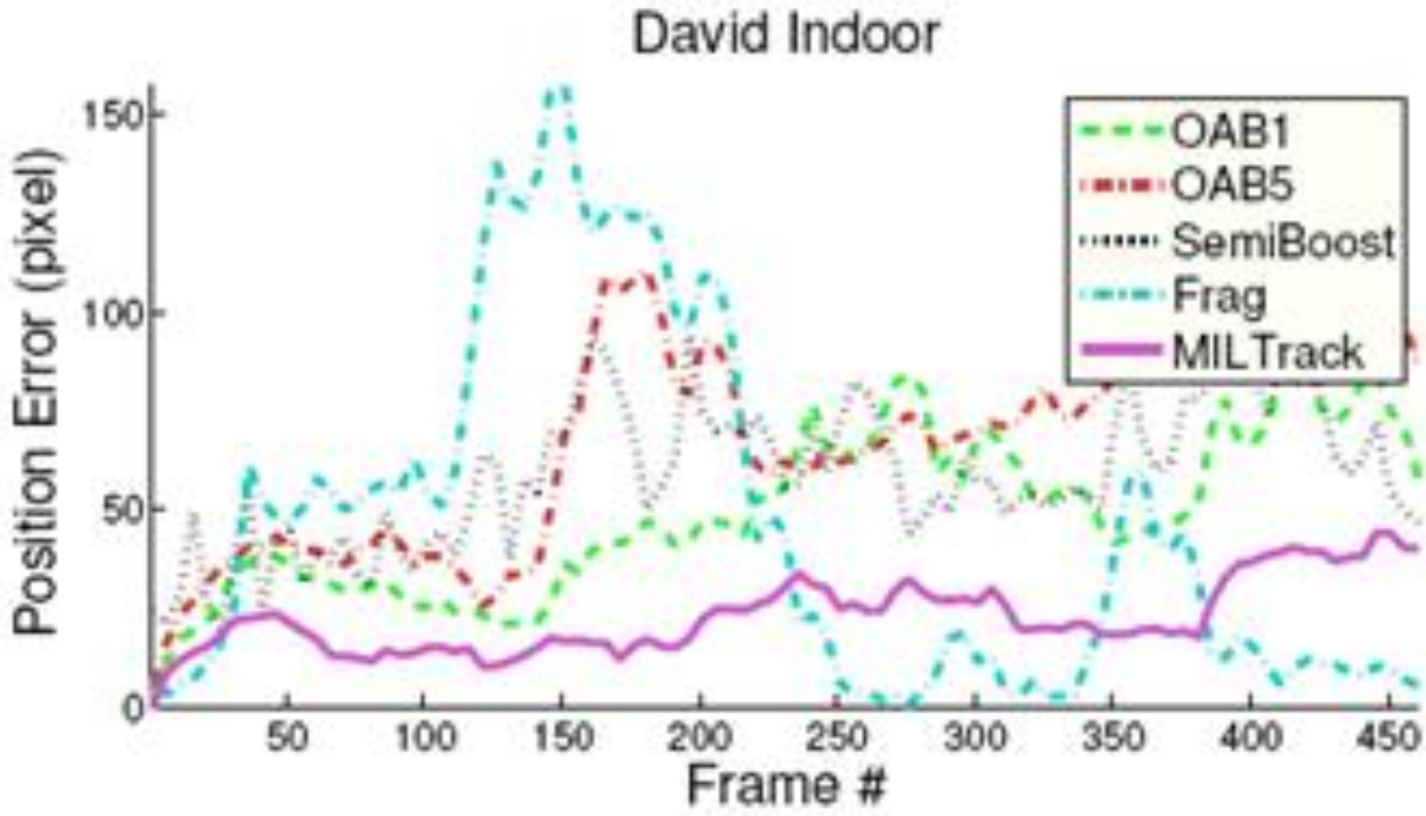
[Video David]

