PRISM: PRincipled Implicit Shape Model

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

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich
Introduction: Object-Class Detection

**real world**

Input: Visual Description

**scene description**

Image I

Search Space \( \Lambda \)

Goal: Semantic Description

a car at position \( \lambda \)
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**

- 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**

- 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

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

  PRinciplced Implicit Shape Model (PRISM)
  + natural voting with + clean reasoning
  - constrained model (negative votes impossible)
  - questionable argument (marginalisation over facts)

Sliding-Window

- “unnatural” algorithm (discriminative learning)
PRISM: Sliding-Window View

<table>
<thead>
<tr>
<th>real world</th>
<th>scene description</th>
<th>scene-independent object description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input: Visual Description</td>
<td>Image I</td>
<td>Goal: Semantic Description</td>
</tr>
</tbody>
</table>

Scene

Search Space $\Lambda$

Footprint

- **fix a single hypothesis** $\Rightarrow$ crop out a sub-image
- **compute scene-independent description** $\Rightarrow$ object footprint
- **not explicitly defined in ISM**
**PRISM: Feature-Object Invariants**

<table>
<thead>
<tr>
<th><strong>real world</strong></th>
<th><strong>scene description</strong></th>
<th><strong>scene-independent object description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Input: Visual Description</td>
<td>Set of Features $f = (f_x, f_y, f_s, f_C)$ (Quantised SIFT)</td>
<td>Hypotheses: $\lambda = (\lambda_x, \lambda_y, \lambda_s)$ (Bounding Boxes)</td>
</tr>
<tr>
<td>Scene</td>
<td><strong>Invariant Space</strong> $\mathbb{I}$</td>
<td>$\mathbb{I}(\lambda, f) = \left[ \frac{f_x - \lambda_x}{\lambda_s} \text{ or } \frac{\lambda_x - f_x}{f_s}, \ldots \right]$</td>
</tr>
</tbody>
</table>

- relative offset, 
- normalised by..

- object scale $\lambda_s$
- feature scale $f_s$

- sliding-window view
- Hough-transform view
PRISM: Footprint & Score

Footprint Function $\phi(\lambda, I)$
- sum of dirac pulses, each
- encoding one invariant $\Pi(\lambda, f)$

Linear Score $S(\lambda) = \langle \phi(\lambda, I), W \rangle = \sum_{f} W( f_{c}, \Pi(\lambda, f))$

Linear Object Model $W$
- compulsory for HT
- no other assumptions
**Sliding-Window \(\leftrightarrow\) Hough-Transform**

Mathematically

(Point-) Score

\[
S(\lambda) = \sum_f W(f_c, \mathbb{I}(\lambda, f))
\]

(fixed \(\lambda\))

(Parallel-) Score

\[
S(\cdot) = \sum_f W(f_c, \mathbb{I}(\cdot, f))
\]

(function of \(\lambda\))

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 \(f \in \mathcal{F}\) : \(S(\lambda) += W(f_c, \mathbb{I}(\lambda, f))\)

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

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

inspired by ISM

- set $\mathcal{W}(c, \mathbb{I}) = p_c(\mathbb{I})$ (occurrence distribution)
- Gaussian mixture models (kernel density estimators)
  → 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 $\Rightarrow$ non-linear distortion

- feature-centric invariant $\Rightarrow$ simple translation & scaling
  $\Rightarrow$ still a Gaussian $\Rightarrow$ explicit voting possible
  $\Rightarrow$ advantages $\Rightarrow$ used in our experiments
Results on *Toyota Pedestrian DB*

- **ISM: baseline** (solid)
- **GMM** \( p_c(I) \) (dashed)
- **modified GMM** \( \tilde{p}_c(I) = \alpha_c \cdot p_c(I) \) (solid)
- **state-of-the-art accuracy** (without ISM’s MDL verification)
  \[ \Rightarrow \text{new theory does not impair quality} \]
Soft-Matching..

- Increases detection quality, but more costly.
- Is not needed during detection.
  \[ \Rightarrow \text{fast } \text{NN-matching sufficient} \quad (4 \times \text{faster than } 5\text{NN}) \]
- Soft-matching \( S \) blurs the footprint \( \phi \).
- \[ \langle S\phi, W \rangle = \langle \phi, S^T W \rangle \Rightarrow \text{regularisation} \]
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*
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

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

(a) \( \phi(I, \lambda^*) \)

(b) \( W \)

(c) \( f^1(\lambda^*, f^1) \)

(d) \( W(\lambda^*, f^1, f^1) \)

(e) \( S(\lambda) \)

Hough Transform

Votes

search space

image plane

visual word 1:

visual word 2: