Learning in Location Space

A New Framework for Object Detection

Damian Eads[1,3], David Helmbold[1], and Edward Rosten[2]

[1]Department of Computer Science
University of California, Santa Cruz

[2]Department of Engineering
University of Cambridge

[3]Space and Remote Sensing Sciences
Los Alamos National Laboratory

ISSDM Day
October 18, 2010

Problem

Find (x, y) locations of objects.

Figure: Predictions marked in red.

Problem

Find (x, y) locations of objects.

Figure: Predictions marked in red.

Standard Methods

Sliding Window Detector

1. Train window detector
2. Slide window over image
   - all positions
   - all orientations
   - all scales
3. Arbitrate overlapping detections

Good for few large objects (face in a portrait)

How to find many smaller objects?

How do we find objects?

Pixel-based

find pixels of many objects

Beamer (2007-2009)

Location-based

make guesses of locations

HoS Boosting (late 2009-2010)

How do we find objects?

Pixel-based

find pixels of many objects

Beamer (2007-2009)

Location-based

make guesses of locations

HoS Boosting (late 2009-2010)

How do we find objects?

Pixel-based

find pixels of many objects

Beamer (2007-2009)

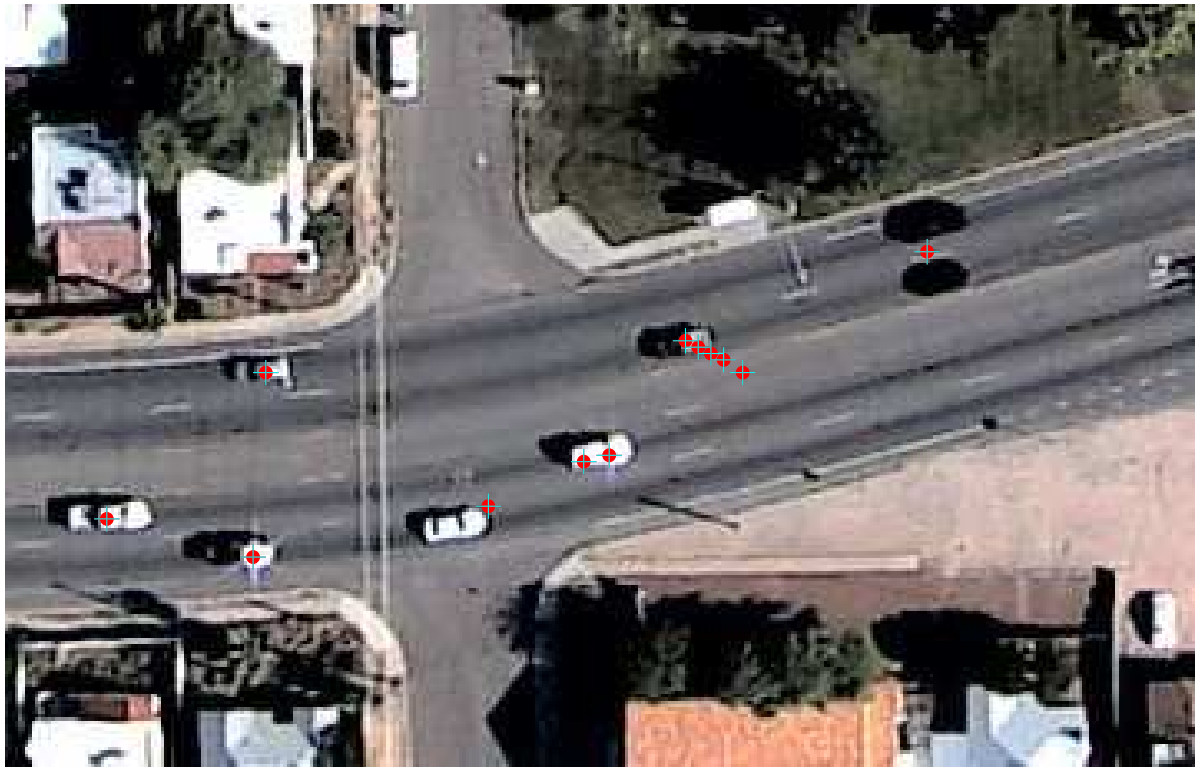
Location-based

make guesses of locations

HoS Boosting (late 2009-2010)

What is object detection anyway?

Is finding 50% of the pixels in all objects the same as finding 100% of the pixels in 50% of the objects?



Location Boosting

Radically different approach

- learns and predicts in (x, y) location space
- combines ensemble of **weak** (x, y) location predictors into **strong** predictor.

Contributions

- **new kind of model**: each weak hypothesis is a meta-object detector.
- **new loss function**: spatially motivated
- **adaBoost variant**: provably minimize loss function every iteration

Location Boosting

Radically different approach

- learns and predicts in (x, y) location space
- combines ensemble of **weak** (x, y) location predictors into **strong** predictor.

Contributions

- **new kind of model**: each weak hypothesis is a meta-object detector.
- **new loss function**: spatially motivated
- **adaBoost variant**: provably minimize loss function every iteration

Hit-or-Shift (HoS) Boosting

Definition

A **weak hypothesis** predicts locations on an image
$h = \{((x_1, y_1), c_1),((x_2, y_2), c_2), \ldots , ((x_n, y_n), c_n)\}$.

We filter away predictions with confidence lower than θ,
$h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}$.

