3D Object Recognition and Scene Understanding from RGB-D Videos

Yu Xiang

Postdoctoral Researcher

University of Washington

Notice and State Estimation Lab



Act in the 3D World



Intelligent System

Understand the 3D World

• Navigation

•

...

• Manipulation





Geometry
 Free space
 Surface

Semantics
 Objects
 Affordances

Recognize Objects in 3D



Semantics
 3D Location
 3D Orientation

Lula Robotics

(H) baxter

6D Object Pose Estimation



Related Work: Feature-based methods



- Lowe, ICCV'99
- Rothganger et al., IJCV'06
- Savarese & Fei-Fei, ICCV'07
- Collet et al., IJRR'11

Related Work: Template-based methods



Texture-less objects
 Symmetry objects
 Occlusion

- Gu & Ren, ECCV'10
- Hinterstoisser et al., ACCV'12
- Xiang & Savarese, CVPR'12
- Cao et al., ICRA'16



Texture-less objects
 Symmetry objects
 Occlusion

Xiang et al. under review

Decouple 3D Translation and 3D Rotation



• 3D Translation



2D center $\mathbf{c} = (c_x, c_y)^T$ Distance T_z

3D Rotation



PoseCNN: Semantic Labeling



PoseCNN: 3D Translation Estimation



PoseCNN: Center Estimation with RANSAC





PoseCNN: 3D Rotation Regression



The YCB-Video Dataset



21 YCB Objects



92 Videos, 133,827 frames

6D Pose Evaluation Metric



Results on the YCB-Video Dataset

















Network

Network + ICP

Network + ICP + Multi-view

Semantic Mapping (Semantic SLAM)



✓ Geometry

✓ Semantics

✓ Camera poses

LATELINE

Semantic Mapping with Data Associated Recurrent Neural Networks (DA-RNNs)



Related Work: 3D Scene Reconstruction



KinectFusion

Geometry
 Data Association
 Semantics

- Newcombe et al., ISMAR'11
- Henry et al., IJRR'12, 3DV'13
- Whelan et al., RSS Workshop'12, RSS'15
- Keller et al., 3DV'13

Related Work: Semantic Labeling



Geometry Data Association Semantics

- Long et al., CVPR'12
- Zheng et al., ICCV'15
- Chen et al., ICLR'15
- Badrinarayanan et al., CVPR'15

Related Work: Semantic Mapping



SemanticFusion

Geometry
 Data Association
 Semantics

- Salas-Moreno et al., CVPR'13
- McCormac et al., ICRA'17
- Bowman et al. ICRA'17

Our Contribution: DA-RNN



Single Frame Labeling with FCNs



Results on RGB-D Scene Dataset [1]



Video Semantic Labeling with DA-RNNs



Data Association from KinectFusion



Data Associated Recurrent Units (DA-RUs)



Weighted Moving Averaging with learnable parameters

Results on RGB-D Scene Dataset [1]

FCN DA-RNN [1] K. Lai, L. Bo and D. Fox. Unsupervised feature learning for 3D scene labeling. In ICRA'14.

Experiments: Datasets

- RGB-D Scene Dataset [1]
 - 14 RGB-D videos of indoor scenes
 - 9 object classes

- ShapeNet Scene Dataset [2]
 - 100 RGB-D videos of virtual table-top scenes
 - 7 object classes

[1] K. Lai, L. Bo and D. Fox. Unsupervised feature learning for 3D scene labeling. In ICRA'14.
[2] Chang et al., ShapeNet: an information-rich 3D model repository. arXiv preprint arXiv:1512.03012, 2015.

Methods	FCN [1]
Background	94.3
Bowl	78.6
Сар	61.2
Cereal Box	80.4
Coffee Mug	62.7
Coffee Table	93.6
Office Chair	67.3
Soda Can	73.5
Sofa	90.8
Table	84.2
MEAN	78.7

RGB-D Scenes

Metric: segmentation intersection over union (IoU)

[1] J. Long, E. Shelhamer and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR'15.³⁵

Methods	FCN [1]	Our FCN
Background	94.3	96.1
Bowl	78.6	87.0
Сар	61.2	79.0
Cereal Box	80.4	87.5
Coffee Mug	62.7	75.7
Coffee Table	93.6	95.2
Office Chair	67.3	71.6
Soda Can	73.5	82.9
Sofa	90.8	92.9
Table	84.2	89.8
MEAN	78.7	85.8

RGB-D Scenes

Metric: segmentation intersection over union (IoU)

[1] J. Long, E. Shelhamer and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR'15.³⁶

Methods	FCN [1]	Our FCN	Our GRU-RNN
Background	94.3	96.1	96.8
Bowl	78.6	87.0	86.4
Сар	61.2	79.0	82.0
Cereal Box	80.4	87.5	87.5
Coffee Mug	62.7	75.7	76.1
Coffee Table	93.6	95.2	96.0
Office Chair	67.3	71.6	72.7
Soda Can	73.5	82.9	81.9
Sofa	90.8	92.9	93.5
Table	84.2	89.8	90.8
MEAN	78.7	85.8	86.4

RGB-D Scenes

Metric: segmentation intersection over union (IoU)

[1] J. Long, E. Shelhamer and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR'15.³⁷

Methods	FCN [1]	Our FCN	Our GRU-RNN	Our DA-RNN
Background	94.3	96.1	96.8	97.6
Bowl	78.6	87.0	86.4	92.7
Сар	61.2	79.0	82.0	84.4
Cereal Box	80.4	87.5	87.5	88.3
Coffee Mug	62.7	75.7	76.1	86.3
Coffee Table	93.6	95.2	96.0	97.3
Office Chair	67.3	71.6	72.7	77.0
Soda Can	73.5	82.9	81.9	88.7
Sofa	90.8	92.9	93.5	95.6
Table	84.2	89.8	90.8	92.8
MEAN	78.7	85.8	86.4	90.1

Metric: segmentation intersection over union (IoU)

RGB-D Scenes

[1] J. Long, E. Shelhamer and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR'15.³⁸

Methods	FCN [1]	Our FCN	Our GRU-RNN	Our DA-RNN	No Data Association
Background	94.3	96.1	96.8	97.6	69.1
Bowl	78.6	87.0	86.4	92.7	3.6
Сар	61.2	79.0	82.0	84.4	9.9
Cereal Box	80.4	87.5	87.5	88.3	14.0
Coffee Mug	62.7	75.7	76.1	86.3	4.5
Coffee Table	93.6	95.2	96.0	97.3	68.0
Office Chair	67.3	71.6	72.7	77.0	13.6
Soda Can	73.5	82.9	81.9	88.7	5.9
Sofa	90.8	92.9	93.5	95.6	35.6
Table	84.2	89.8	90.8	92.8	20.1
MEAN	78.7	85.8	86.4	90.1	24.4

Metric: segmentation intersection over union (IoU)

RGB-D Scenes

[1] J. Long, E. Shelhamer and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR'15.³⁹

Our DA-RNN

RGB Image

Our FCN

Experiments: Analysis on Network Inputs

RGB Images

Depth Images

Semantic Mapping

42

Conclusion & Future Work

- 6D Object Pose Estimation and Semantic Mapping from RGB-D videos
- Deep neural networks with geometric representations

Acknowledgements

Thank you!