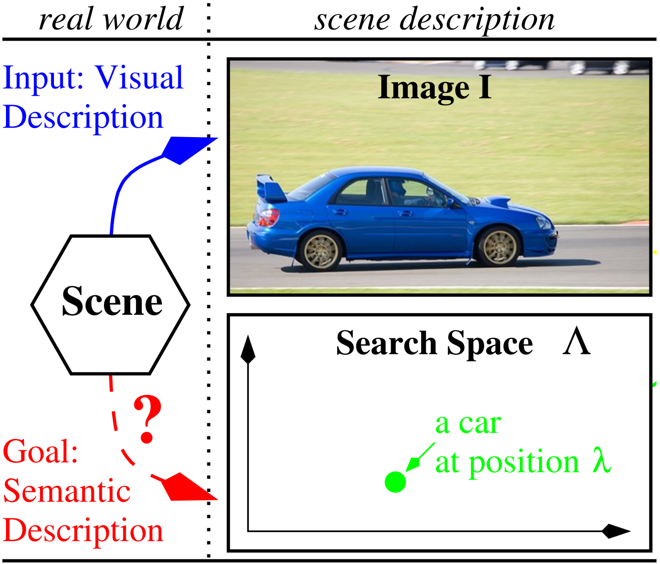
PRISM: PRincipled Implicit Shape Model

Alain Lehmann
Bastian Leibe and Luc Van Gool
September 9th, BMVC 2009, London

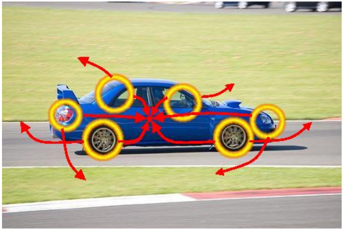
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Introduction: Object-Class Detection



Object-Class Detection Paradigms

Hough-Transform



- Implicit Shape Model (ISM) [Leibe et al., 2008]

+ natural voting

– constrained model (negative votes impossible)

– questionable argument (marginalisation over facts)

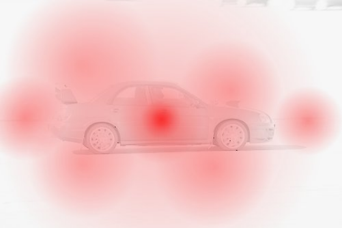
Sliding-Window

+ clean reasoning

+ flexible model (discriminative learning)

– “unnatural” algorithm

Object-Class Detection Paradigms

Hough-Transform	**Sliding-Window**
	

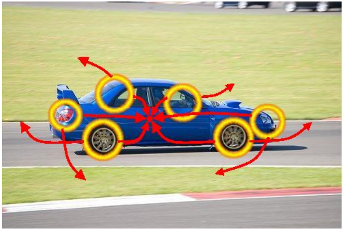
- Implicit Shape Model (ISM) [Leibe et al., 2008]

+ natural voting
– constrained model (negative votes impossible)
– questionable argument (marginalisation over facts)

+ clean reasoning
+ flexible model (discriminative learning)
– "unnatural" algorithm

Object-Class Detection Paradigms

Hough-Transform



- Implicit Shape Model (ISM) [Leibe et al., 2008]

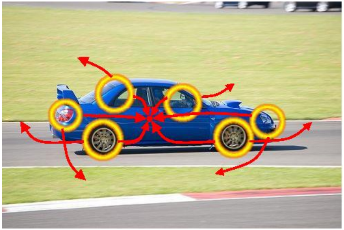
+ natural voting
– constrained model (negative votes impossible)
– questionable argument (marginalisation over facts)

Sliding-Window

+ clean reasoning
+ flexible model (discriminative learning)
– "unnatural" algorithm

Object-Class Detection Paradigms

Hough-Transform



Sliding-Window

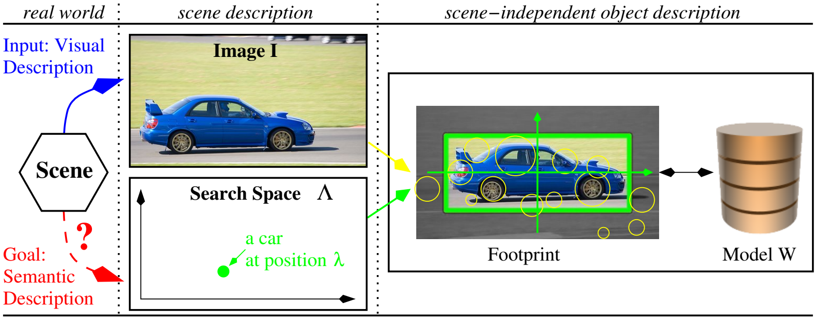
- Implicit Shape Model (ISM) [Leibe et al., 2008]