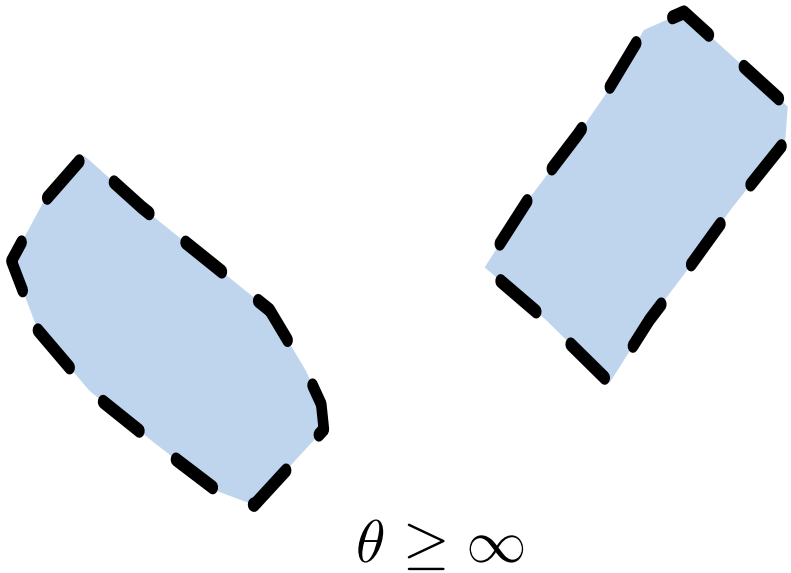


Figure: Illustrates how $h(\theta)$ changes as θ is lowered.

Definition

A **weak hypothesis** predicts locations on an image
$h = \{((x_1, y_1), c_1),((x_2, y_2), c_2), \ldots , ((x_n, y_n), c_n)\}$.

We filter away predictions with confidence lower than θ,
$h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}$.

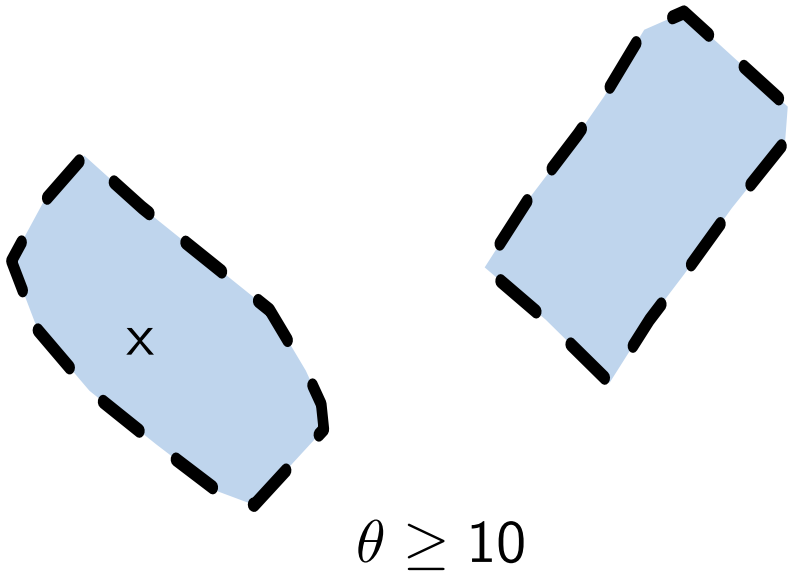


Figure: Illustrates how $h(\theta)$ changes as θ is lowered.

Definition

A **weak hypothesis** predicts locations on an image
$h = \{((x_1, y_1), c_1),((x_2, y_2), c_2), \ldots , ((x_n, y_n), c_n)\}$.

We filter away predictions with confidence lower than θ,
$h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}$.

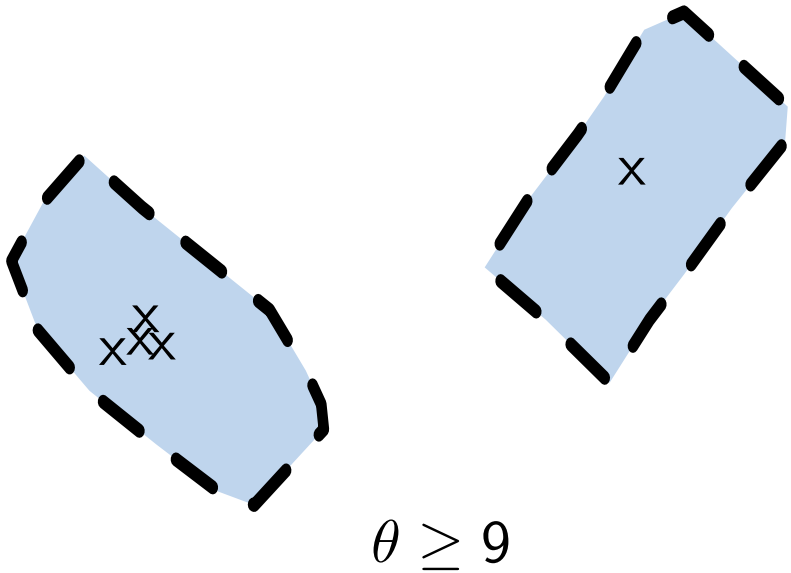


Figure: Illustrates how $h(\theta)$ changes as θ is lowered.

Definition

A **weak hypothesis** predicts locations on an image
$h = \{((x_1, y_1), c_1),((x_2, y_2), c_2), \ldots , ((x_n, y_n), c_n)\}$.

We filter away predictions with confidence lower than θ,
$h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}$.

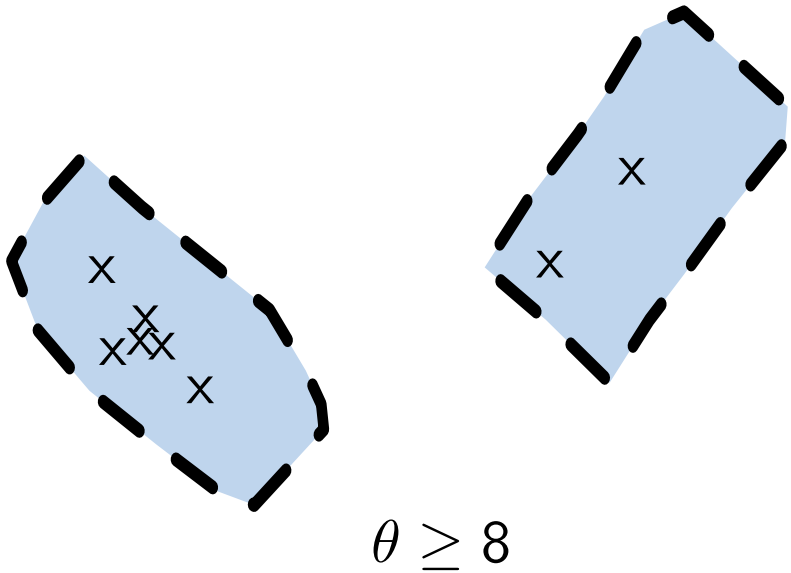


Figure: Illustrates how $h(\theta)$ changes as θ is lowered.

