Visual Classification of Images by Learning Geometric Appearances through Boosting

Martin Antenreiter, Christian Savu-Krohn, Peter Auer

Chair of Information Technology
University of Leoben
Austria

August 30, 2006

Introduction

Classification of images through boosting
- Feature types and preprocessing steps
- LPBoost
- Weak learner

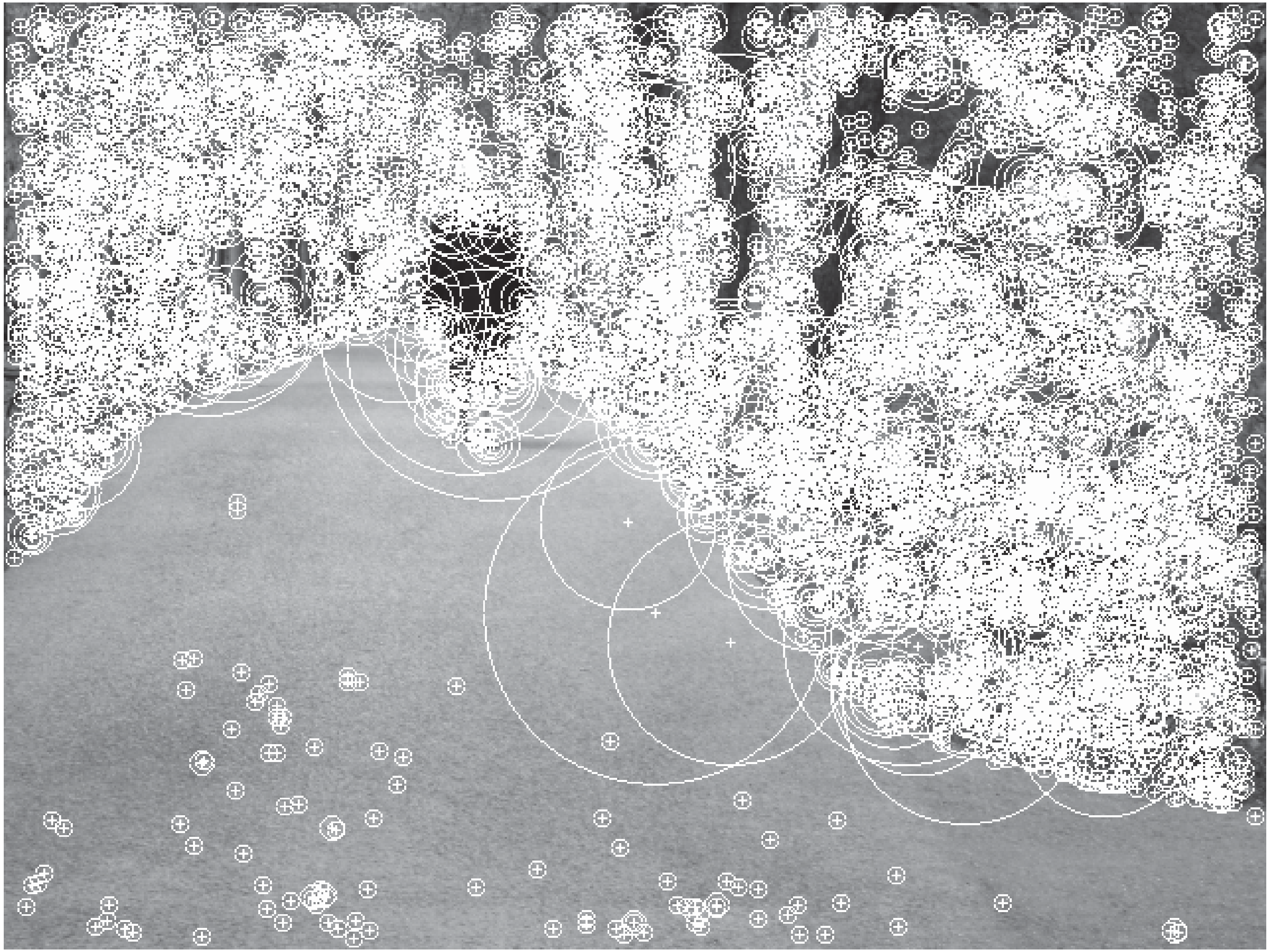
Multiclass image classification
- Weight optimization method

Evaluation and results
- Xerox dataset
- PASCAL Visual object classes challenge 2006

