SemanticPaint

Adam Kosiorek

Advisor: M.Eng. Keisuke Tateno

10.12.2015

Outline

Introduction

Introduction

State of the Art

Acquisition and Reconstruction

Who	What	How
Levoy et. al. 2000	world-scale 3D models	offline, from online images
Newcombe et. al. 2011	online 3D scanning	KinectFusion
Pradeep et.al. 2013	3D reconstructiona with 1 RGB camera	sparse tracking and stereo reconstruction on par with KinectFusion

Scene Understanding

Who	What	How
Kim et. al. 2013	reconstruction segmentation	Voxel-based CRF with visibility contraints
Herbst et. al. 2014	registration segmentation	online change detection
Valentin et. al. 2013	inference on mesh from TSDF	RGB and geom. features CRF segmentation

Pipeline

Voxel Oriented Patch features

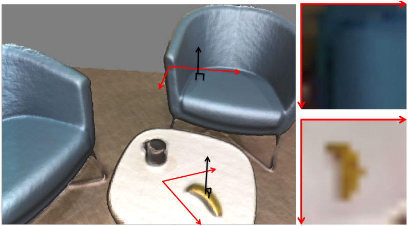


Figure: Colours shown in RGB for illustration purposes.

$(\mathbf{p} - \mathbf{p}_i) \cdot (n)_i = 0$
$r \times r$, $r = 13px$ with $10\frac{mm}{pixel}$
CIELab
Rotated to dominant gradient direction

Random Forest

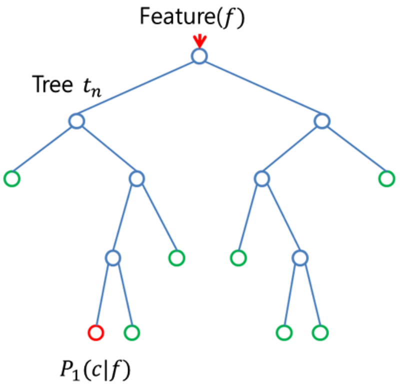


Figure: Single tree

bagged trees
bootstraped data
voting for final result
greedy training
off-line, all data at once
$(i, l) \in \mathcal{S}$ - (voxel, label) pairs
$f(i, \theta)$ - split functions
Θ - distribution of split functions
$P_F(x_i = l|\mathbf{D})$

Streaming Random Forest

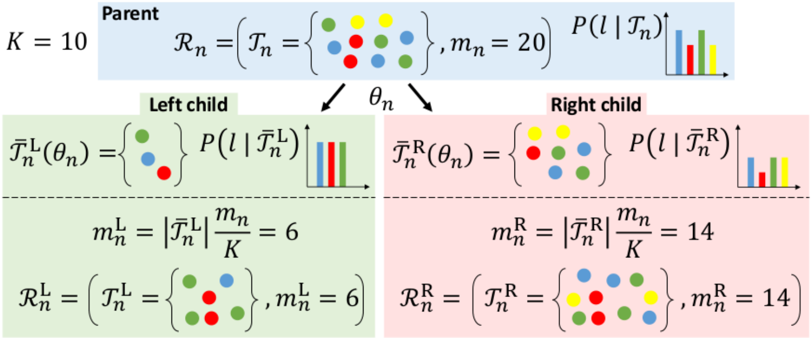
Information Gain:

$$G(S, S^L, S^R) = H(S) - \sum_{d \in \{L,R\}} \frac{|S^d|}{|S|} H(S^d) \tag{1}$$

Shannon Entropy:

$$H(S) = - \sum_{(l,i) \in S} p(c_i = l) \log p(c_i) \tag{2}$$

SRF - Reservoir Splitting



Dynamic Conditional Random Field

$$P(\mathbf{x}|\mathbf{D}) = \prod_{i \in \mathcal{V}} \left(\psi_i(x_i) \prod_{j \in \mathcal{E}_i} \psi_{ij}(x_i, x_j) \right) \tag{3}$$

$$E_t(\mathbf{x}) = \sum_{i \in \mathcal{V}} \left(\phi_i(x_i) + \sum_{j \in \mathcal{E}_i} \phi_{ij}(x_i, x_j) \right) + K \tag{4}$$

CRF - User Interactions

Touching:

$$\phi_i(l) = \begin{cases} 0 & \text{if } l = l_S \\ \infty & \text{otherwise} \end{cases} \tag{5}$$

Encircling:

$$\phi_i(l) = \begin{cases} \log P_E(fg|\mathbf{a}_i) & \text{if } l = \text{fg} \\ \log(1 - P_E(fg|\mathbf{a}_i)) & \text{if } l = \text{bg} \end{cases} \tag{6}$$

CRF - Predictions and Smoothnes

Predictions:

$$\phi_i(l) = -\log P_F(x_i = l|\mathbf{D}) \tag{7}$$

Smoothnes:

$$\phi_{ij}(x_i, x_j) = \theta_p e^{-||\mathbf{p}_i - \mathbf{p}_j||} + \theta_a e^{-||\mathbf{a}_i - \mathbf{a}_j||} + \theta_n e^{-||\mathbf{n}_i - \mathbf{n}_j||} \tag{8}$$