Definition

A **weak hypothesis** predicts locations on an image
$h = \{((x_1, y_1), c_1),((x_2, y_2), c_2), \ldots , ((x_n, y_n), c_n)\}$.

We filter away predictions with confidence lower than θ,
$h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}$.

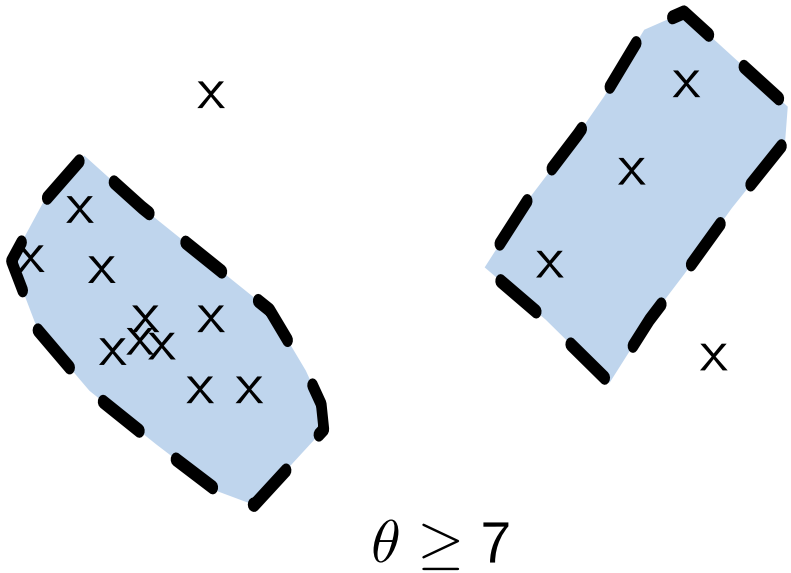


Figure: Illustrates how $h(\theta)$ changes as θ is lowered.

Definition

A **weak hypothesis** predicts locations on an image
$h = \{((x_1, y_1), c_1),((x_2, y_2), c_2), \ldots , ((x_n, y_n), c_n)\}$.

We filter away predictions with confidence lower than θ,
$h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}$.

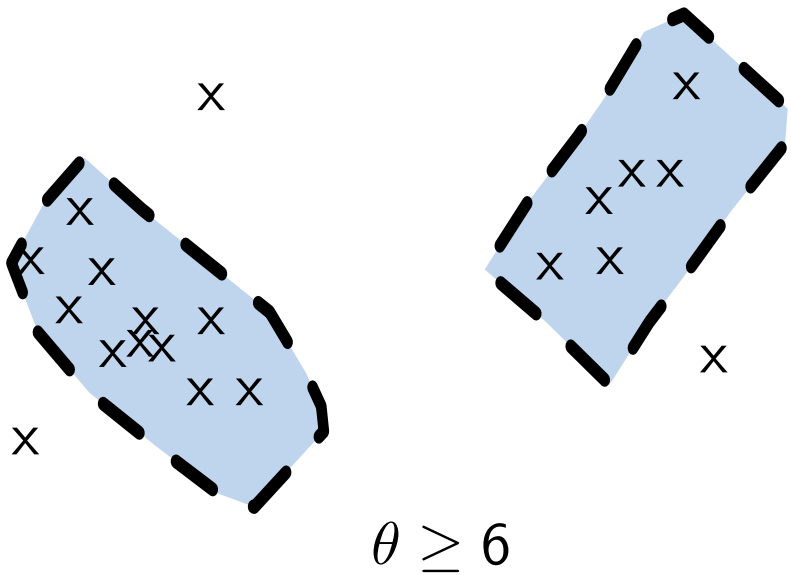


Figure: Illustrates how $h(\theta)$ changes as θ is lowered.

Definition

A **weak hypothesis** predicts locations on an image
$h = \{((x_1, y_1), c_1),((x_2, y_2), c_2), \ldots , ((x_n, y_n), c_n)\}$.

We filter away predictions with confidence lower than θ,
$h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}$.

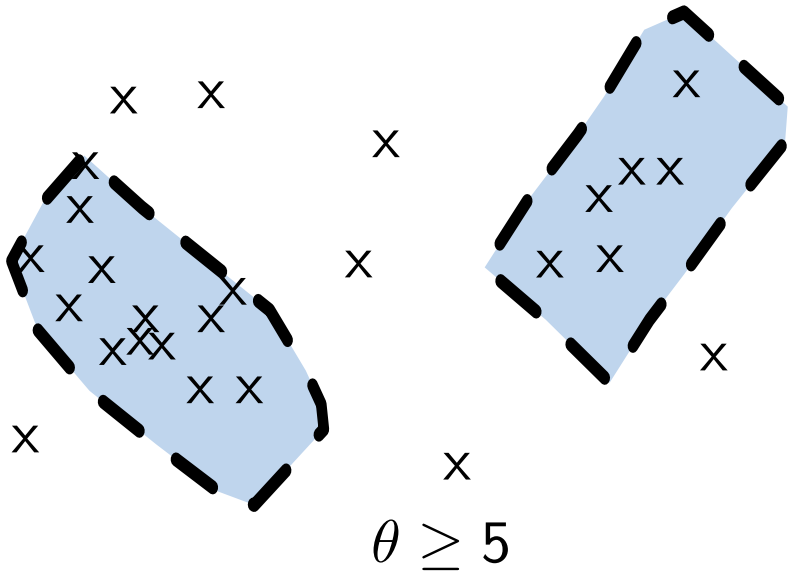


Figure: Illustrates how $h(\theta)$ changes as θ is lowered.