Example

Example

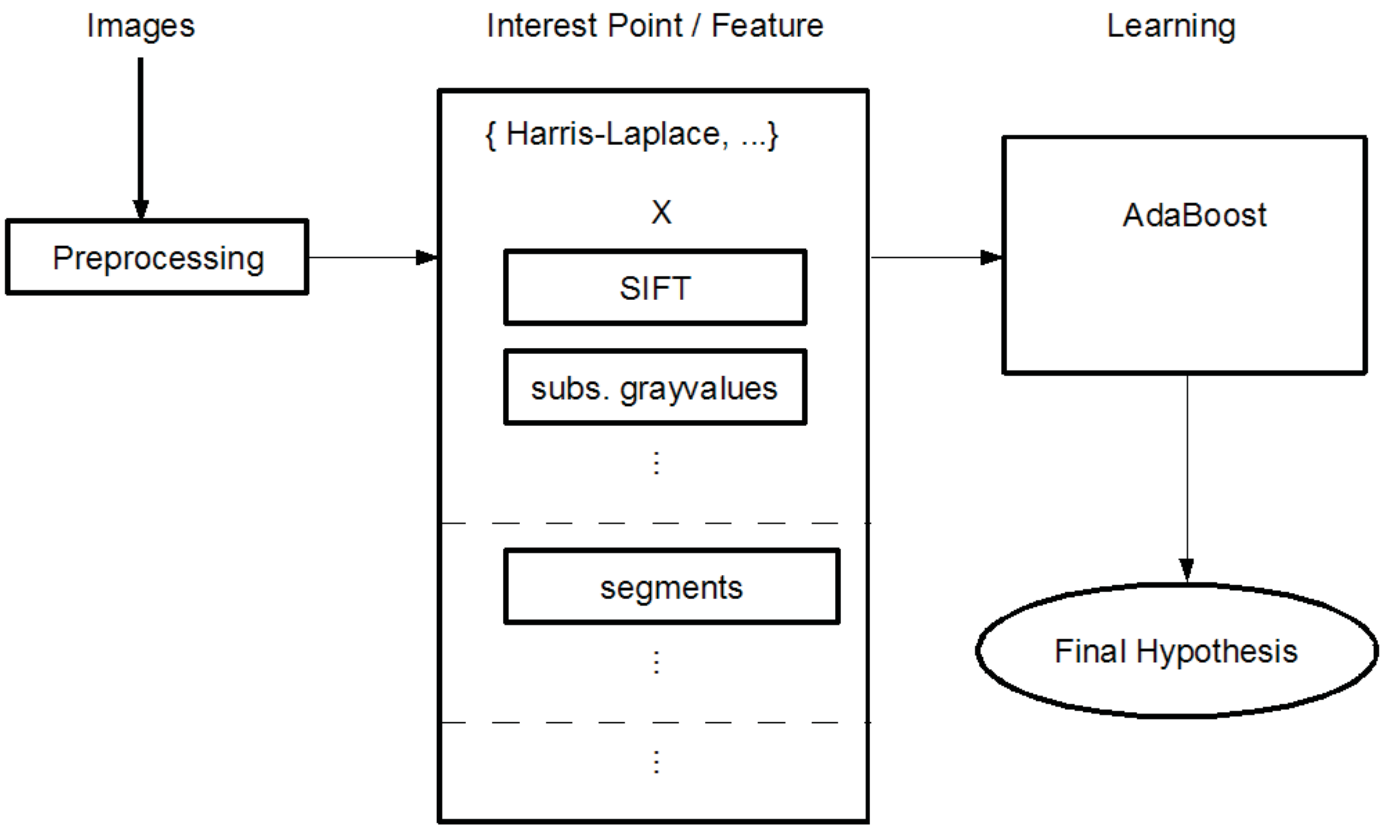
Example



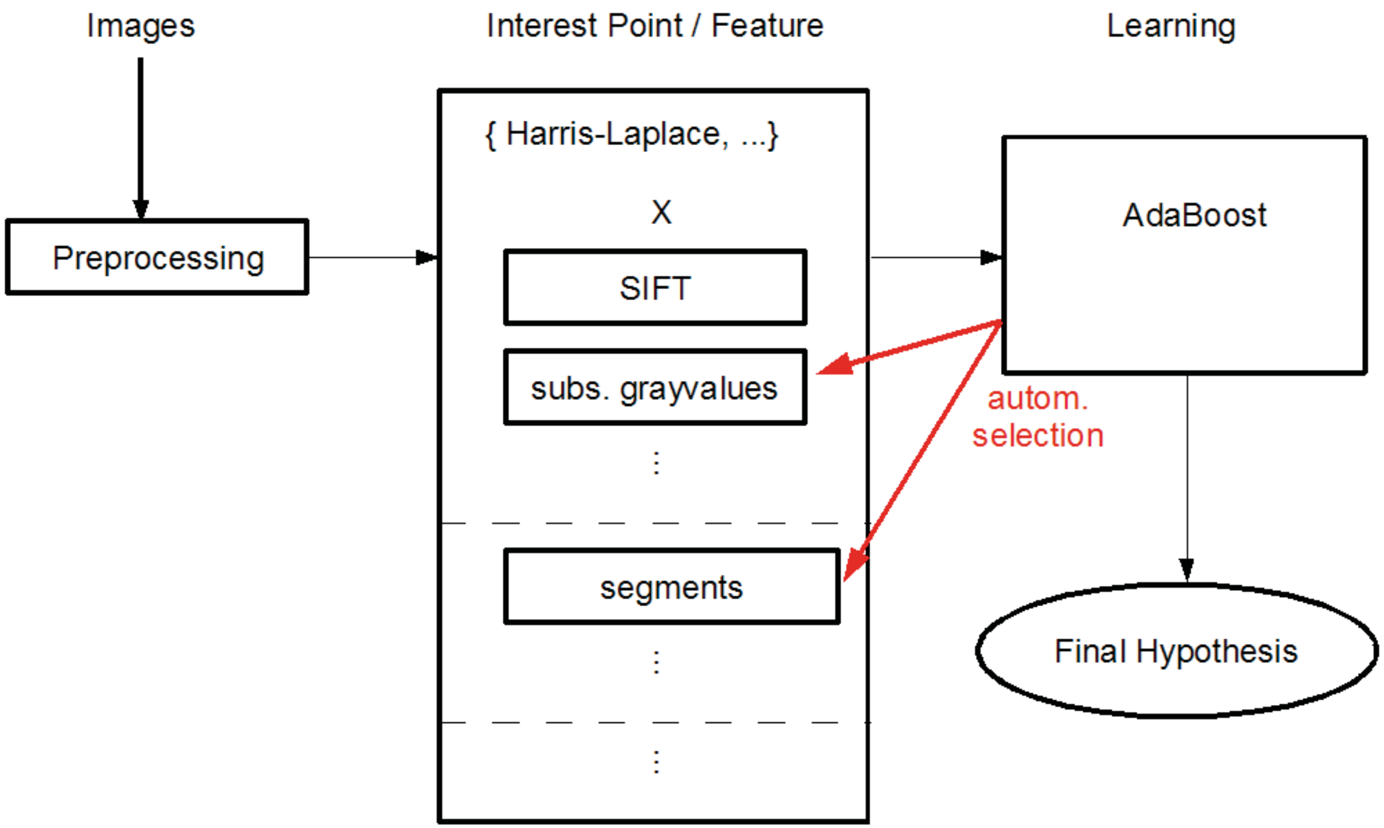
Example

Example

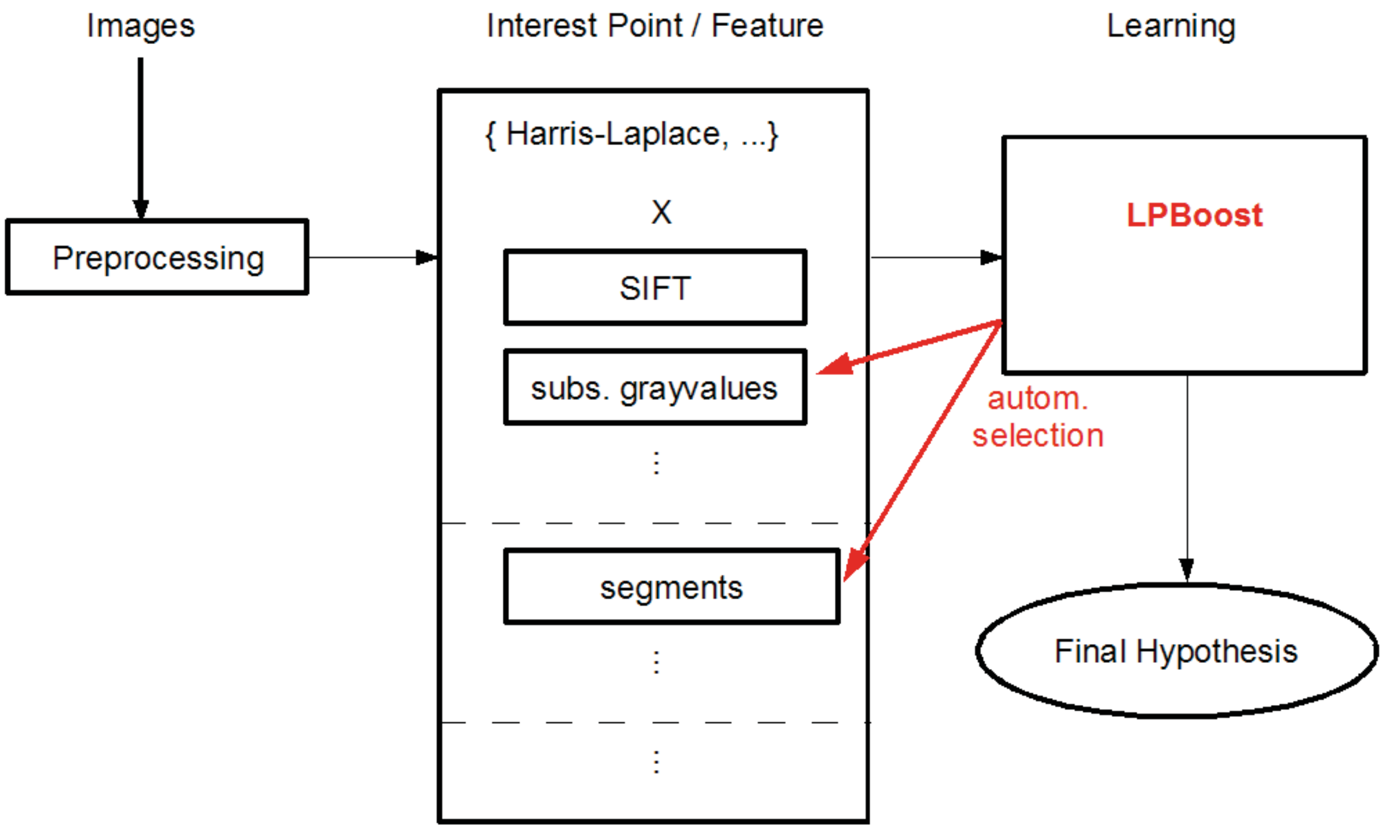
Framework


Images
Interest Point / Feature
Learning
Preprocessing
{ Harris-Laplace, ...}
X
SIFT
subs. grayvalues
segments
AdaBoost
Final Hypothesis

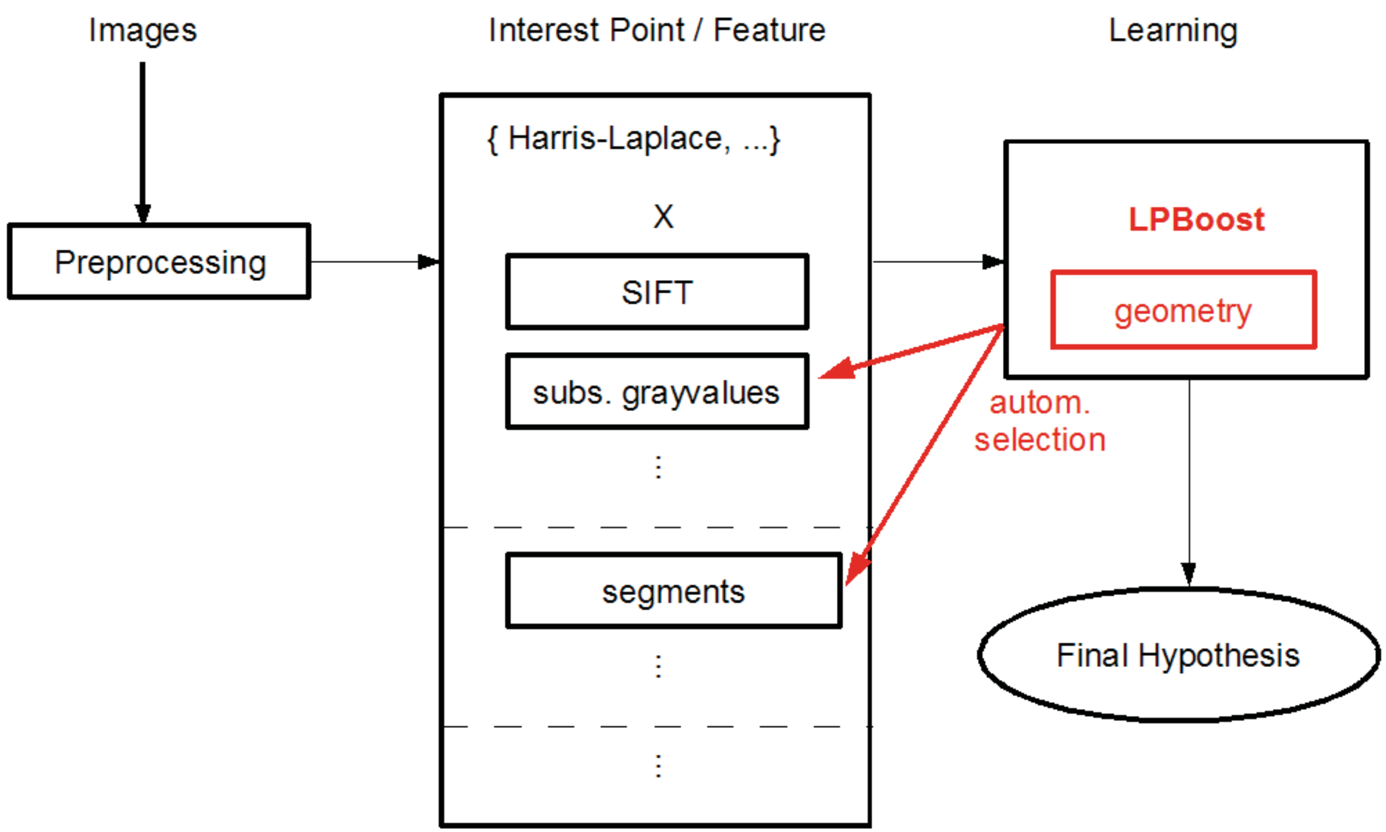
Framework


Images
Interest Point / Feature
Learning
Preprocessing
{ Harris-Laplace, ...}
X
SIFT
subs. grayvalues
segments
AdaBoost
autom.
selection
Final Hypothesis

