3-D Scene Analysis via Sequenced Predictions over Points and Regions

Xuehan Xiong

Daniel Munoz  Drew Bagnell  Martial Hebert

Carnegie Mellon
THE ROBOTICS INSTITUTE
Problem: 3D Scene Understanding
Solution: Contextual Classification
Classical Approach: Graphical Models

Graphical models

- Intractable inference
- Difficult to train
- Limited success

Belief propagation
Mean field
MCMC

Kulesza NIPS 2007
Wainwright JMLR 2006
Finley & Joachims ICML 2008

Anguelov, et al. CVPR 2005
Triebel, et al. IJCAI 2007
Munoz, et al. CVPR 2009

Fig. from Anguelov, et al. CVPR 2005
Classical Approach: Graphical Models

- Intractable inference
- Difficult to train
- Limited success

Belief propagation
Mean field
MCMC

Kulesza NIPS 2007
Wainwright JMLR 2006
Finley & Joachims ICML 2008

Anguelov, et al. CVPR 2005
Triebel, et. al. IJCAI 2007
Munoz, et al. CVPR 2009

Fig. from Anguelov, et al. CVPR 2005
Classical Approach: Graphical Models

- Intractable inference
- Difficult to train
- Limited success

Belief propagation
Mean field
MCMC

Kulesza
Wainwright
Finley & Joachims ICML 2008

Anguelov, et al. CVPR 2005
Triebel, et. al. IJCAI 2007
Munoz, et al. CVPR 2009

Fig. from Anguelov, et al. CVPR 2005
Our Approach: Inference Machines

• Train an inference **procedure**, not a model.
  – To encode spatial layout and long range relations
Our Approach: Inference Machines

• Train an inference **procedure**, not a model.
  – To encode spatial layout and long range relations

• Inference via sequential prediction

E.g. Viola-Jones 2001
Our Approach: Inference Machines

• Train an inference procedure, not a model.
  – To encode spatial layout and long range relations

• Inference via sequential prediction
Example features

\[ \hat{\mathbf{X}}_i^{(0)} : \text{point features} \]

\[ \hat{Y}^{(0)} = \text{LogReg}^{(0)}(X^{(0)}) \]
$\text{arg max}(\hat{Y}^{(0)})$

$\hat{X}_i^{(0)}$: point features
point features

Contextual features

$\bar{X}_i^{(0)}$

top  mid  bottom
\[
\hat{Y}^{(1)} = \logreg^{(1)}(X^{(1)})
\]
\[
\hat{Y}^{(1)} = \text{LogReg}^{(1)}(X^{(1)})
\]

\[
\arg\max(\hat{Y}^{(1)})
\]
\[ \hat{Y}^{(1)} = \text{LogReg}^{(1)}(X^{(1)}) \]
\[ \hat{\mathbf{X}}^{(2)}(i) \]

\[ \hat{Y}^{(2)} = \text{LogReg}^{(2)}(X^{(2)}) \]
Local features only
Round 1
Round 2
Round 3
Create regions

Level 2

Level 1
$\bar{X}_i^{(2)}$,

point features

Point level

Region level
With Regions
Learned Relationships

\[ \bar{x}_i : \text{point features} \]

Neighbor contextual feature

Learned weights

top  mid  bottom

veg

top  middle  bottom
Learned Relationships

Neighbor contextual feature

Learned weights
Experiments

• 3 large-scale datasets
  – CMU (26M), Moscow State (10M), Univ. Wash (10M)

• Multiple classes (4 to 8)
  – car, building, veg, wire, fence, people, trunk, pole, ground, street sign

• Different sensors
  – SICK (ground), ALTM 2050 (aerial), Velodyne (ground)

• Comparisons
  – Graphical models, exemplar based
Quantitative Results

Average F1 Score

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>0.8</td>
<td>0.5</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td>Related Work</td>
<td>0.7</td>
<td>0.6</td>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>LogReg</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
<td>0.3</td>
</tr>
</tbody>
</table>


* Use additional semi-supervised data not leveraged by other methods.
CMU Dataset

Ours

Max Margin CRF [1]

CMU Dataset

Ours

Max Margin CRF [1]

CMU Dataset

Ours

Max Margin CRF [1]
Moscow State Dataset

Ours

Logistic regression
Conclusion

• Simple and fast approach for scene labeling
  – No graphical model
  – Labeling via 5x logistic regression predictions

• Support flexible contextual features
  – Learning rich relationships
Thank you! And Questions?

- Acknowledgements
  - US Army Research Laboratory, Collaborative Technology Alliance
  - QinetiQ North America Robotics Fellowship