Coupled Joint Registration and Co-segmentation for Indoor Rigid Object Sets

Siyu Hu*

Abstract

Keywords: Co-segmantion, Joint Registration

Concepts: •**Computing methodologies** → **Image manipulation;** *Computational photography;*

1 Introduction

In many researches and applications of indoor scenes the data of segmented and even annotated 3D indoor scenes are required as either data base or training data (e.g.[Dema and Sari-Sarraf 2012][Fisher et al. 2012][Chen et al. 2014][Fisher et al. 2015]).
One way to build such data base is to interactively compose scenes from 3D shape models resulting in scenes with object segmentation and annotation naturally available, or to mannually segment and annotate existing scenes. This procedure can be tedious and time consuming, despite the efforts to improve the interaction experience(e.g.[Merrell et al. 2011][Xu et al. 2013]).
Another way is to automatically generate scenes from 3D shape models according to the input RGB or RGB-D images(e.g.[Liu et al. 2015][Chen et al. 2014]). In such methods, a retrievial procedure is usually needed and inevitably limit the result to a certain set of 3D models despite the actual 3D model in the input images.
We prefer a approach that helps us build such data set directly from the captured data. One of the major gap between the required data set (with object level segmentations) and available scene capturing framework(e.g.[Izadi et al. 2011]) is the general object level segmentation. We want to highlight that a general object segmentation should not be treated as an equivalence of multilabel classification. For 3D data, [Jia et al. 2015] used while .The work of [Xu et al. 2015] proposes a practical and a rather complete framework to close the gap. One of the key idea in [Xu et al. 2015] is that the motion consistency of rigid object can be strong evidence of general objectness. To exploit this, they employ a robot to do proactive push and use the movement tracking to verify and iteratively improve their object level segmentation result. Our work presented in this paper is trying to exploit the same idea with a different approach.

Need to find papers that uses data base and training data to back up my claims

*e-mail:sy891228@mail.ustc.edu.cn

SIGGRAPH 2016 Posters, July 24-28, 2016, Anaheim, CA
ISBN: 978-1-4503-ABCD-E/16/07
DOI: http://doi.acm.org/10.1145/9999997.9999999

2 Related Work

2.1 Functional Mapping

The Coupled Joint Registration and Co-segmentation problem addressed here is essentially a problem of point-to-point correspondence problem. A series of work based on the functional maps representation advocated in [Ovsjanikov et al. 2012] have be done. In one of the most recent work [Maron et al. 2016], a convex relaxation technique was used to better approximate the global minimal for both rigid and non-rigid registration problem.

3 Method Overview

3.1 Problem Statement

Given a set of point clouds which record the same group of rigid indoor objects with different layout. We intend to samutaneously partition the point clouds into objects and align the points of same object to recover layouts for corresponding object. Figure **??** shows an example of input point clouds set.

3.2 Formulation

To formulate the relation between the unknown object set and the input point clouds. We come up with a generation model as follows:

$$P(v_{mi}) = \sum_{n=1}^{N} p_n \sum_{k=1}^{K_n} p_k N(\phi_{mn}(v_{mi})|x_k, \Sigma_k) \tag{1}$$

which means, The observed point clouds are generated by N object model. Each object model is represented by a gaussian mixture model with K_n centroids. Our goal is to maximize the probability of the expected compelete-data log-likelihood. The object function can be written as:

$$\Theta = \text{argmax} \sum_{Z} P(Z|V, \Theta) \ln P(V, Z; \Theta) \tag{2}$$

in which:

$$\Theta = \{\{p_n\}_{n=1}^{N}, \{p_k, x_k, \Sigma_k\}_{k=1}^{\sum K_n}, \{\phi_{mn}\}_{m=1,n=1}^{MN}\}$$

is the parameters of the generation model.
p_n is the prior probability that the point is generated by the n-th object.
p_k is the weight of the k-th Gaussian.
x_k is the center of the k-th Gaussian.
Σ_k is the standard deviation of the k-th Gaussian.
There are $\sum K_n$ Gaussian model in total and among them, K_n Gaussian models belongs to object n.
V is the M input point clouds.
v_{mi} is the i-th point of the m-th point cloud.
Z is a latent variable set defined as:

$$Z = \{z_{ij}|j = 1...M, i = 1...N_j\}$$

among which if $z_{ij} = k(k = 1...\sum K_n)$ assign the observation of $\phi_{mn}(v_{mi})$ to the k-th component of Gaussian mixture model.Such formulation can be seen as an extention of joint registration formulation in [**?**], upon which we add several gaussian mixture model together to express a group of objects. By solving this new problem we simutaneously solve the object co-segmentation of given observation.

4 Algorithms and Implementation

4.1 Expectation Conditional Maximization

Assuming the observed point clouds $\{V_m\}$ are independent and identically distributed, we can then write the (2) as:

$$\varepsilon(\Theta|V,Z) = \sum_{m,i,k} \alpha_{mik}(\log p_n + \log p_k + \log P(\phi_{nm}(v_{mi})|z_{ji} = k;\Theta)) \tag{3}$$

In which the $\alpha_{mik} = P(z_{mi} = k|v_{mi};\Theta)$,

Algorithm 1 Joint Registration and Co-segmentation (JRCS)

Input:
$\{V_m\}$:Observed point clouds
$\{\alpha^0_{mik}\}$:Initial posterior probabilities
Output:
Θ^q:Final parameter set
1. $q \leftarrow 0$
2. **repeat**
3. CM-step-a: Use α^q_{mik}, x^{q-1}_k to estimate $\{R^q_{mn}\}$ and $\{t^q_{mn}\}$
4. CM-step-b: Use α^q_{mik}, $\{R^q_{mn}\}$ and $\{t^q_{mn}\}$ to estimate the Gaussian centers x^q_k
5. CM-step-c: Use α^q_{mik}, $\{R^q_{mn}\}$ and $\{t^q_{mn}\}$ to estimate the covariances Σ^q_k
6. CM-step-d: Use α^q_{mik} to estimate the priors p^q_k,p^q_n
7. E-step: Use Θ^{q-1} to estimate posterior probabilities. $\alpha^q_{mik} = P(z_{mi}|v_{mi};\Theta^{q-1})$
8. $q \leftarrow q + 1$
9. **until** Convergence
10. **return** Θ^q

4.2 Initialization Techniques

A key advantage motivates our formulation is that the soft correspondence can be initialized more flexiblely comparing to the typical initialization techniques such as landmark point pairs in registration.