Babenko et al., 09

Results

Position error versus Frame #, Video David





Babenko et al., 2009

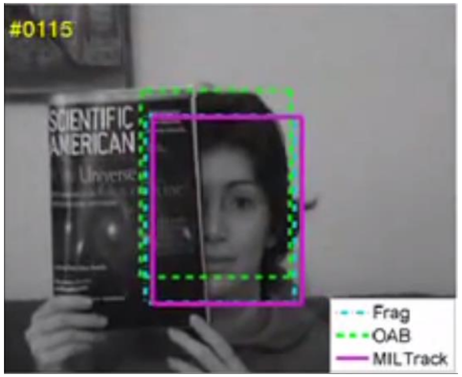
Results

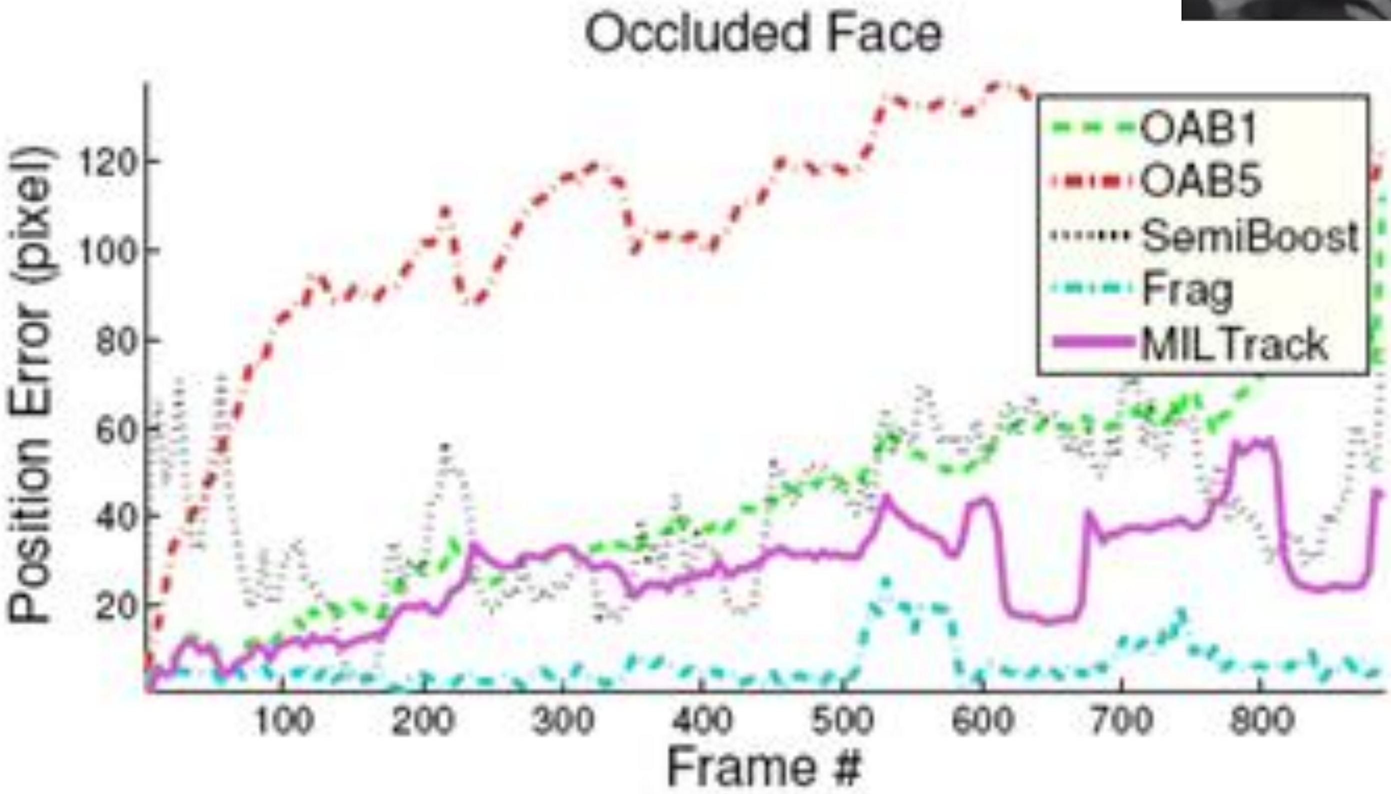
Video Occluded Face

Babenko et al., 09

Results

Position error versus Frame #,
Video Occluded Face





Babenko et al., 2009

Results

Average Center location errors (pixels)