Definition

A **weak hypothesis** predicts locations on an image
$h = \{((x_1, y_1), c_1),((x_2, y_2), c_2), \ldots , ((x_n, y_n), c_n)\}$.

We filter away predictions with confidence lower than θ,
$h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}$.

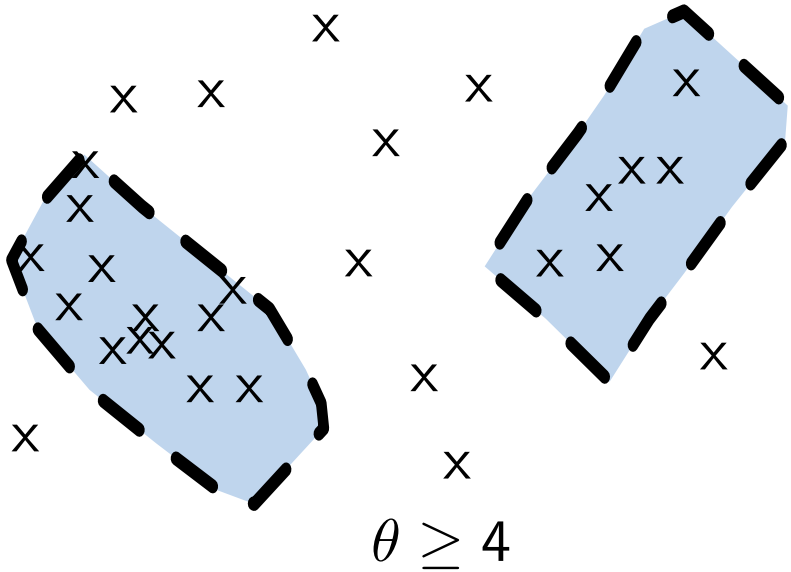


Figure: Illustrates how $h(\theta)$ changes as θ is lowered.

Definition

A **weak hypothesis** predicts locations on an image
$h = \{((x_1, y_1), c_1),((x_2, y_2), c_2), \ldots , ((x_n, y_n), c_n)\}$.

We filter away predictions with confidence lower than θ,
$h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}$.

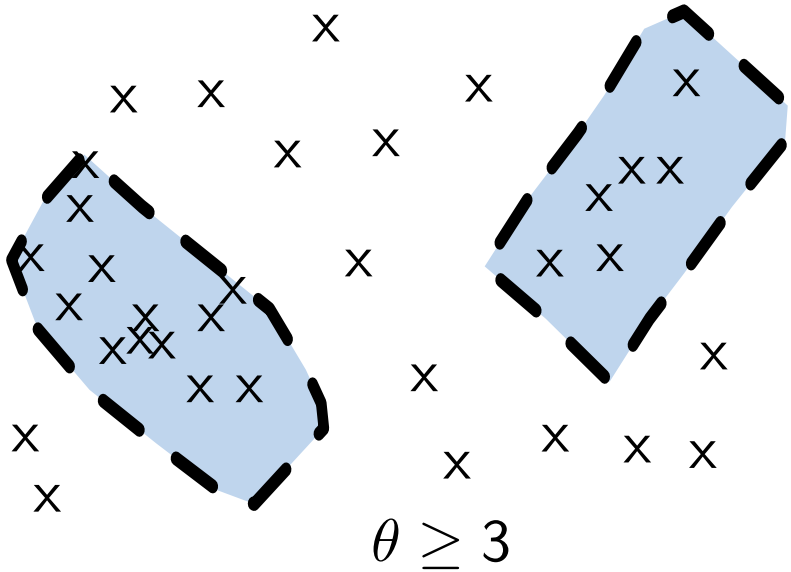


Figure: Illustrates how $h(\theta)$ changes as θ is lowered.

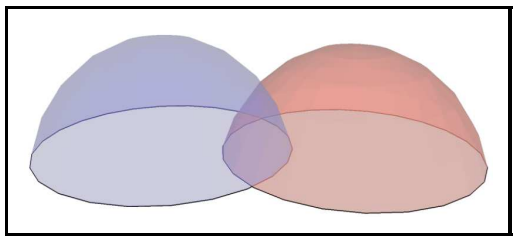
Correlation

Definition

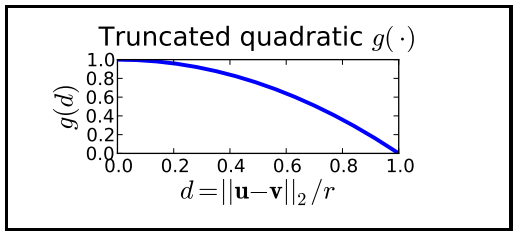
The correlation function is given by

$$C(\mathbf{u},\mathbf{v}) = \int g\left(\frac{\|\mathbf{u}-\mathbf{w}\|_2}{r}\right) g\left(\frac{\|\mathbf{v}-\mathbf{w}\|_2}{r}\right) d\mathbf{w}. \tag{1}$$

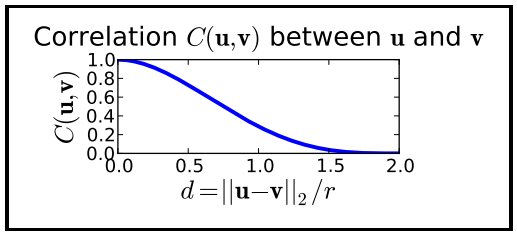
We require $0 \leq C(\mathbf{u},\mathbf{v}) \leq 1$ and $C(\mathbf{u},\mathbf{v}) = C(\mathbf{v},\mathbf{u})$.



(a)



(b)



(c)

Figure: Plots of (a) two overlapping quadratic bumps with centers $\mathbf{u}$ and $\mathbf{v}$, (b) the truncated quadratic kernel g as a function of distance $d = ||\mathbf{u}-\mathbf{v}||_2/r$, and (c) the correlation C.