Initial Segment based on Planar Fitting

Block Based Feature Extraction and Clustering

The result of Clustering:

$$P(B_{mj} \in C_n)$$

Soft Correspondence Initialization

Then the α is initialized as:

$$\alpha_{ijk} = P(B_{mj} \in C_n)$$

on the condition that:

$$v_{ij} \in B_{mj} \wedge x_k \in O_n$$

5 Experiments and Discussion

5.1 Debugging The Smooth Step

5.2 Current Result

5.3 Initialization Experiments

Acknowledgements

To Robert, for all the bagels.

References

BOUAZIZ, S., TAGLIASACCHI, A., AND PAULY, M. 2013. Sparse iterative closest point. In *Proceedings of the Eleventh Eurographics/ACMSIGGRAPH Symposium on Geometry Processing*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, SGP '13, 113–123.

BRONSTEIN, M. M., AND KOKKINOS, I. 2010. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 1704–1711.

CHEN, K., LAI, Y.-K., WU, Y.-X., MARTIN, R., AND HU, S.-M. 2014. Automatic semantic modeling of indoor scenes from low-quality rgb-d data using contextual information. *ACM Trans. Graph. 33*, 6 (Nov.), 208:1–208:12.

DEMA, M. A., AND SARI-SARRAF, H. 2012. 3d scene generation by learning from examples. In *Multimedia (ISM), 2012 IEEE International Symposium on*, 58–64.

EVANGELIDIS, G. D., KOUNADES-BASTIAN, D., HORAUD, R., AND PSARAKIS, E. Z. 2014. *A Generative Model for the Joint Registration of Multiple Point Sets*. Springer International Publishing, Cham, 109–122.

FISHER, M., RITCHIE, D., SAVVA, M., FUNKHOUSER, T., AND HANRAHAN, P. 2012. Example-based synthesis of 3d object arrangements. *ACM Trans. Graph. 31*, 6 (Nov.), 135:1–135:11.

FISHER, M., SAVVA, M., LI, Y., HANRAHAN, P., AND NIE, M. 2015. Activity-centric scene synthesis for functional 3d scene modeling. *ACM Trans. Graph. 34*, 6 (Oct.), 179:1–179:13.

IZADI, S., KIM, D., HILLIGES, O., MOLYNEAUX, D., NEWCOMBE, R., KOHLI, P., SHOTTON, J., HODGES, S., FREEMAN, D., DAVISON, A., AND FITZGIBBON, A. 2011. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, ACM, New York, NY, USA, UIST '11, 559–568.

JIA, Z., GALLAGHER, A. C., SAXENA, A., AND CHEN, T. 2015. 3d reasoning from blocks to stability. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37*, 5 (May), 905–918.

KRÄHENBÜHL, P., AND KOLTUN, V. 2011. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, Eds. Curran Associates, Inc., 109–117.

LEVY, B. 2006. Laplace-beltrami eigenfunctions towards an algorithm that "understands" geometry. In *IEEE International Conference on Shape Modeling and Applications 2006 (SMI'06)*, 13–13.

LIU, Z., ZHANG, Y., WU, W., LIU, K., AND SUN, Z. 2015. Model-driven indoor scenes modeling from a single image. In *Graphics Interface Conference*.

MARON, H., DYM, N., KEZURER, I., KOVALSKY, S., AND LIPMAN, Y. 2016. Point registration via efficient convex relaxation. *ACM Trans. Graph. 35*, 4 (July), 73:1–73:12.

MERRELL, P., SCHKUFZA, E., LI, Z., AGRAWALA, M., AND KOLTUN, V. 2011. Interactive furniture layout using interior design guidelines. *ACM Trans. Graph. 30*, 4 (July), 87:1–87:10.

OVSJANIKOV, M., BEN-CHEN, M., SOLOMON, J., BUTSCHER, A., AND GUIBAS, L. 2012. Functional maps: A flexible representation of maps between shapes. *ACM Trans. Graph. 31*, 4 (July), 30:1–30:11.

RODOL, E., BUL, S. R., AND CREMERS, D. 2014. Robust region detection via consensus segmentation of deformable shapes. *Computer Graphics Forum 33*, 5, 97–106.

WANG, F., HUANG, Q., AND GUIBAS, L. 2013. Image co-segmentation via consistent functional maps. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, 849–856.

XU, K., CHEN, K., FU, H., SUN, W.-L., AND HU, S.-M. 2013. Sketch2scene: Sketch-based co-retrieval and co-placement of 3d models. *ACM Trans. Graph. 32*, 4 (July), 123:1–123:15.

XU, K., HUANG, H., SHI, Y., LI, H., LONG, P., CAICHEN, J., SUN, W., AND CHEN, B. 2015. Autoscanning for coupled scene reconstruction and proactive object analysis. *ACM Trans. Graph. 34*, 6 (Oct.), 177:1–177:14.