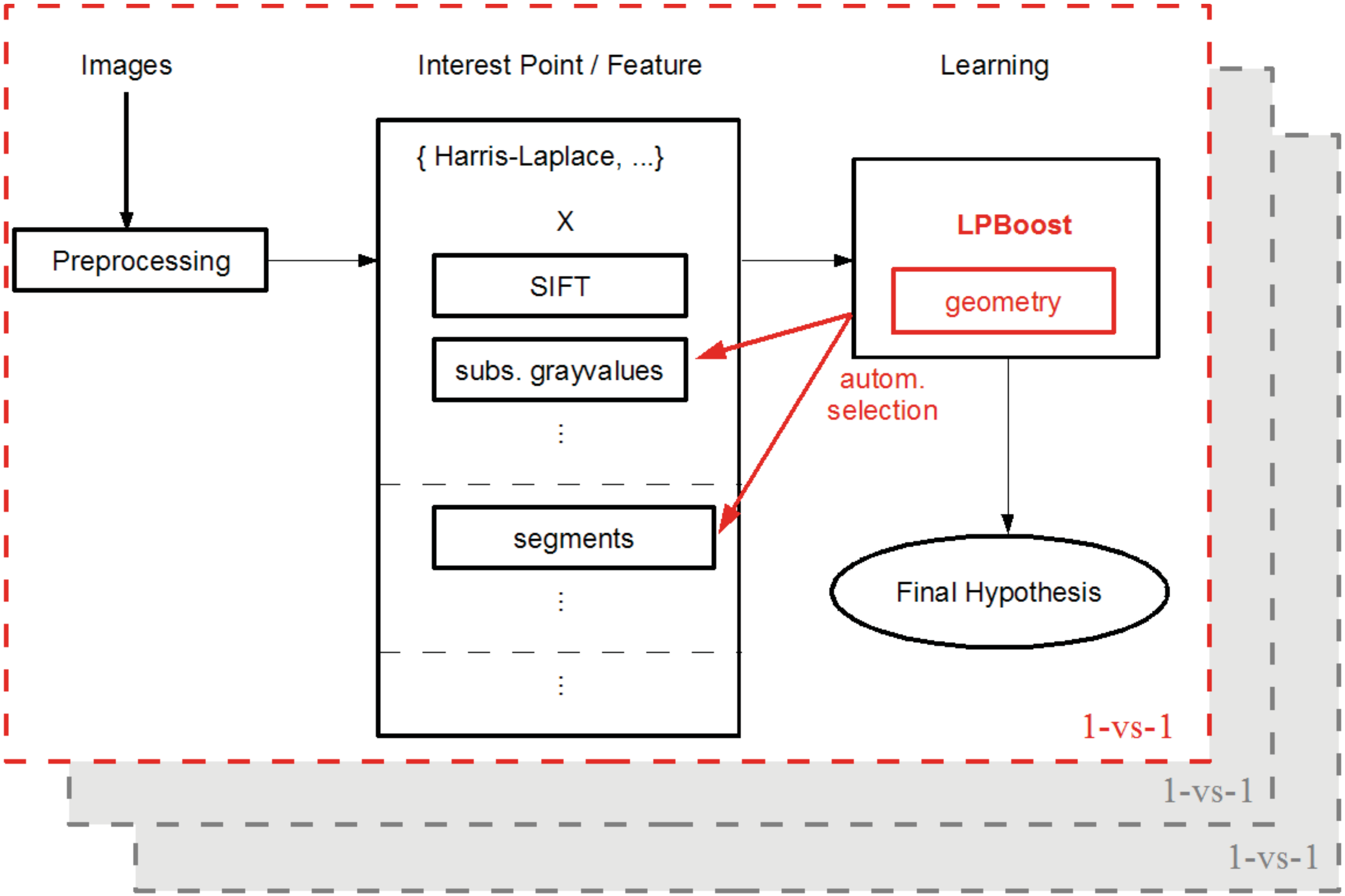
Framework


Images
Interest Point / Feature
Learning
{ Harris-Laplace, ...}
X
Preprocessing
SIFT
subs. grayvalues
segments
LPBoost
autom.
selection
Final Hypothesis

Framework


Images
Interest Point / Feature
Learning
Preprocessing
{ Harris-Laplace, ...}
X
SIFT
subs. grayvalues
segments
LPBoost
geometry
autom.
selection
Final Hypothesis

Framework


Images
Interest Point / Feature
Learning
{ Harris-Laplace, ...}
X
Preprocessing
SIFT
subs. grayvalues
segments
LPBoost
geometry
autom.
selection
Final Hypothesis
1-vs-1
1-vs-1
1-vs-1

Outline

Introduction

Classification of images through boosting
- Feature types and preprocessing steps
- LPBoost
- Weak learner

Multiclass image classification
- Weight optimization method

Evaluation and results
- Xerox dataset
- PASCAL Visual object classes challenge 2006

Feature types and preprocessing steps

ϕ	feature type	int. norm.	whitening	$k_\phi = \lfloor 2\sqrt{m_\phi} \rfloor$
1	subsampled grayval.		x	1 848
2		x	x	1 848
3	basic moments		x	1 846
4		x	x	1 848
5	moment invariants [3]		x	1 848
6		x	x	1 848
7	SIFTS [4]		x	1 798
8			PCA 40	1 798
9	segments [2]		x	1 661

$\sum$ 16 343

"reference features"

Outline

Introduction

Classification of images through boosting
- Feature types and preprocessing steps
- LPBoost
- Weak learner

Multiclass image classification
- Weight optimization method

Evaluation and results
- Xerox dataset
- PASCAL Visual object classes challenge 2006

- combines weak hypotheses to a strong hypothesis:

$$f(x_i) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \in \{+1, -1\}$$

- Primal:

$$\begin{array}{lll} \max_{\rho,a,\xi} & \rho - D\sum_{i=1}^{m} \xi_i & \\ s.t. & y_i \sum_{t=1}^{T} \alpha_t h_t(x_i) + \xi_i \geq \rho & i = 1, \dots, m \\ & \sum_{t=1}^{T} \alpha_t = 1 & \alpha_t \geq 0 \\ & \xi_i \geq 0 & i = 1, \dots, m \end{array}$$

- Dual:

$$\begin{array}{lll} \min_{\beta,w} & \beta & \\ s.t. & \sum_{i=1}^{m} y_i w_i h_t(x_i) \leq \beta & t = 1, \dots, T \\ & \sum_{i=1}^{m} w_i = 1 & 0 \leq w_i \leq D \end{array}$$

\+ has a well-defined stopping criterion
\+ is a soft-margin classifier

- combines weak hypotheses to a strong hypothesis:

$$f(x_i) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \in \{+1, -1\}$$

- Primal:

$$\begin{array}{lll} \max_{\rho,a,\xi} & \rho - D\sum_{i=1}^{m} \xi_i & \\ s.t. & y_i \sum_{t=1}^{T} \alpha_t h_t(x_i) + \xi_i \geq \rho & i = 1, \dots, m \\ & \sum_{t=1}^{T} \alpha_t = 1 & \alpha_t \geq 0 \\ & \xi_i \geq 0 & i = 1, \dots, m \end{array}$$

- Dual:

$$\begin{array}{lll} \min_{\beta,w} & \beta & \\ s.t. & \sum_{i=1}^{m} y_i w_i h_t(x_i) \leq \beta & t = 1, \dots, T \\ & \sum_{i=1}^{m} w_i = 1 & 0 \leq w_i \leq D \end{array}$$

+ has a well-defined stopping criterion
+ is a soft-margin classifier

- combines weak hypotheses to a strong hypothesis:

$$f(x_i) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \in \{+1, -1\}$$

- Primal:

$$\begin{array}{lll} \max_{\rho,a,\xi} & \rho - D\sum_{i=1}^{m} \xi_i & \\ s.t. & y_i \sum_{t=1}^{T} \alpha_t h_t(x_i) + \xi_i \geq \rho & i = 1, \dots, m \\ & \sum_{t=1}^{T} \alpha_t = 1 & \alpha_t \geq 0 \\ & \xi_i \geq 0 & i = 1, \dots, m \end{array}$$