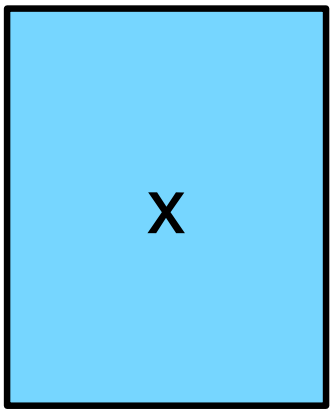
Objectness

Definition

The **objectness** of a location **u** for h's predictions with confidence at least θ is

$$f(\mathbf{u};\theta) = \sum_{\mathbf{v}\in h(\theta)} C(\mathbf{u},\mathbf{v}) \tag{2}$$

where $C(\mathbf{u},\mathbf{v})$ quantifies the relatedness **u** and **v**.



Objectness at **x**: 0.0

Ideally

- **high** objectness on **objects**
- **low** objectness elsewhere

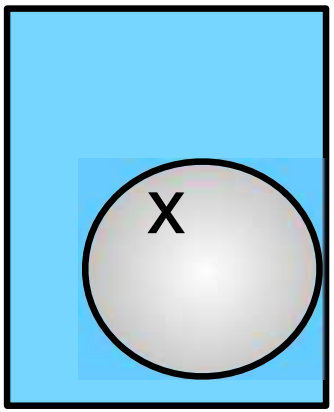
Objectness

Definition

The **objectness** of a location **u** for h's predictions with confidence at least θ is

$$f(\mathbf{u};\theta) = \sum_{\mathbf{v}\in h(\theta)} C(\mathbf{u},\mathbf{v}) \tag{2}$$

where $C(\mathbf{u},\mathbf{v})$ quantifies the relatedness **u** and **v**.



Objectness at **x**: 0.2

Ideally

- **high** objectness on **objects**
- **low** objectness elsewhere

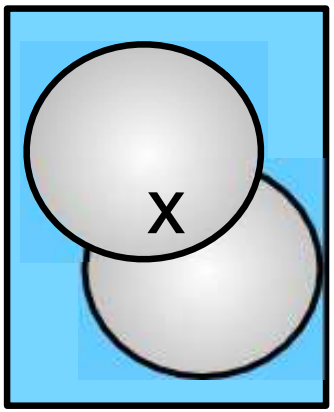
Objectness

Definition

The **objectness** of a location **u** for h's predictions with confidence at least θ is

$$f(\mathbf{u};\theta) = \sum_{\mathbf{v}\in h(\theta)} C(\mathbf{u},\mathbf{v}) \tag{2}$$

where $C(\mathbf{u},\mathbf{v})$ quantifies the relatedness **u** and **v**.



Objectness at **x**: 0.5

Ideally

- **high** objectness on **objects**
- **low** objectness elsewhere

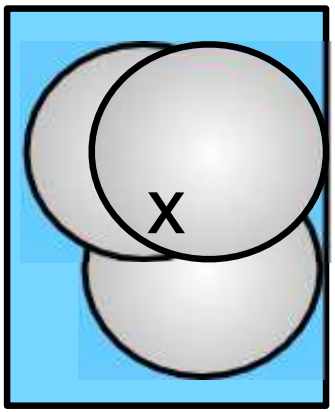
Objectness

Definition

The **objectness** of a location **u** for h's predictions with confidence at least θ is

$$f(\mathbf{u};\theta) = \sum_{\mathbf{v}\in h(\theta)} C(\mathbf{u},\mathbf{v}) \tag{2}$$

where $C(\mathbf{u},\mathbf{v})$ quantifies the relatedness **u** and **v**.



Objectness at **x**: 0.8

Ideally

- **high** objectness on **objects**
- **low** objectness elsewhere

Training Algorithm

Boosting-based learning, each iteration:

- generate many weak hypotheses (grammar)
- pick best and add to master rule
- re-weight training data
- repeat

Shift

Issues

- false positives - "objectness" never reduced
- lack of detections suggests absence of objects

Shift

Issues

- false positives - "objectness" never reduced
- lack of detections suggests absence of objects

Solution

- let weak hypotheses predict negative objectness
- same "shift value" s at all uncorrelated locations
- shift parameter s found analytically

Hit-or-Shift (HoS) Framework

Definition

A **hit-or-shift (HoS) weak hypothesis** predicts positive carness when $f(\mathbf{x};\theta)$ is positive, otherwise $-s$.

$$f'(\mathbf{x}) = \begin{cases} \alpha f(\mathbf{x};\theta) & \text{if } f(\mathbf{x};\theta) > 0, \\ -s & \text{if } f(\mathbf{x};\theta) = 0. \end{cases} . \tag{3}$$

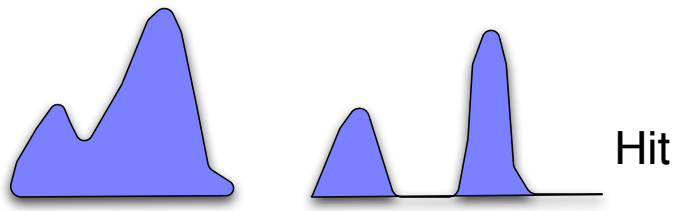


Figure: Hit-or-shift Weak Hypothesis

Hit-or-Shift (HoS) Framework

Definition

A **hit-or-shift (HoS) weak hypothesis** predicts positive carness when $f(\mathbf{x};\theta)$ is positive, otherwise $-s$.

$$f'(\mathbf{x}) = \begin{cases} \alpha f(\mathbf{x};\theta) & \text{if } f(\mathbf{x};\theta) > 0, \\ -s & \text{if } f(\mathbf{x};\theta) = 0. \end{cases} . \tag{3}$$

