Autoscanning for Coupled Scene Reconstruction and Proactive Object Analysis

Kai Xu, Hui Huang, Yifei Shi, Hao Li, Pinxin Long, Jianong Caichen, Wei Sun, Baoquan Chen

Shenzhen VisuCA Key Lab / SIAT
National University of Defense Technology
Shandong University
Outline

- Background & motivation
- Method
- Results & evaluation
- Conclusion
- Future works
Background & motivation

- Fast development of
- Commodity depth camera
- Real-time reconstruction

KinectFusion [Nießner et al. 2013]

Microsoft Kinect
Background & motivation

- **Real-time reconstruction** allows for ...

  Instant visual feedback

  Online structural analysis [Zhang et al. 2014]

  Online analysis!!
What difference does it make?

- Real world
- Point cloud
- 3D models

Scanning → Reconstruction → Analysis
- Why is it so exciting?

**Online analysis guides autonomous scanning !!**

- Online analysis tells the robot *where to scan* and *when to stop*.
- Analysis will benefit from more and more complete reconstruction.
Background & motivation

- What’s more exciting?

**Online analysis guides autonomous scanning!!**
Autoscanning replacing human scanning!

- What’s more exciting?

Online analysis guides autonomous scanning!!
How is the scene composited?
Object extraction
Background & motivation

[Zhang et al. 2014]
Scene segmentation is not easy!

- Low quality depth maps
- Drifting issue with KinectFusion
- Objects in varying scales
- Cluttered objects
Our solution:
Proactive analysis
Method
Problem statement

Input:
- Raw KinectFusion recon.

Output:
- Full scene reconstruction
- Object-level segmentation
- Object-wise fidelity
Problem statement

Input:
- Raw KinectFusion recon.

Output:
- Object-aware recon.
Method overview

Initialization

Validation

Object analysis
Algorithms

Object-level segmentation

Entropy-based validation

Validation

Object analysis
Algorithm I
Object-level segmentation
Object-level segmentation - Pipeline

Over-segmentation

Graph-cut and graph contraction

Detected objects

Patch graph

Object graph
Object-level segmentation

**Generating** object hypothesis

- Multiple binary graph-cuts

**Selecting** object hypothesis

- One multi-class graph-cut
Generating object hypotheses

Seed patch

Input scene

Binary graph-cut [Golovinskiy et al. 2009]

Foreground background

A hypothetical object

SA2015.SIGGRAPH.ORG
Generating object hypotheses
Generating object hypotheses
Generating object hypotheses

So we get three hypothetical objects:

The next step is to label the scene using the three labels above ...
Selecting object hypotheses

Selecting the most prominent hypotheses with a multi-class graph-cut
Selecting object hypotheses

Hypothesis a

Hypothesis b

Hypothesis c

Multi-class labeling

\[ E(L) = \sum_{u \in V_p} E_d(l_u; P_u) + \sum_{(u,v) \in E_p} E_s(l_u, l_v) \]

multi-class graph-cut
Selecting object hypotheses

\[ E(L) = \sum_{u \in V_p} E_d(l_u; P_u) + \sum_{(u,v) \in E_p} E_s(l_u, l_v) \]

\[ E_d(l_u = \alpha; P_u) = -\ln P(u \leftarrow \alpha) \]
Selecting object hypotheses

Hypothesis a

Hypothesis b

Hypothesis c

Multi-class labeling

\[
E(L) = \sum_{u \in V_p} E_d(l_u; P_u) + \sum_{(u, v) \in E_p} E_s(l_u, l_v)
\]

\[
E_s(l_u \neq l_v) = \eta(1 - \cos(\theta_{uv}))
\]
Object-level segmentation

Ours

RANSAC
Learn segmentation from history ...

Can we learn from such case, and apply the learned knowledge in segmenting the rest part of the scene?
Online Learning of cut cost

\[ E(L) = \sum_{u \in V_p} E_d(l_u; P_u) + \sum_{(u,v) \in E_p} E_s(l_u, l_v) \]

\[ E_s(l_u \neq l_v) = 1 - p(l_u \neq l_v | x(P_u, P_v)) \]

Predicted by learned SVM classifier
Online Learning of cut cost

Feature representation

\[ \mathbf{x}(P_u, P_v) \]

Linear classifier
Multiple Kernel Learning
Online Learning of cut cost

Collect training examples from pushes:

Positive examples: must cut

Negative examples: cannot cut
Online Learning of cut cost

- Incremental update of the classifier
Online Learning of cut cost
Algorithm II

Entropy-based proactive validation
Where to push?