PRincipled Implicit Shape Model (PRISM)

+ natural voting — with — + clean reasoning

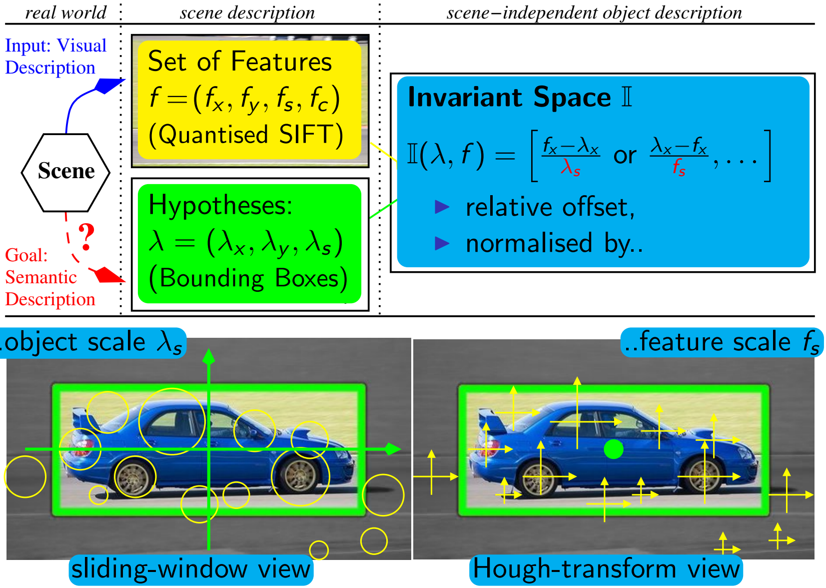
Hough-Transform	Sliding-Window
+ natural voting	+ clean reasoning
– constrained model (negative votes impossible)	+ flexible model (discriminative learning)
– questionable argument (marginalisation over facts)	– “unnatural” algorithm

PRISM: Sliding-Window View

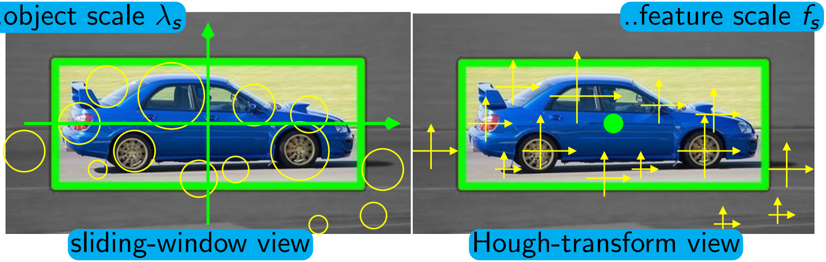


- fix a single hypothesis ⇒ crop out a sub-image
- compute scene-independent description ⇒ object footprint
- not explicitly defined in ISM

PRISM: Feature-Object Invariants



PRISM: Footprint & Score



Footprint Function $\phi(\lambda, I)$

- sum of dirac pulses, each
- encoding one invariant $\mathbb{I}(\lambda, f)$

Linear Object Model W

- compulsory for HT
- no other assumptions

Linear Score $$S(\lambda) = \underbrace{\langle \phi(\lambda, I), W \rangle}_{\textit{linear model}} = \sum_f \underbrace{W(\overbrace{f_c}^{\textit{visual word id}}, \mathbb{I}(\lambda, f))}_{\textit{point evaluations}}$$

Sliding-Window $\mapsto$ Hough-Transform

Mathematically

(Point-) Score	$S(\lambda) = \sum_f \overbrace{W(f_c, \mathbb{I}(\lambda, f))}^{\textit{point evaluation}}$	(fixed λ)
(Parallel-) Score	$S(\bullet) = \sum_f \underbrace{W(f_c, \mathbb{I}(\bullet, f))}_{\textit{voting pattern}}$	(**function of λ**)

Voting Pattern $W(f_c, \mathbb{I}(\cdot, f))$

- transformation of W defined by invariants $\mathbb{I}, f$
- no constraints on W, *i.e.* can be positive & negative
  $\Rightarrow$ ICCV'09

Algorithmically

SW: **for** $\lambda \in \Lambda$: **for** $f \in \mathcal{F}$: $S(\lambda) \mathrel{+}= W(f_c, \mathbb{I}(\lambda, f))$

HT: **for** $f \in \mathcal{F}$: **for** $\lambda \in \Lambda$: $S(\lambda) \mathrel{+}= W(f_c, \mathbb{I}(\lambda, f))$