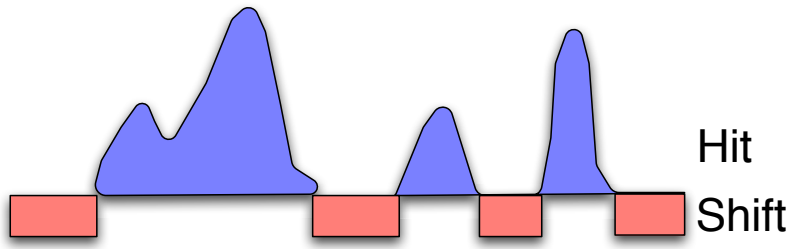


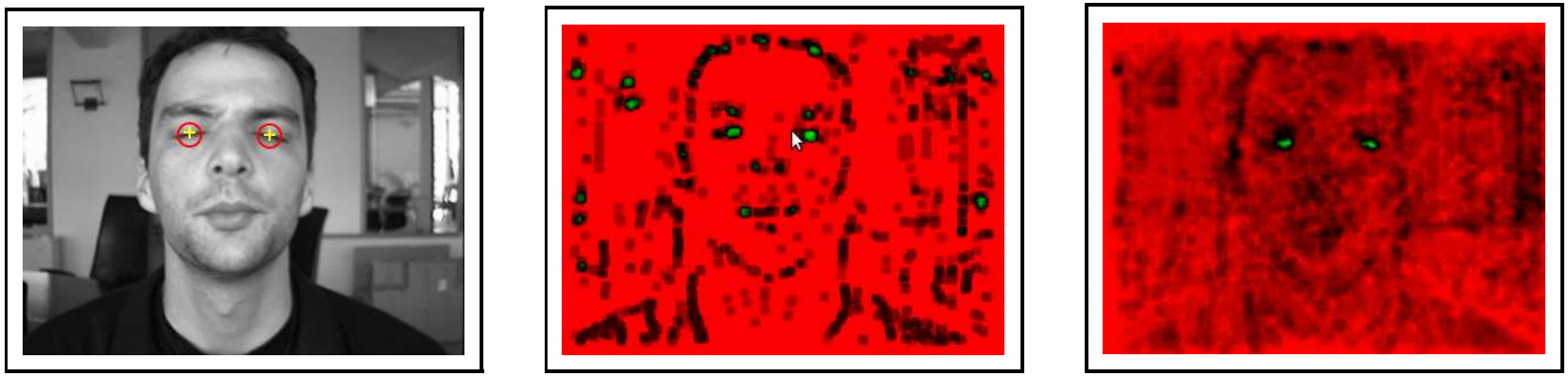
Figure: Hit-or-shift Weak Hypothesis

Hit-or-Shift (HoS) Framework

Definition

A **hit-or-shift (HoS) master hypothesis** is simply the cumulative objectness given by all weak hypotheses,

$$H_t(\mathbf{x}) = \sum_{i=1}^{t} f_i'(\mathbf{x}). \tag{4}$$



(a) (b) (c)

Figure: Plots of (a) an image and its master hypothesis after (b) 10 iterations and (c) 100 iterations.

Loss: Object and Background

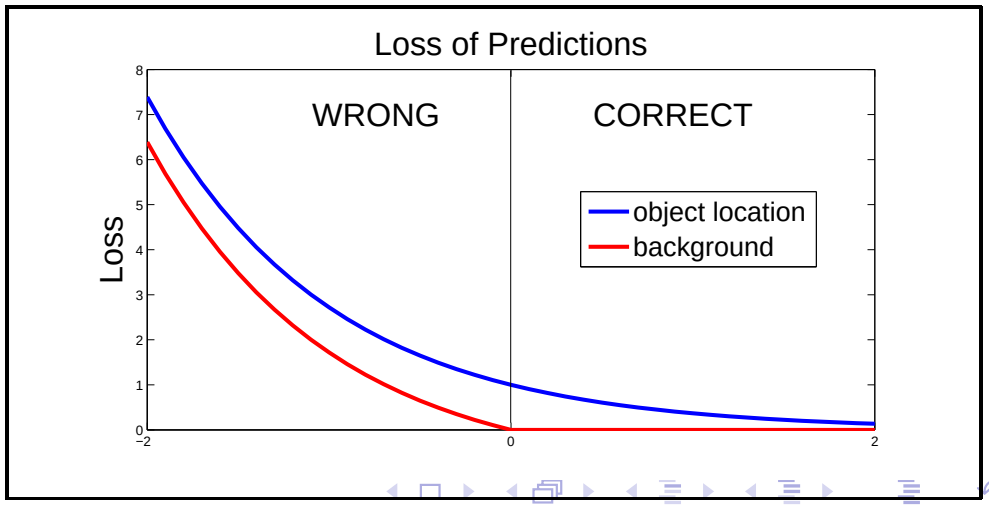
Each boosting iteration minimizes a two-part loss, the loss at **object locations**

$$\mathcal{L}^{\text{obj}} = \sum_{\mathbf{x} \in \text{obj}} \exp(-H_t(\mathbf{x})) \tag{5}$$

and the loss at the **background**,

$$\mathcal{L}^{\text{bg}} = b \sum_{\mathbf{x} \in \text{bg}} \max\{0, \exp(H_t(\mathbf{x})) - 1\}, \tag{6}$$

where b is a trade-off parameter.