Mean-Field Inference

$P(\mathbf{x})$ approximated by $Q(\mathbf{x})$ under $KL(Q||P)$:

$$Q_i^t(l) = \frac{1}{Z_i} e^{M_i(l)},\ t = 1, \ldots, T \tag{9}$$

$$M_i(l) = \phi_i(l) + \sum_{l' \in \mathcal{L}} \sum_{j \in \mathcal{N}(i)} Q_j^{t-1}(l') \phi_{ij}(l, l') \tag{10}$$

Frame at time t initialized with:

$$\widetilde{Q}_i^t(x_i) = \gamma Q_i^{t-1}(x_i) + (1-\gamma) P_F^{t-1}(x_i = l|\mathbf{D}),\ \gamma \in [0,1] \tag{11}$$

Results

Segmentation

Table: Segmentation Results

Component	LivingRoom	Bedroom	Kitchen	Desk	Average
User Interaction	99.35%	97.61%	96.09%	97.73%	97.7%
Forest Prediction	94.57%	88.31%	82.58%	90.29%	88.94%
Final Inference	96.26%	95.19%	90.69%	95.55%	94.42%

Features

Table: Feature Comparison

Feature	LivingRoom	Bedroom	Kitchen	Desk	Average
VOP	**94.57%**	**88.31%**	82.58%	**90.29%**	**88.94%**
△ RGB mean	80%	71.84%	76.29%	73.42%	75.39%
Depth Probe	77.54%	61.79%	**84.9%**	68.9%	73.06%
Color Probe	56.39%	65.68%	60.77%	60.74%	60.9%
SURF	43.74%	67.12%	57%	58.13%	56.5%
SPIN	58.77%	43.22%	48.41%	36.1%	46.63%

Streaming Random Forest

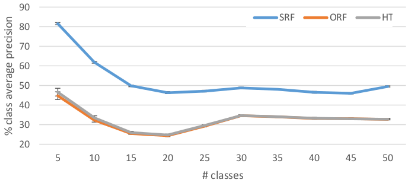


Figure: Average Precision

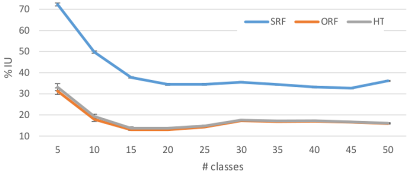


Figure: Intersection/Union

Data:
300 objects
51 classes
full revolution
3 points of view

SRF - Streaming Random Forest
ORF - Online Random Forest
HT - Hoeffding Tree

Discussion and Outlook

Summary

- customized models of 3D enviornments
- fully interactive
- online and real time
- no pretraining

Failures

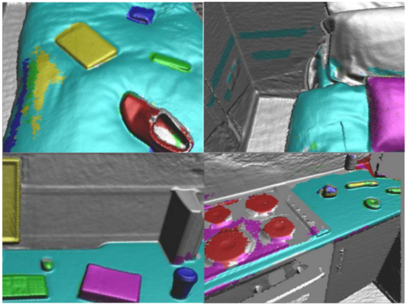


Figure: Failure cases.

- bleeding
- illumination change
- viewpoint change

Future Work

- class priors for different enviornments
- priors for class properties (vertical walls)
- discriminative geometrical features
- outdoor enviornments
- better scalability

References

- Roberts, L. G. 1963. Machine perception of three-dimensional solids. Ph.D. thesis, Massachusetts Institute of Technology.
- Kim, B.-S. et. al. 2013. 3D scene understanding by voxel-CRF. In Proc. ICCV.
- Pradeep, V. et. al. 2013. Monofusion: Real-time 3D reconstruction of small scenes with a single web camera. In Proc. ISMAR.
- Herbst, E. et.al. 2014. Toward online 3-d object segmentation and mapping. In IEEE International Conference on Robotics and Automation (ICRA).
- Valentin, J. P. et. al. 2013. Mesh based semantic modelling for indoor and outdoor scenes. In Proc. CVPR.
- Levoy , M. et. al. 2000. The digital Michelangelo project: 3D scanning of large statues. In Proc. SIGGRAPH. ACM.
- Newcombe , R. A. et. al. 2011. KinectFusion: Real-time dense surface mapping and tracking. In Proc. ISMAR.
- Curless , B. et. al. 1996. A volumetric method for building complex models from range images. In Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. ACM, 303312.
- Niessner , M. et. al. 2013. Real-time 3D reconstruction at scale using voxel hashing. ACM TOG 32, 6

References cont'd

- Saffari , A. et. al. 2009. On-line random forests. In IEEE ICCV Workshop.
- Vitter , J. S. 1985. Random sampling with a reservoir. ACM TOMS 11, 1.
- Lower , D. G. 1999. Object recognition from local scale-invariant features. In Proc. ICCV.
- Lafferty , J. et. al. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
- Ktahenbl, P. et. al. 2011. Efficient inference in fully connected CRFs with Gaussian edge potentials. In NIPS.
- Koller , D. et.al , N. 2009. Probabilistic Graphical Models: Principles and Techniques. MIT Press
- Domingos, P. et. al. 2000. Mining high-speed data streams. In Proc. SIGKDD.
- Lai, K. et. al. 2011. A large-scale hierarchical multi-view rgb-d object dataset. In Proc. ICRA.
- Valentin, J. et. al. 2015. SemanticPaint: Interactive 3D Labeling and Learning at your Fingertips. SIGGRAPH.