Best
Second Best

Video Clip	OAB1	OAB5	SemiBoost	Frag	MILTrack
David Indoor	49	72	59	**46**	**23**
Sylvester	25	79	22	**11**	**11**
Occluded Face	44	105	41	**6**	**27**
Occluded Face 2	**21**	93	43	45	**20**
Girl	48	68	52	**27**	**32**
Tiger 1	**35**	58	46	40	**15**
Tiger 2	34	**33**	53	38	**17**
Coke Can	**25**	57	85	63	**21**

Babenko et al., 2009

Conclusion

- Online MILBoost:
  Online algorithm to update MIL-based classifier

- Performance of “MILTrack” is stable

Babenko et al., 2009

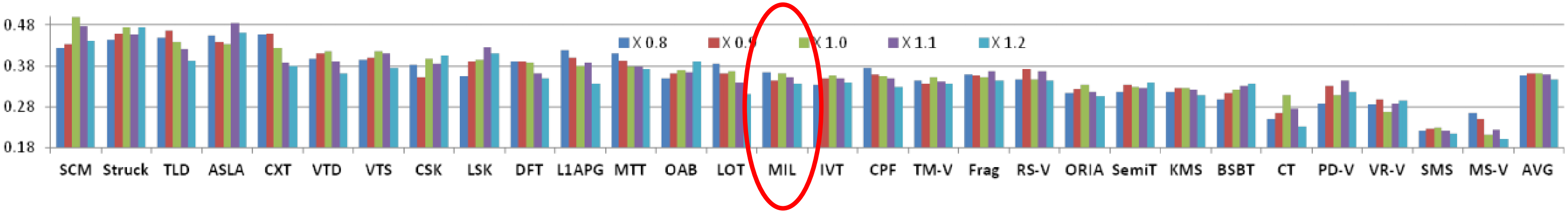
Discussion

- Why it can handle occlusion?

- Possible improvements
  - Motion Model
  - Features
  - Part based representation