- Dual:

$$\begin{array}{lll} \min_{\beta,w} & \beta & \\ s.t. & \sum_{i=1}^{m} y_i w_i h_t(x_i) \leq \beta & t = 1, \dots, T \\ & \sum_{i=1}^{m} w_i = 1 & 0 \leq w_i \leq D \end{array}$$

+ has a well-defined stopping criterion
+ is a soft-margin classifier

- combines weak hypotheses to a strong hypothesis:

$$f(x_i) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \in \{+1, -1\}$$

- Primal:

$$\begin{array}{lll} \max_{\rho,a,\xi} & \rho - D\sum_{i=1}^{m} \xi_i & \\ s.t. & y_i \sum_{t=1}^{T} \alpha_t h_t(x_i) + \xi_i \geq \rho & i = 1, \ldots, m \\ & \sum_{t=1}^{T} \alpha_t = 1 & \alpha_t \geq 0 \\ & \xi_i \geq 0 & i = 1, \ldots, m \end{array}$$

- Dual:

$$\begin{array}{lll} \min_{\beta,w} & \beta & \\ s.t. & \sum_{i=1}^{m} y_i w_i h_t(x_i) \leq \beta & t = 1, \ldots, T \\ & \sum_{i=1}^{m} w_i = 1 & 0 \leq w_i \leq D \end{array}$$

+ has a well-defined stopping criterion
+ is a soft-margin classifier

- combines weak hypotheses to a strong hypothesis:

$$f(x_i) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \in \{+1, -1\}$$

- Primal:

$$\begin{array}{lll} \max_{\rho,a,\xi} & \rho - D\sum_{i=1}^{m} \xi_i & \\ s.t. & y_i \sum_{t=1}^{T} \alpha_t h_t(x_i) + \xi_i \geq \rho & i = 1, \ldots, m \\ & \sum_{t=1}^{T} \alpha_t = 1 & \alpha_t \geq 0 \\ & \xi_i \geq 0 & i = 1, \ldots, m \end{array}$$

- Dual:

$$\begin{array}{lll} \min_{\beta,w} & \beta & \\ s.t. & \sum_{i=1}^{m} y_i w_i h_t(x_i) \leq \beta & t = 1, \ldots, T \\ & \sum_{i=1}^{m} w_i = 1 & 0 \leq w_i \leq D \end{array}$$

+ has a well-defined stopping criterion
+ is a soft-margin classifier

Outline

Introduction

Classification of images through boosting
Feature types and preprocessing steps
LPBoost
Weak learner

Multiclass image classification
Weight optimization method

Evaluation and results
Xerox dataset
PASCAL Visual object classes challenge 2006

Weak learner

- in every boosting round $t = 1, ..., T$ the weak learner selects $h^* \in \mathcal{H}$:

$$\max_{h \in \mathcal{H}} \left(\sum_{i=1}^{m} h(x_i) y_i w_i \right) = \sum_{i=1}^{m} h^*(x_i) y_i w_i. \qquad (1)$$

- kinds of weak learners:

'none' selects reference feature of type ϕ and an optimal threshold to it w.r.t. current **w**

'relations A' uses geometric primitives 'up', 'down', 'left', 'right', relating up to three (*) reference features.

'relations B' .... same, but uses eight sections.

(*) if an object category needs more than three features, our search algorithms builds hierarchies modelled as trees

- in every boosting round $t = 1, ..., T$ the weak learner selects $h^* \in \mathcal{H}$:

$$\max_{h \in \mathcal{H}} \left(\sum_{i=1}^{m} h(x_i) y_i w_i \right) = \sum_{i=1}^{m} h^*(x_i) y_i w_i. \tag{1}$$

- kinds of weak learners:

'none' selects reference feature of type ϕ and an optimal threshold to it w.r.t. current **w**

'relations A' uses geometric primitives 'up', 'down', 'left', 'right', relating up to three $^{(*)}$ reference features.

'relations B' .... same, but uses eight sections.

$^{(*)}$if an object category needs more than three features, our search algorithms builds hierarchies modelled as trees

Weak learner

- in every boosting round $t = 1, ..., T$ the weak learner selects $h^* \in \mathcal{H}$:

$$\max_{h \in \mathcal{H}} \left(\sum_{i=1}^{m} h(x_i) y_i w_i \right) = \sum_{i=1}^{m} h^*(x_i) y_i w_i. \quad (1)$$

- kinds of weak learners:

'none' selects reference feature of type ϕ and an optimal threshold to it w.r.t. current **w**

'relations A' uses geometric primitives 'up', 'down', 'left', 'right', relating up to three $^{(*)}$ reference features.

'relations B' .... same, but uses eight sections.

$^{(*)}$if an object category needs more than three features, our search algorithms builds hierarchies modelled as trees

- in every boosting round $t = 1, ..., T$ the weak learner selects $h^* \in \mathcal{H}$:

$$\max_{h \in \mathcal{H}} \left(\sum_{i=1}^{m} h(x_i) y_i w_i \right) = \sum_{i=1}^{m} h^*(x_i) y_i w_i. \qquad (1)$$

- kinds of weak learners:

’none’ selects reference feature of type ϕ and an optimal threshold to it w.r.t. current **w**

’relations A’ uses geometric primitives ’up’, ’down’, ’left’, ’right’, relating up to three $^{(*)}$ reference features.

’relations B’ .... same, but uses eight sections.

$^{(*)}$if an object category needs more than three features, our search algorithms builds hierarchies modelled as trees

- in every boosting round $t = 1, ..., T$ the weak learner selects $h^* \in \mathcal{H}$:

$$\max_{h \in \mathcal{H}} \left(\sum_{i=1}^{m} h(x_i) y_i w_i \right) = \sum_{i=1}^{m} h^*(x_i) y_i w_i. \tag{1}$$

- kinds of weak learners:

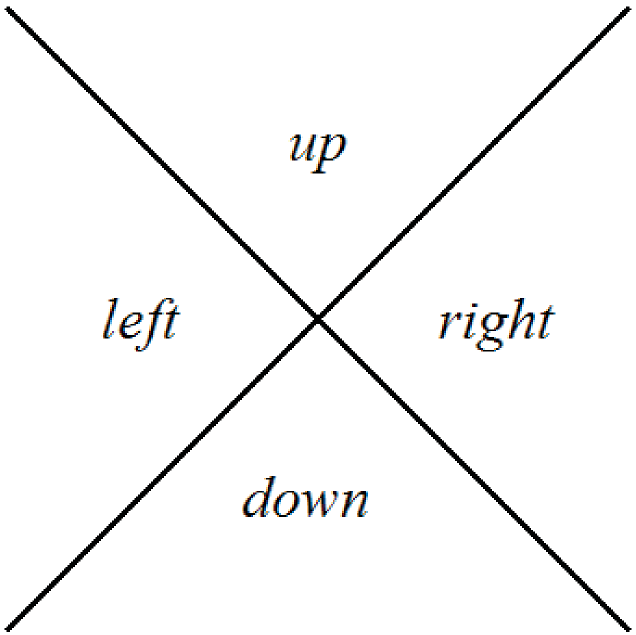
'none' selects reference feature of type ϕ and an optimal threshold to it w.r.t. current **w**

'relations A' uses geometric primitives 'up', 'down', 'left', 'right', relating up to three $^{(*)}$ reference features.

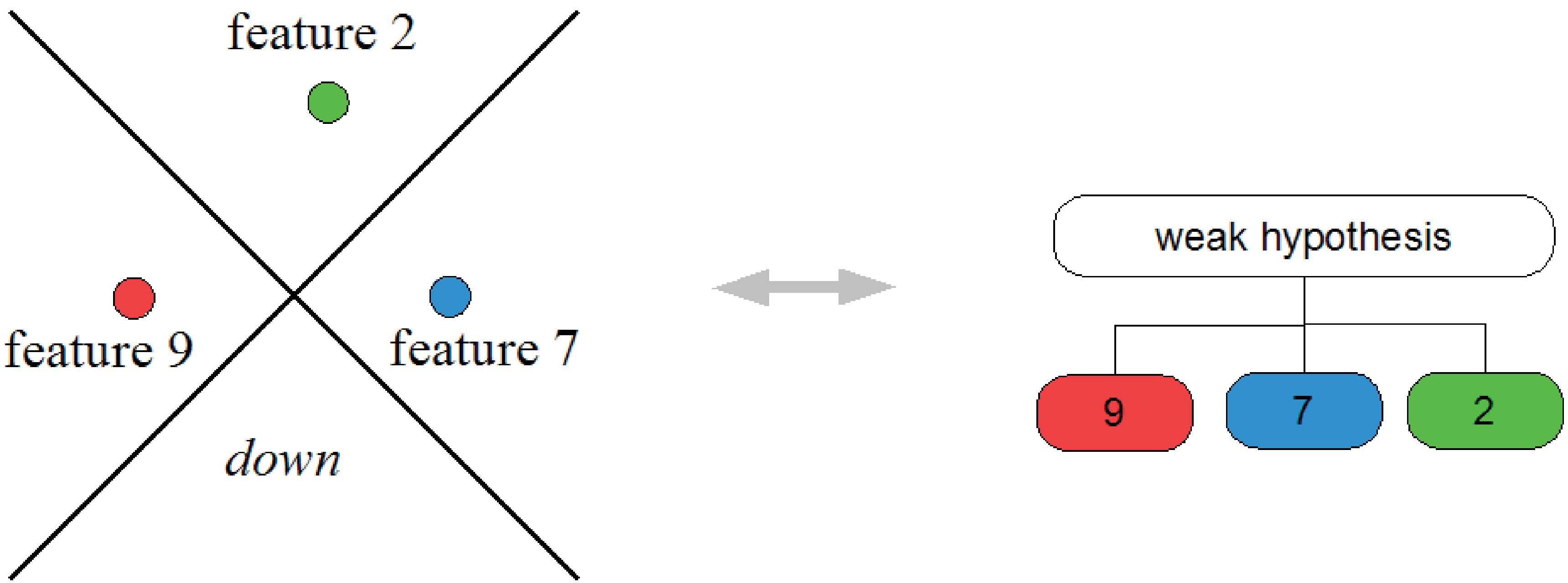
'relations B' .... same, but uses eight sections.