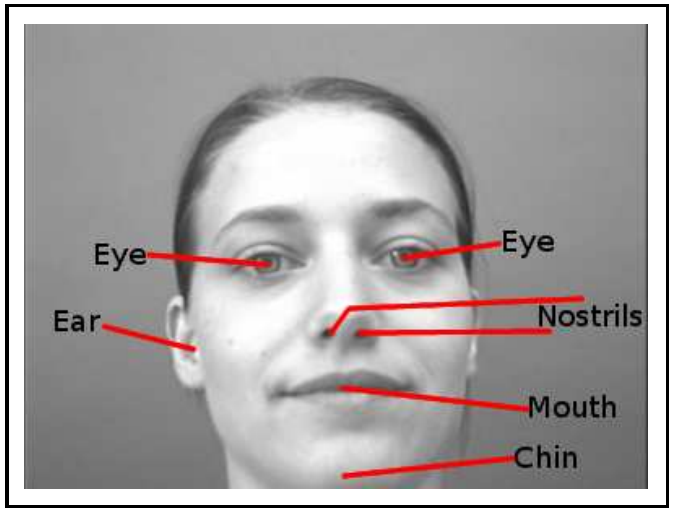


Experiments

Car Detection

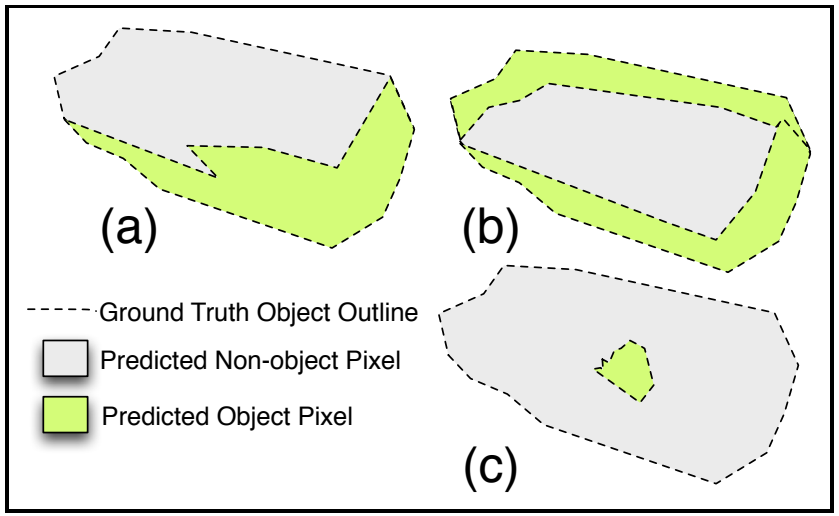
- 12 large images of Phoenix, AZ
- 300 cars labeled
- split into three partitions

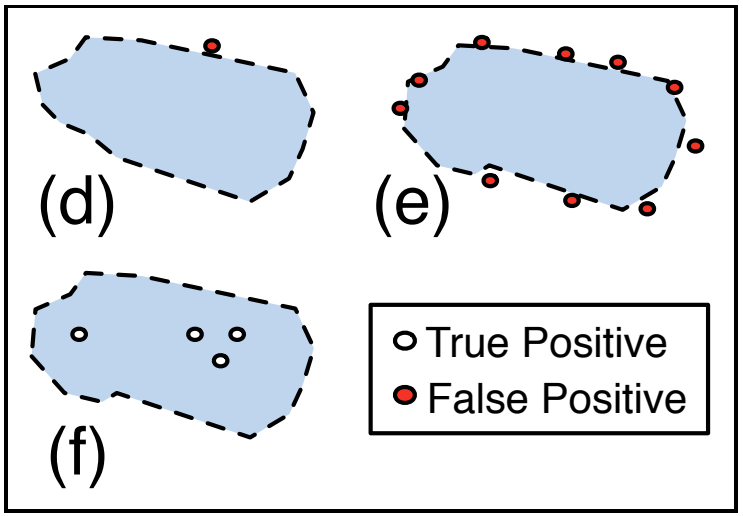
Face Labeling



- 1520 images of human faces
- label parts of face
- split into three partitions

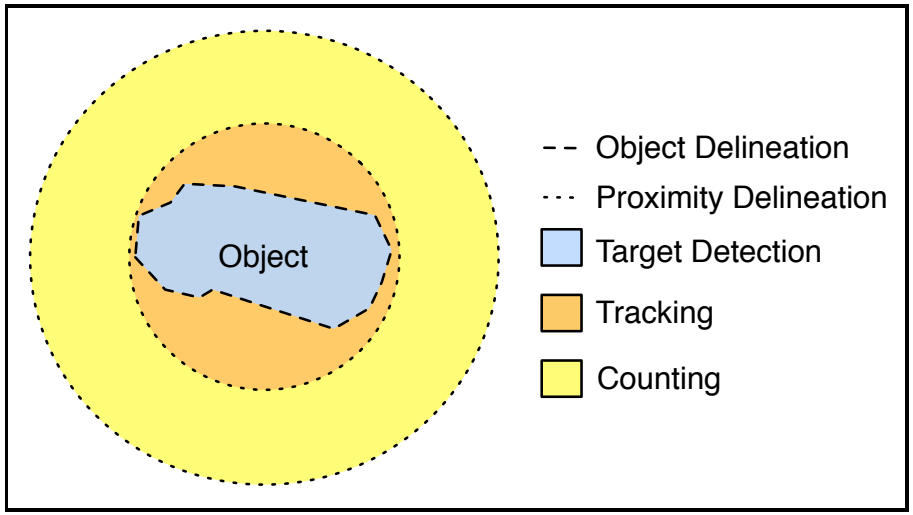
Scoring is ill-posed





Pixel classifications in (a) and (b) have about the **same number** of correct pixels, is one better? (a) and (b) have many more pixels classified correctly than (c), but (c) finds the center of the object.
Is one near miss (d) better than 10 near misses (e)? Are 5 correct hits (f) better than 1 near miss (d)?

Scoring Metric



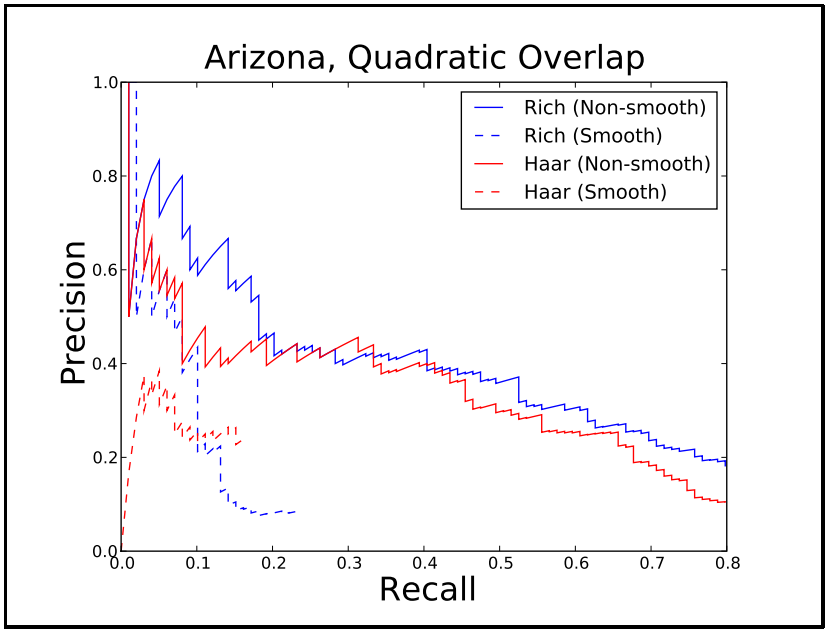
Two scoring attributes:

1. **delineation boundary:** (a) object delineation (polygon) or (b) proximity delineation (circle)
2. **multiple detections penalty:** whether to treat **multiple detections** as **false positives**

For our scoring we used a **circular delineation boundary** and **penalized multiple detections**.