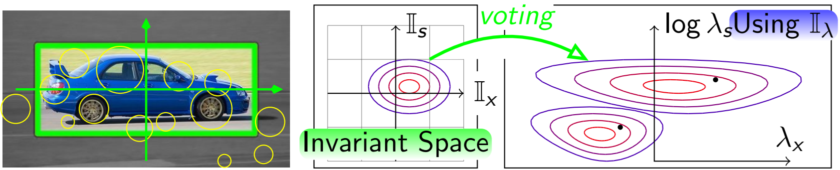
avoid: summing over $W(f_c, \mathbb{I}(\lambda, f)) = 0$

A Concrete Algorithm

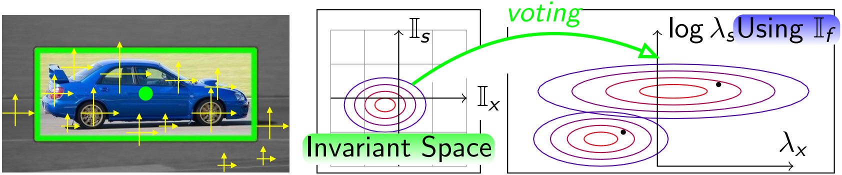
inspired by ISM (recovering ISM)

- set $W(c, \mathbb{I}) = p_c(\mathbb{I})$ (occurrence distribution)
- Gaussian mixture models (kernel density estimators)
  $\rightarrow$ better scaling (scale linear with training data)
- EM-based learning
- gradient-based search (mean-shift in ISM)

What happens to a Gaussian during voting?

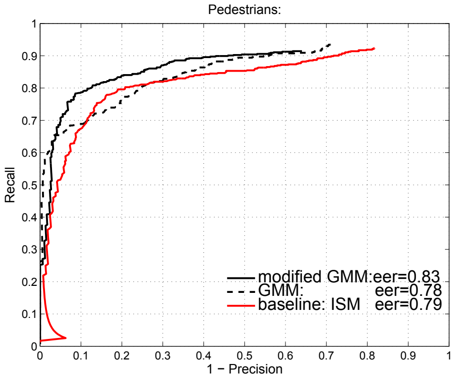


- object-centric invariant ⇒ non-linear distortion



- feature-centric invariant ⇒ simple translation & scaling
  ⇒ still a Gaussian ⇒ explicit voting possible
  ⇒ advantages ⇒ used in our experiments

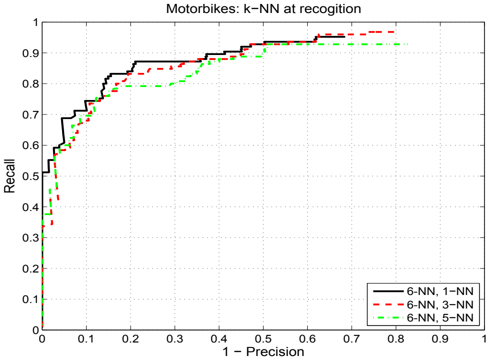
Results on *Toyota Pedestrian DB*



- ISM: baseline (solid)
- GMM $p_c(\mathbb{I})$ (dashed)
- modified GMM $\tilde{p}_c(\mathbb{I}) = \alpha_c \cdot p_c(\mathbb{I})$ (solid)
- state-of-the-art accuracy (without ISM's MDL verification)
  $\Rightarrow$ new theory does not impair quality

Soft-Matching..

- ..increases detection quality, but more costly
- ..is not needed during detection
  $\Rightarrow$ fast NN-matching sufficient (4$\times$ faster than 5NN)
- soft-matching S *blurs* the footprint ϕ
- $\langle S\phi, W\rangle = \langle \phi, S^T W\rangle \Rightarrow$ regularisation



Conclusion

PRISM: PRincipled Implicit Shape Model

- sound justification for Hough voting
  $\Rightarrow$ *resolve theoretical problems of ISM*
- object footprint & invariants
- duality: Hough-transform $\Leftrightarrow$ linear sliding-window
- soft-matching causes regularisation
  $\Rightarrow$ *fast NN-matching at detection time*

Conclusion

PRISM: PRincipled Implicit Shape Model

- sound justification for Hough voting
  ⇒ *resolve theoretical problems of ISM*
- object footprint & invariants
- duality: Hough-transform ⇔ linear sliding-window
- soft-matching causes regularisation
  ⇒ *fast NN-matching at detection time*

Feature-Centric Efficient Subwindow Search [ICCV'09]

- PRISM + discriminative learning + branch and bound
- advantages over ESS:
  - true-scale invariance
  - less memory usage
  - no on-line pre-processing

demo code available at
www.vision.ee.ethz.ch/lehmanal/iccv09

Questions?

PRISM: Full 1D Example