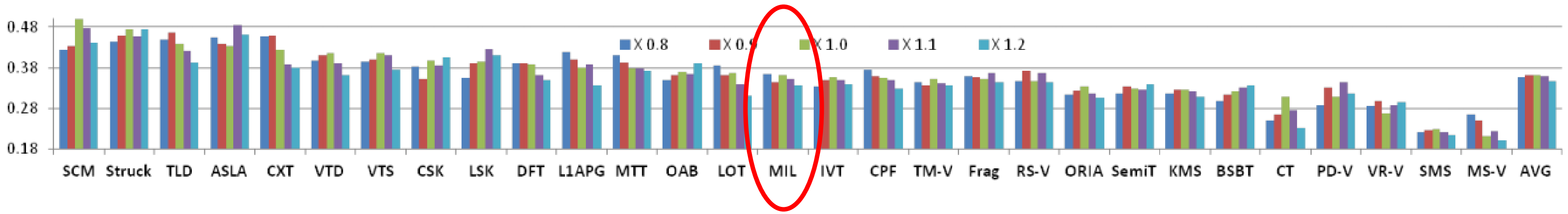
Babenko et al., 2009

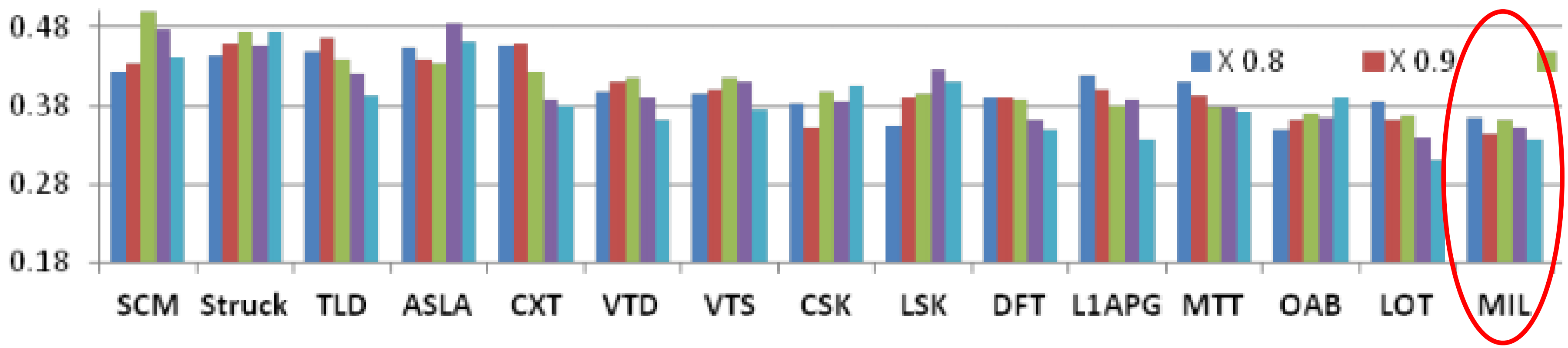
Note…



Wu et al., 2013

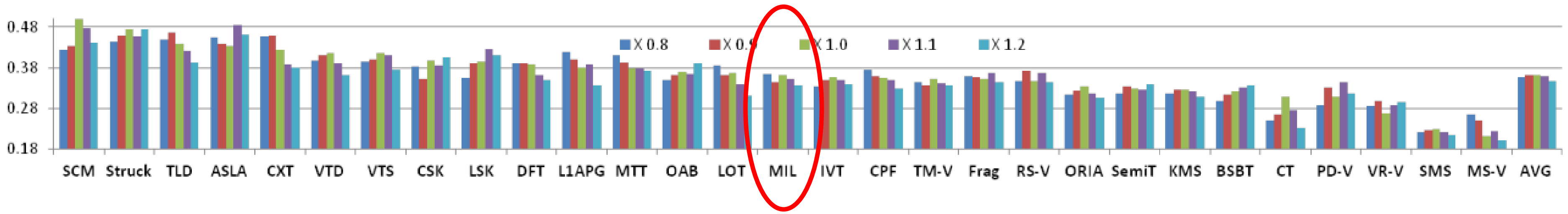
Note…

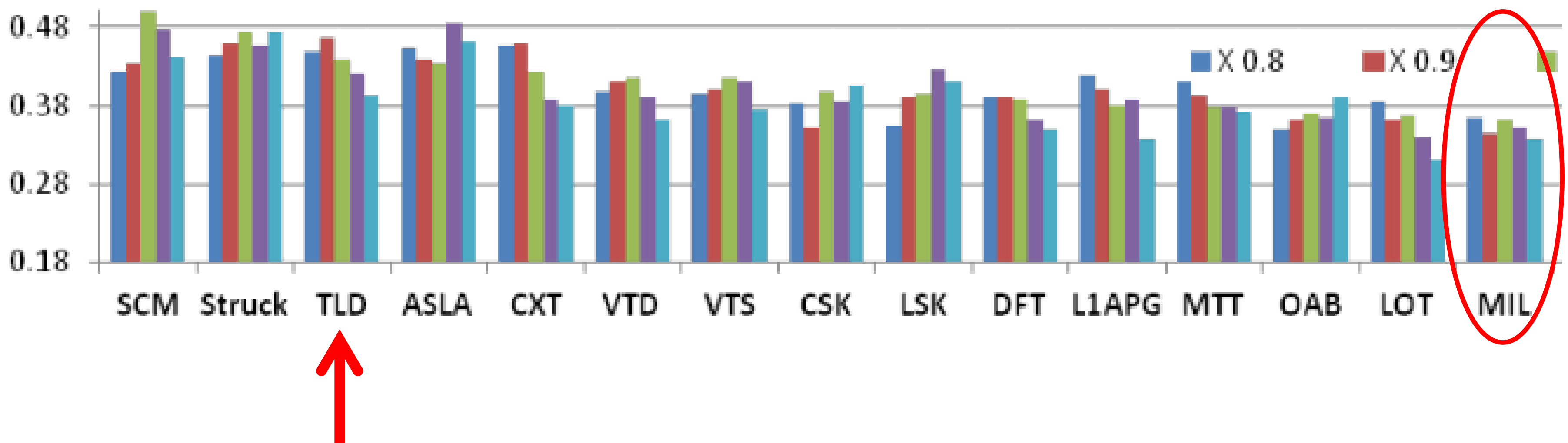




Wu et al., 2013

Note…





Project 2 use this!

Wu et al., 2013

Thank You!
Q&A