$^{(*)}$if an object category needs more than three features, our search algorithms builds hierarchies modelled as trees

Example for a geometric hypothesis



Example for a geometric hypothesis



Greedy search strategy

1. Select h^* (eq. 1)
2. For all previously generated hypotheses h_p, $p = 1, \dots, t-1$ do:
   1. **Create a hypothesis with a logical AND**: $h_{and} = h^*$ AND h_p.
   2. **Search geometric relations**: The two sub hypotheses from h_{and} are applied on every image yielding two point sets. We seek a common geometric relation between these sets, yielding a geometric hypothesis h_{geom}.
3. Compare performance (eq. 1) of h^* and h_{geom}, output the better.

Greedy search strategy

1. Select h^* (eq. 1)
2. For all previously generated hypotheses h_p, $p = 1, \ldots, t-1$ do:
   2.1 **Create a hypothesis with a logical AND**:
   $h_{and} = h^*$ AND h_p.
   2.2 **Search geometric relations**:
   The two sub hypotheses from h_{and} are applied on every image yielding two point sets. We seek a common geometric relation between these sets, yielding a geometric hypothesis h_{geom}.
3. Compare performance (eq. 1) of h^* and h_{geom}, output the better.

Greedy search strategy

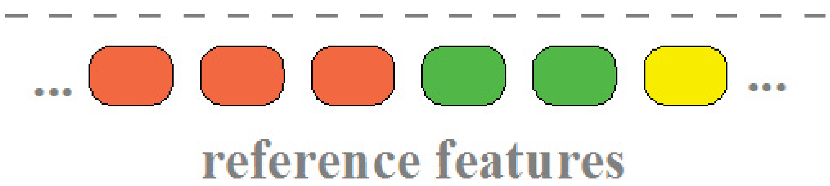
1. Select h^* (eq. 1)
2. For all previously generated hypotheses h_p, $p = 1, \ldots, t-1$ do:
   2.1 **Create a hypothesis with a logical AND**:
   $h_{and} = h^*$ AND h_p.
   2.2 **Search geometric relations**:
   The two sub hypotheses from h_{and} are applied on every image yielding two point sets. We seek a common geometric relation between these sets, yielding a geometric hypothesis h_{geom}.
3. Compare performance (eq. 1) of h^* and h_{geom}, output the better.

Greedy search strategy

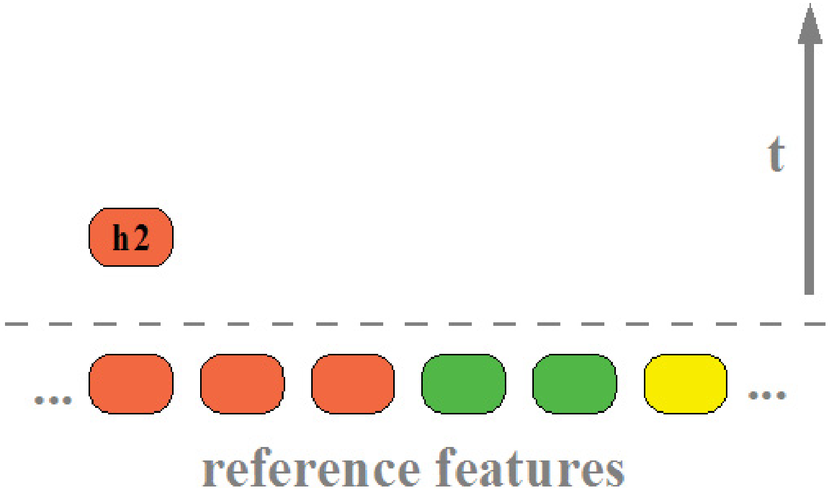
1. Select h^* (eq. 1)
2. For all previously generated hypotheses h_p, $p = 1, \ldots, t - 1$ do:
   2.1 **Create a hypothesis with a logical AND**: $h_{and} = h^*$ AND h_p.
   2.2 **Search geometric relations**: The two sub hypotheses from h_{and} are applied on every image yielding two point sets. We seek a common geometric relation between these sets, yielding a geometric hypothesis h_{geom}.
3. Compare performance (eq. 1) of h^* and h_{geom}, output the better.

