Learning Hough Forest with Depth-Encoded Context for Object Detection

Tao Wang  Xuming He  Nick Barnes

NICTA & Australian National University

27 November 2012
Outline

- Problem & Background: Detecting Objects with Context and Depth, Class-specific Hough Forest
- Our Approach: Hough Forest with Depth-Encoded Context
- Experiments: Datasets and Evaluation
Problem: Indoor Object Detection

- Using a RGB-D sensor
- Encoding the surrounding context

Heavy occlusion
Typical setup
Limited visual cue
All the above
Context at Various Levels
Divvala et al., CVPR 2009

Local pixel context
- window surround, image neighbourhoods, object boundary/shape

2D scene “gist” context
- global image statistics

3D geometric context
- 3D scene layout, support surface, surface orientations, occlusions, contact points, etc.

Semantic context
- event/activity depicted, scene category, objects present in the scene and their spatial extents, keywords

More...
- photogrammetric context, illumination context, weather, geographic context, temporal context, cultural context
Depth-augmented Object Detection

Depth features and channel fusion
- Features: Histograms of Oriented Depths (HOD), Spin-images, etc.
- Early fusion v.s. late fusion
- Papers: Lai et al., ICRA 2011 and Spinello et al., ICRA 2012

Depth-encoded Hough voting (DEHV)
- Generative depth-estimation model
- 3D object model reconstruction
- Feedback loop with layout estimator and support region segmenter
- Papers: Sun et al., ECCV 2010 / BMVC 2010
Class-specific Hough Forest
Gall et al., CVPR 2009

Uncertainty measures

- Class-label uncertainty: \( U_1(A) = |A| \cdot Entropy\{c_i\} \)
- Offset uncertainty: \( U_2(A) = \sum_{i: c_i = 1} (d_i - d_A)^2 \)

Node-splitting criterion

- Minimising uncertainty (random choice)
- Search over all binary tests

Scoring function

\[
E(x|\mathcal{I}(y); T) = \frac{C_L}{|D_L|\sigma^2} \sum_{d \in D_L} \exp\left(-\frac{||y - x - d||^2}{2\sigma^2}\right)
\]
Overview to Our Approach

Training
- Object location $x$
- Image patches $L_i$

Testing
- Search over all patches $y$
- Cast votes $E(x|I(y); L_i)$
Offset variance minimisation on contextual patches
• \( U_2'(A) = \sum_{i:c_i=1} (d_i - d_{iA})^2 + \alpha \sum_{j:c_j=0} (d_j - d_{jA})^2 \)
• Only one term has main effect (low class-entropy criterion)

Overall vote score

\[
E(x|I(y); T) \propto E_f(x|I(y); T) + \beta E_b(x|I(y); T)
\]

\[
= \frac{C_L}{|D_L|\sigma^2} \sum_{d \in D_L} \exp \left( - \frac{||y - x - d||^2}{2\sigma^2} \right) + \\
\beta \frac{1 - C_L}{|D'_L|\sigma^2} \sum_{d \in D'_L} \exp \left( - \frac{||y - x - d||^2}{2\sigma^2} \right)
\]

weighted sum

votes from foreground

votes from context
Depth-augmented Context

Minimising 3D offset uncertainty

- \( U''_2(A) = \sum_{i:c_i=1} (d_i - d_{iA})^2 + \alpha \sum_{j:c_j=0} (d'_j - d'_{jA})^2 \)

- \( d_i = (x_i, y_i) \) and \( d'_j = (x_j, y_j, z_j) \), 2D for object, 3D for context
Depth-augmented Context (cont.)

Patch rescaling based on depth - training
Patch rescaling based on depth - testing

Depth-augmented Context (cont.)
Depth modulated contextual voting confidence
Overall scoring function

\[ E_b(x|I(y), s_x; T) = \frac{1 - C_L}{|D'_L|\sigma^2} \cdot \sum_{d \in D'_L, s \in S'_L, w \in W'_L} \delta(s_x = s) \cdot we^{-\frac{||y-x-d||^2}{2\sigma^2}} \]

- Matching the recorded scale: \( s_x \) is the associated scale of position \( x \), \( S_L \) is the target voting scales.
- Heat kernel based weighting: \( w_j = \exp\left(-\frac{||d_j||^2}{t}\right) \), where \( d_j = \sqrt{x_j^2 + y_j^2 + z_j^2} \).
Experiments

Specifications and Setup

- Datasets: Berkeley 3D Objects (v1) and NYU Depth (v2)
- 15 trees in a cascading fashion (5 trees per layer)
- 20000 binary tests per node, $d_{\text{Max}} = 15$, $N_{\text{min}} = 20$.
- Features: Lab colour channels, $1^{st}$ and $2^{nd}$ order image derivatives, HOG

Example Detections and Hough Maps
Performance Comparisons

Precision-recall curves and average precision

- Deformable Parts Model performs better than Class-specific Hough Forest
- Simple 2D geometric context provides limited benefit
- Depth augmented context helps build a better appearance codebook
- Each of the 3 components in our model has its own merit
Diagnostic Results

Benefits

- More sensible voting origin (some objects still difficult to localise)
- Encodes various aspects of context (e.g., supporting plane, typical scene setup)
- Hough maps generally smoother, robust to local noise
Thank you

Thanks for your time!

Questions or Comments?

For details please refer to the following publication:
Tao Wang, Xuming He, Nick Barnes. Learning Hough forest with depth-encoded context for object detection.