Object-level segmentation
Validation pipeline

- Uncertainty estimation
- Next-Best-Push
- Movement tracking
- Scan refinement
- New recon.
- Next-Best-View
Validation pipeline

Uncertainty estimation → Next-Best-Push → Movement tracking

New recon. ← Scan refinement ← Next-Best-View
Information gain maximization

- Maximize the information gain

\[ I(X) = H(X) - H'(X) \]

info. gain \hspace{2cm} entropy \hspace{2cm} entropy

before push \hspace{2cm} after push

How to measure the entropy?

Shannon entropy

\[ - \sum_i p(x_i) \log p(x_i) \]
How to measure entropy?

- Our goal is object-aware reconstruction
- Two aspects:
  - Uncertainty of Segmentation
  - Uncertainty of Reconstruction

\[ H = H(\ ,\ ) \quad \text{Joint entropy} \]

\[ H(S, R) = H(S) + H(R|S) \quad \text{Conditioned entropy} \]
How to measure entropy?

\[ H(S, R) = H(S) + H(R | S) \]

\[ H(S) = -2 \sum_{e \in \mathcal{E}^p} p_c(e) \log p_c(e) \]

Cut probability
How to measure entropy?

\[ H(S, R) = H(S) + H(R | S) \]

\[ H(R) = - \sum_{s \in \Omega} g(c(s)) \log g(c(s)) \]

\[ c(s) = \Gamma(s) \cdot n_s \]

recon. *certainty* = sharpness of Poisson field

[Kazhdan et al. 2006]
How to measure entropy?

Given an object segmentation, how much uncertainty is there in the object-wise reconstruction?
How to measure entropy?

\[ H(S, R) = H(S) + H(R|S) \]

\[ H(R|S) = - \sum_{e \in \mathcal{E}(S)} \sum_{s \in \hat{\Omega}(e)} p_e(e) g(c(s)) \log g(c(s)) \]
How to select the Next Best Push?

Search over all push points:

$$u^* = \arg \max_u I(S, R|\langle p_u, d_u \rangle)$$

$$I(S, R|\langle p_u, d_u \rangle) = H(S, R) - H'(S, R|\langle p_u, d_u \rangle)$$

**Posterior entropy**

before push is performed

after push is performed
How to compute the posterior entropy

Given a push, how much uncertainty can be reduced by it?

How many occluded iso-points can be exposed by it?
Plot of information gain

Object-level segmentation

Information gain
Results and evaluation
Results

- Test on real-life scenes
Results

- Test on real-life scenes
Quantitative evaluation

- Ground-truth data
A quantitative measure

- **bilateral support** between ground-truth and reconstructed surfaces

\[
\Pi(S, T) = \frac{\sum_{(p_i, n_i) \in T} \pi(p_i, n_i, S)}{|T|} + \frac{\sum_{(p_j, n_j) \in S} \pi(p_j, n_j, T)}{|S|}
\]
Quantitative evaluation

- Bilateral support

reconstructed surface

ground-truth surface

Imperfect reconstruction
Quantitative evaluation

- Bilateral support

Reconstructed with *under*-segmentation
Quantitative evaluation

- Bilateral support

Reconstructed with over-segmentation
Quantitative evaluation

- Compare to [Zhang et al. 2014]
Quantitative evaluation

- Comparing to alternative robot active analysis
Limitations

- Extraction rate: 50%~80%
- Push scheduling
- Vertical support
- Object stack
- Non-rigid objects
Conclusion

- A new paradigm of 3D acquisition
  - Autonomous scene scanning
    - Scene level: Large scale
    - Object level: Detailed
  - Proactive validation
    - Physically-validated segmentation
Future works

- Multiple robots?
Future works

- Recognizing while scanning

What is this?

Data-driven!
Acknowledgement

- Anonymous reviewers
- Hao (Richard) Zhang and Daniel Cohen-Or
- Research grants
  - NSFC
  - National 973 Program
  - Guangdong Science and Technology Program
  - Shenzhen VisuCA Key Lab
  - The Fundamental Research Funds of Shandong University
  - SIAT Innovation Program for Excellent Young Researchers
Thank you!

Code of object analysis:
www.kevinkaixu.net
Physically feasible push

- Heuristic rules