Example for a hierarchy

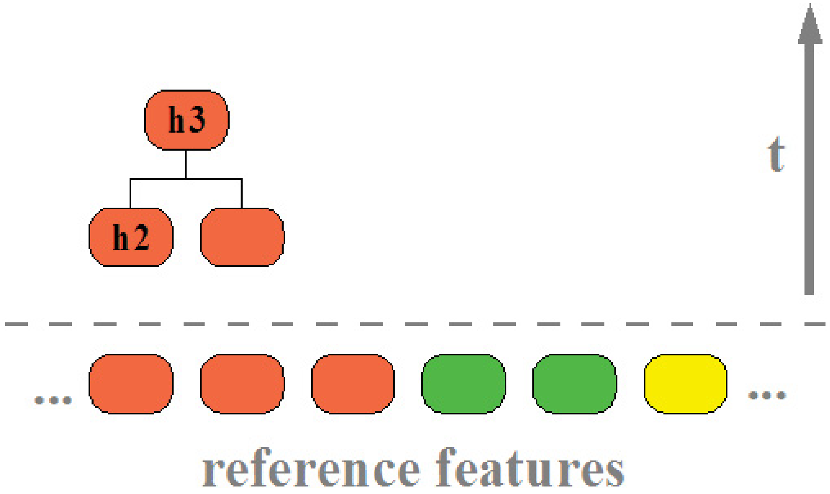
Example for a hierarchy



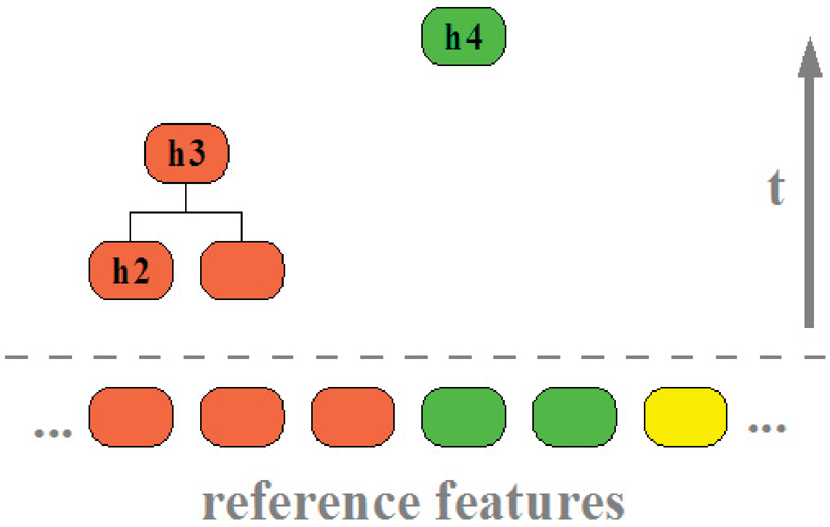
Example for a hierarchy



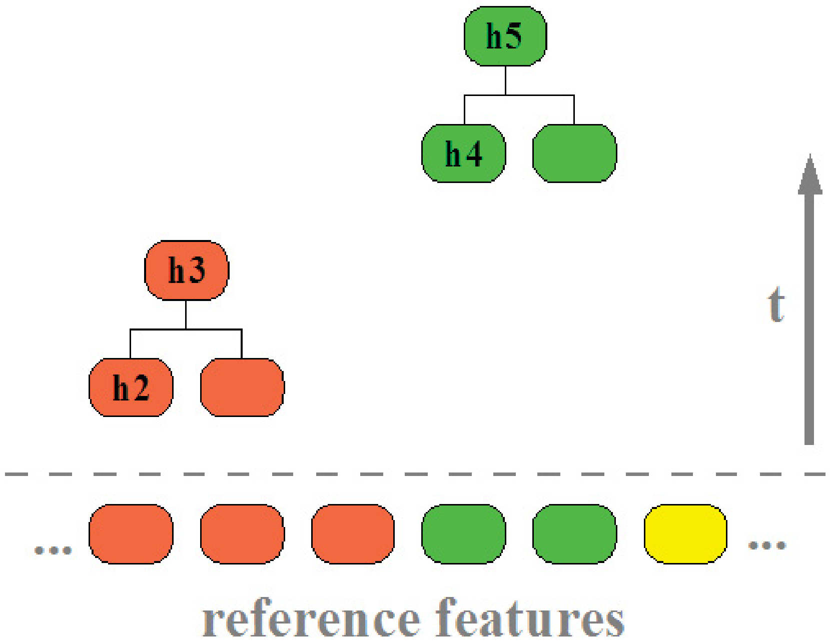
Example for a hierarchy



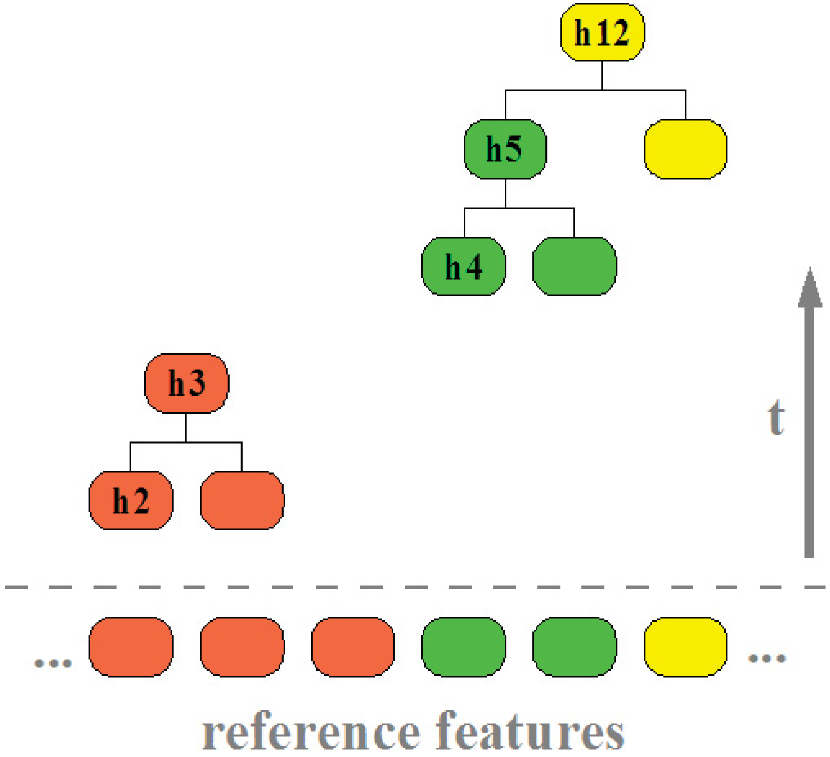
Example for a hierarchy



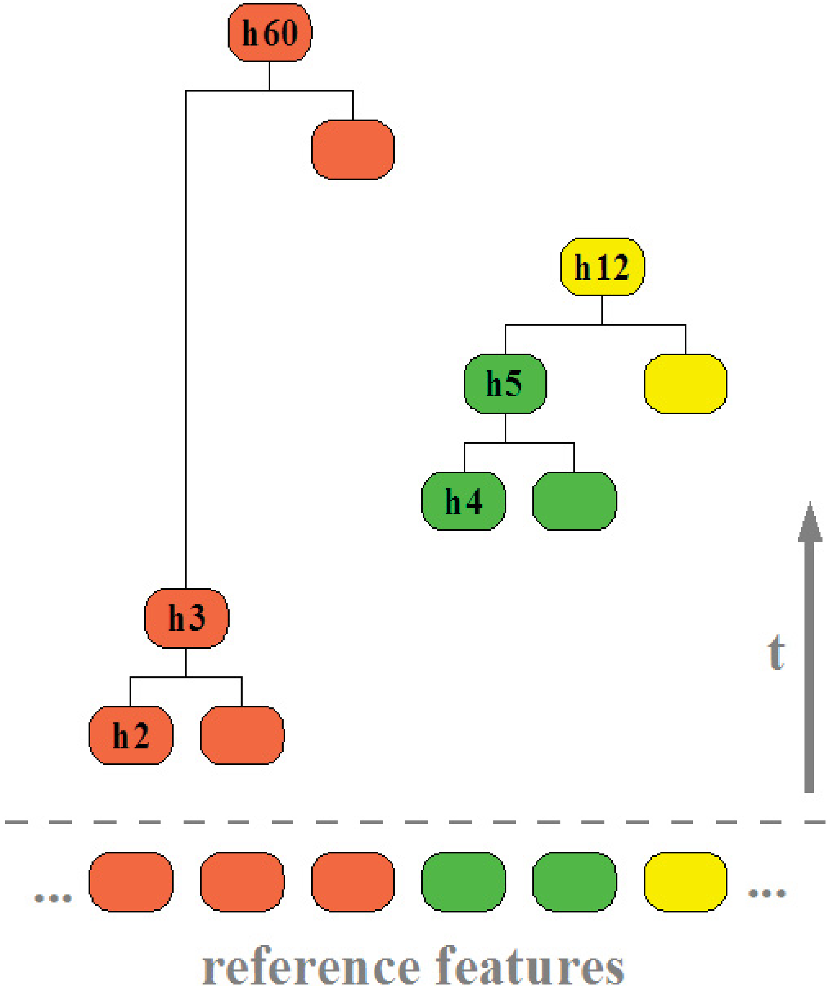
Example for a hierarchy



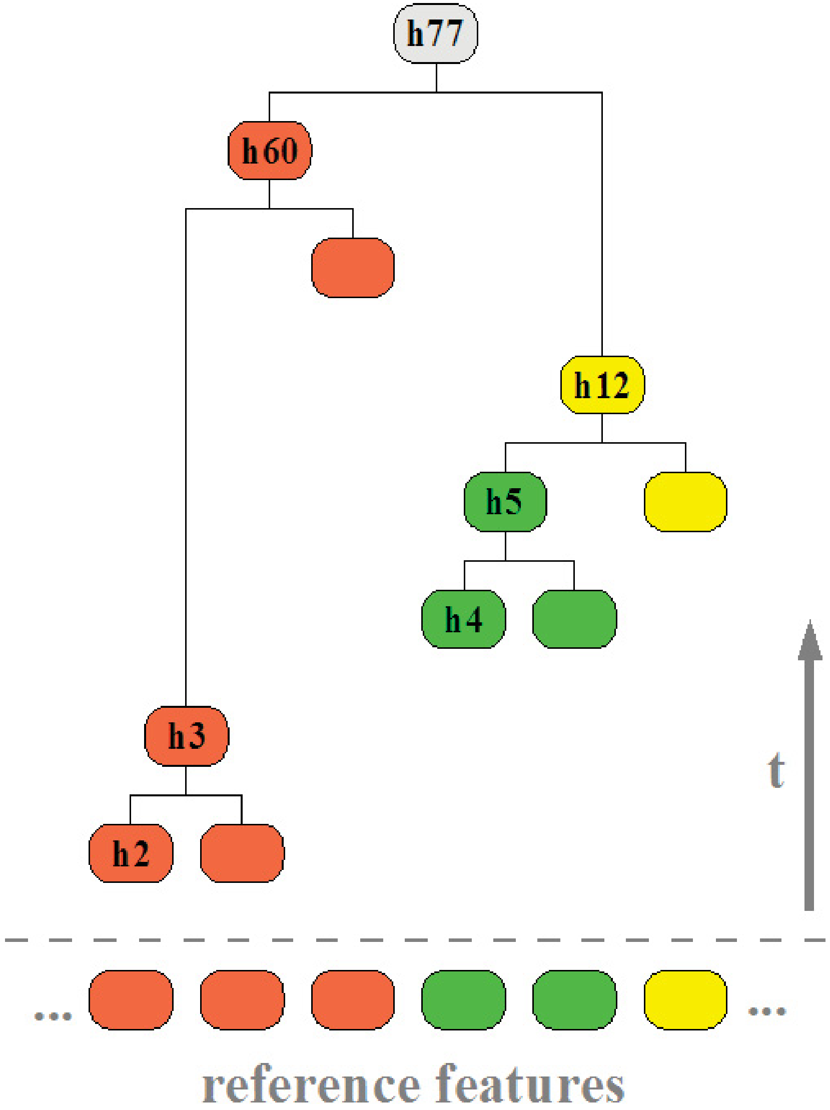
Example for a hierarchy



Example for a hierarchy



Example for a hierarchy



Outline

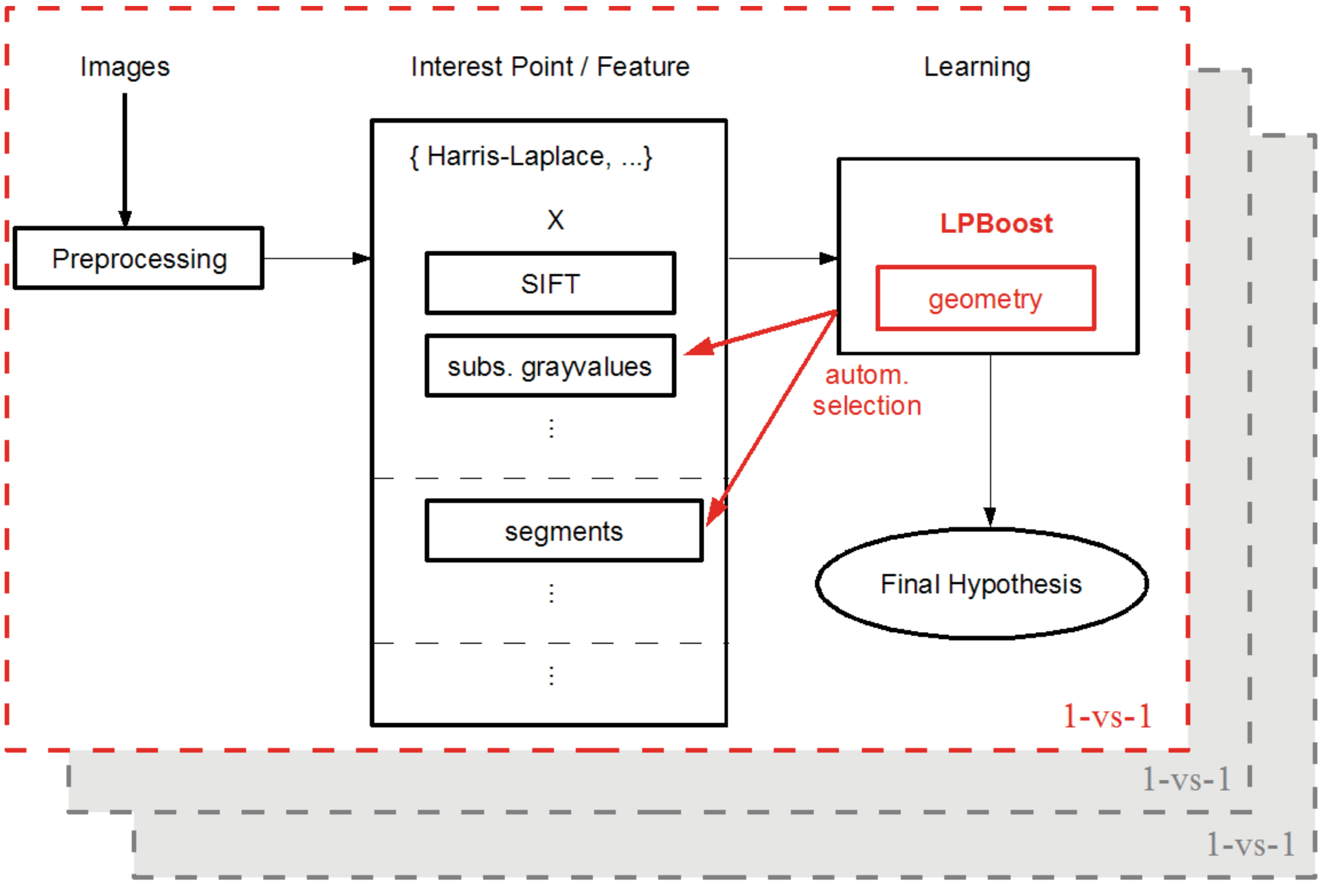
Introduction

Classification of images through boosting
- Feature types and preprocessing steps
- LPBoost
- Weak learner

Multiclass image classification
- Weight optimization method

Evaluation and results
- Xerox dataset
- PASCAL Visual object classes challenge 2006

Framework



- r categories $\Rightarrow r \cdot (r-1)$ classifiers

1. signed distance to the decision boundary of a 1-vs-1 classifier

$$\delta(x_i) = \sum_{t=1}^{T} \alpha_t h_t(x_i)$$

2. build $\mathbf{c}_i = \left(\delta_{1,2}(x_i), \delta_{2,1}(x_i), ..., \delta_{r-1,r}(x_i), \delta_{r,r-1}(x_i)\right)^T$

3. search weights $\mathbf{w}_l \ (l = 1, ..., r)$ such that

$$class(x_i) = \underset{l}{\text{argmax}} \ \mathbf{w}_l \cdot \mathbf{c}_i + b_l$$

⇒ e.g. formulate following SVM

$$\begin{aligned} \min \quad & \| (\mathbf{w}_1, ..., \mathbf{w}_r) \|^2 + C \cdot \sum_i \xi_i \\ s.t. \quad & \mathbf{w}_l \cdot \mathbf{c}_i + b_l \geq 1 - \xi_i, \quad && l = class(x_i) \\ & -\mathbf{w}_l \cdot \mathbf{c}_i - b_l \geq 1 - \xi_i, \quad && \forall l : l \neq class(x_i) \\ & \xi_i \geq 0 \quad && i = 1, \ldots, m, \\ & && l = 1, \ldots, r \end{aligned}$$

1. signed distance to the decision boundary of a 1-vs-1 classifier

$$\delta(x_i) = \sum_{t=1}^{T} \alpha_t h_t(x_i)$$

2. build $\mathbf{c}_i = \left(\delta_{1,2}(x_i), \delta_{2,1}(x_i), \ldots, \delta_{r-1,r}(x_i), \delta_{r,r-1}(x_i)\right)^T$

3. search weights $\mathbf{w}_l$ $(l = 1, \ldots, r)$ such that

$$class(x_i) = \operatorname*{argmax}_l \; \mathbf{w}_l \cdot \mathbf{c}_i + b_l$$

$\Rightarrow$ e.g. formulate following SVM