Results: Arizona Test Set

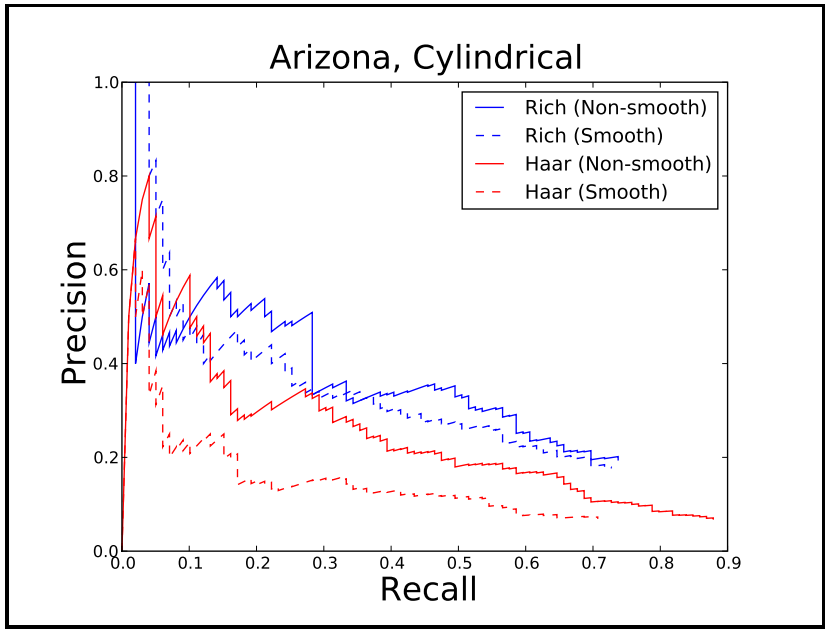
Quadratic Overlap (a)



Average Precision

	Rich	**Haar**
Kink	0.355	0.138
Smooth	0.321	0.059

Cylindrical (b)



Average Precision

	Rich	**Haar**
Kink	0.330	0.295
Smooth	0.256	0.148

Figure: Columns (a) and (b) show the precision/recall curves and average precisions for quadratic overlap and cylindrical kernels.

Results: Face Labelling

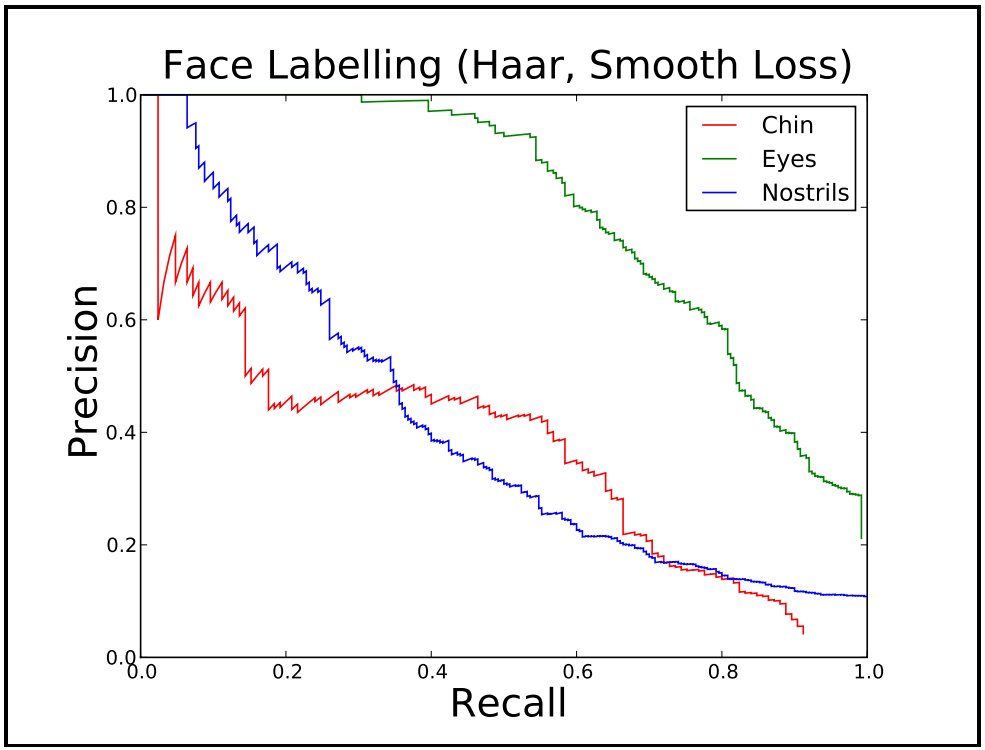


Figure: Shows precision recall curves using Haar features for labeling different parts of human faces.

Conclusions

- **location-based** approach more natural for **object detection**
  - easier data labeling
  - uniform treatment of weak/master hypotheses
  - uniform treatment of image (no subsampling)
- quickly learns and sifts through uninteresting background
- loss function directly tied to finding good locations
- **HoS weak hypotheses** are structured for efficient optimization
- can be used as a part detector
- **new algorithm**: *provably minimizes the loss* at every iteration given a new feature

Future Work

Our latest work involves significant adaptations for **large objects**

- new HoS detector based on SIFT features and vocabulary trees
- polar offset learning: exploits scale information to quickly learn offsets
- apply technique to PASCAL and CalTech data sets.