$$\begin{aligned} \min \quad & \|(\mathbf{w}_1, \ldots, \mathbf{w}_r)\|^2 + C \cdot \sum_i \xi_i \\ s.t. \quad & \mathbf{w}_l \cdot \mathbf{c}_i + b_l \geq 1 - \xi_i, & l = class(x_i) \\ & -\mathbf{w}_l \cdot \mathbf{c}_i - b_l \geq 1 - \xi_i, & \forall l : l \neq class(x_i) \\ & \xi_i \geq 0 & i = 1, \ldots, m, \\ & & l = 1, \ldots, r \end{aligned}$$

1. signed distance to the decision boundary of a 1-vs-1 classifier

$$\delta(x_i) = \sum_{t=1}^{T} \alpha_t h_t(x_i)$$

2. build $\mathbf{c}_i = \left(\delta_{1,2}(x_i), \delta_{2,1}(x_i), ..., \delta_{r-1,r}(x_i), \delta_{r,r-1}(x_i)\right)^T$

3. search weights $\mathbf{w}_l$ $(l = 1, ..., r)$ such that

$$class(x_i) = \operatorname*{argmax}_{l} \; \mathbf{w}_l \cdot \mathbf{c}_i + b_l$$

⇒ e.g. formulate following SVM

$$\begin{aligned} \min \quad & \| (\mathbf{w}_1, ..., \mathbf{w}_r) \|^2 + C \cdot \sum_i \xi_i \\ s.t. \quad & \mathbf{w}_l \cdot \mathbf{c}_i + b_l \geq 1 - \xi_i, & l = class(x_i) \\ & -\mathbf{w}_l \cdot \mathbf{c}_i - b_l \geq 1 - \xi_i, & \forall l : l \neq class(x_i) \\ & \xi_i \geq 0 & i = 1, \ldots, m, \\ & & l = 1, \ldots, r \end{aligned}$$

1. signed distance to the decision boundary of a 1-vs-1 classifier

$$\delta(x_i) = \sum_{t=1}^{T} \alpha_t h_t(x_i)$$

2. build $\mathbf{c}_i = \left(\delta_{1,2}(x_i), \delta_{2,1}(x_i), ..., \delta_{r-1,r}(x_i), \delta_{r,r-1}(x_i)\right)^T$
3. search weights $\mathbf{w}_l$ $(l = 1, ..., r)$ such that

$$class(x_i) = \underset{l}{\operatorname{argmax}}\ \mathbf{w}_l \cdot \mathbf{c}_i + b_l$$

⇒ e.g. formulate following SVM

$$\begin{aligned} \min \quad & \| (\mathbf{w}_1, ..., \mathbf{w}_r) \|^2 + C \cdot \sum_i \xi_i \\ s.t. \quad & \mathbf{w}_l \cdot \mathbf{c}_i + b_l \geq 1 - \xi_i, \qquad l = class(x_i) \\ & -\mathbf{w}_l \cdot \mathbf{c}_i - b_l \geq 1 - \xi_i, \qquad \forall l : l \neq class(x_i) \\ & \xi_i \geq 0 \qquad i = 1, \ldots, m, \\ & \qquad\qquad l = 1, \ldots, r \end{aligned}$$

Outline

Introduction

Classification of images through boosting
- Feature types and preprocessing steps
- LPBoost
- Weak learner

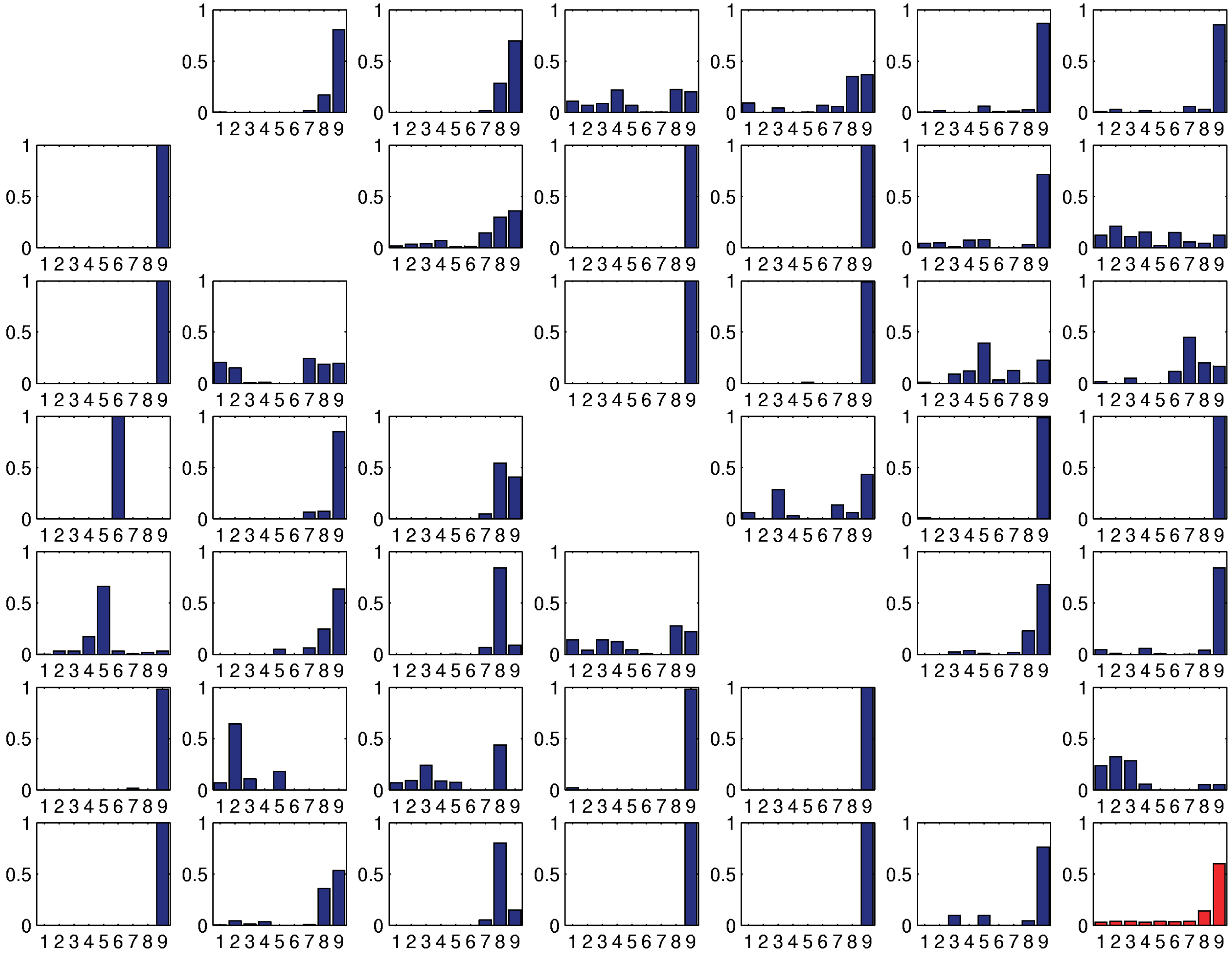
Multiclass image classification
- Weight optimization method

Evaluation and results
- Xerox dataset
- PASCAL Visual object classes challenge 2006

- 1774 real-world images
- $r = 7$ categories, uneven class distribution
  - faces
  - buildings
  - trees
  - cars
  - phones
  - bikes
  - books

1. preliminary 50-50-split of data
2. optimize parameter D (LPBoost) and C (SVM) upon test-set
3. fix parameters
4. stratified 10-fold cross-validation

Selected feature types (50-50-split + ’none’)



Weak hypotheses learned

- correct detections for buildings-vs-trees

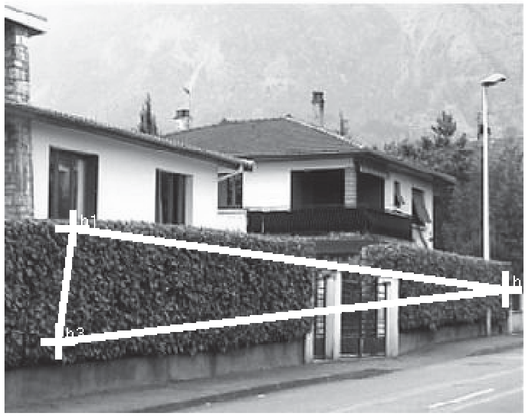
(a) (b) (c)

- ... misclassified examples

(d) (e) (f)

Weak hypotheses learned (cont.)

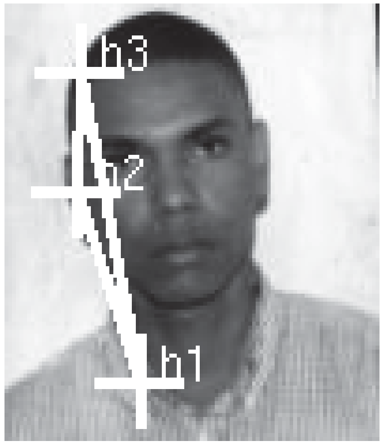
(g)



(h)

(i)

(j)



(k)

(l)

(m)

Accuracy upon 10CV

voting	geometry	parameter	mean	(std)
majority voting	none	–	64.25	(3.21)
majority voting	relations A	–	74.78	(2.92)
majority voting	relations B	–	75.08	(2.51)
[1]	–	–	85	n/a
SVM	none	$C = 0.2583$	90.60	(2.06)
SVM	relations A	$C = 0.7622$	90.90	(2.16)
SVM	relations B	$C = 0.1666$	**91.28**	(2.28)

Confusion matrix for ’relations B’ upon 10CV

$\longrightarrow$	faces	bldgs	trees	cars	phones	bikes	books
faces	**98.99**	0.66	1.33	8.47	2.64	0	0.71
bldgs	0	**70.66**	8.00	0	0	2.84	8.92
trees	0	10.00	**87.33**	0	0	0.83	1.42
cars	0.50	0	0.66	**84.09**	9.41	0	0
phones	0.50	0	0	7.42	**87.93**	0	0
bikes	0	2.67	2.66	0	0	**94.65**	2.14
books	0	16.00	0	0	0	1.66	**86.78**

Outline

Introduction

Classification of images through boosting
- Feature types and preprocessing steps
- LPBoost
- Weak learner

Multiclass image classification
- Weight optimization method

Evaluation and results
- Xerox dataset
- PASCAL Visual object classes challenge 2006

- 5304 real-world images
- $r = 10$ categories, uneven class distribution
- additional new feature: color-based segments
- data split: 25% training, 25% validation, 50% test
  1. learning 1-vs-1 on training set
  2. optimize parameter D (LPBoost) and C (SVM) for validation set
  3. fix parameters
  4. eval on test set using area under ROC-curve (AUC)

AUC on test set

(†)	INRIA		QMUL		XRCE	MUL
	Marsz.	Moosm.	HSLS	LSPCH		1vs1
bicycle	0.929	0.903	0.944	**0.948**	0.943	0.864
bus	**0.984**	0.933	**0.984**	0.981	0.978	0.945
car	0.971	0.957	**0.977**	0.975	0.967	0.928
cat	0.922	0.883	0.936	**0.937**	0.933	0.826
cow	0.938	0.895	0.936	0.938	**0.940**	0.789
dog	0.856	0.825	0.874	**0.876**	0.866	0.764
horse	0.908	0.824	0.922	**0.926**	0.925	0.733
motorbike	0.964	-	0.966	**0.969**	0.957	0.906
person	0.845	0.780	0.845	0.855	**0.863**	0.718
sheep	0.944	0.930	0.946	**0.956**	0.951	0.872

(†)Selection of all participants having a top rank w.r.t. the AUC reported at the PASCAL VOC challenge workshop, ECCV 2006

References and further reading I

- Gabriela Csurka, Cedric Bray, Christopher Dance, and Lixin Fan.
  Visual categorization with bags of keypoints.
  In European Conference on Computer Vision, ECCV'04, Prague, Czech Republic, May 2004.
- Michael Fussenegger, Andreas Opelt, Axel Pinz, and Peter Auer.
  Object recognition using segmentation for feature detection.
  In ICPR (3), pages 41–44, 2004.

References and further reading II

- Luc J. Van Gool, Theo Moons, and Dorin Ungureanu.
  Affine/photometric invariants for planar intensity patterns.
  In ECCV '96: Proceedings of the 4th European Conference on Computer Vision-Volume I, pages 642–651.
  Springer-Verlag, 1996.
- D.G. Lowe.
  Object recognition from local scale-invariant features.
  In Seventh International Conference on Computer Vision, pages 1150–1157, 1999.