PROPOSAL: Interior/Residential Design with RGBD Data

Abstract

RGBD cameras is becoming more and more popular for common users to capture the environment where they live. In this paper, we present a *form and function* exploration system to reconstruct the 3D geometry and mine furniture functions in interior environments from dynamic RGBD data. We capture a sequence of RGBD images in a long period of time, in which objects may change their positions and poses frequently. To reconstruct the geometry of the cluttered indoor scene, we first cluster the objects by their motions: static objects or dynamic objects. Object motion is computed from correspondence detection from both appearance and geometry. During the motion clustering, the interrelation between different objects can be discovered from the spatial relationship at different times.

Keywords: Interior design, appearance, editing, material

1 Introduction

3D indoor scenes are popular in many applications, such as games, robotics, virtual reality, etc. Modeling indoor scenes has been attracted large amount of attentions for decades in computer graphics. Recently, many techniques have been presented to generate static indoor environments, including dense modeling from RGBD data [Henry et al. 2012; Izadi et al. 2011; Xiao and Furukawa 2012; Yan et al. 2014], combing object classification and modeling [Shao et al. 2012; Nan et al. 2012; Kim et al. 2012], and synthesizing of 3D indoor scenes from large collection of examples [Fisher et al. 2012a; Xu et al. 2013].

Comparing with static scenes, dynamic scene analysis has significant value for artists in interior design, animation making, etc. The manners of how furniture objects interact with each other and how furniture objects interact with users play a very important rule in interior design. Typically, the function (behavior) of an object is interrelated with its form (geometry). In a dynamic scene, the function of an object is reflected its different forms. However, the dynamic indoor scene analysis has not been investigated much in computer graphics.

In this paper, we present a framework to explore the forms and functions of dynamic indoor scenes. Because the raw RGBD data is noisy and incompleted due to occlusions in a cluttered environment, it is very challenging to reconstruct a perfect 3d model for the scene. Ambiguous boundaries between objects make it nontrivial to generate clear motion representation. We do not separate the geometry and behavior of objects in our system. The scene is firstly roughly segmented and modeled. Then we build a structure graph to describe the object behavior in the scene. To make sure the semantics consistency between objects during their movements, we simultaneously optimize the geometry and the structure graph.

The contributions of our system is two-fold.

1. To the best of our knowledge, our system performs, for the first time, behavior analysis in a dynamic indoor scene from raw RGBD data.

2. The learned behaviors provide a set of fundamental rules for a wide range of applications. We apply our behavior model to room layout suggestions, incident detection, indoor scene synthesis to demonstrate the power of our model.

2 Related Work

Many techniques have been proposed to generate static 3D indoor scenes in computer graphics. Though none of them focus on dynamic scene analysis like our system, they provide valuable reference on the underlying techniques.

Reconstruction from RGBD Data For static scenes, KinectFusion [Izadi et al. 2011] enables the real-time reconstruction by holding and moving a depth camera. For large-scale indoor scenes with multiple rooms, reconstructing a dense 3D model from the noisy and incomplete scanned range data typically involves registration of point clouds in different views and a global optimization to reduce gaps in a large scene [Xiao and Furukawa 2012; Henry et al. 2012]. Their goal is mainly to generate high-quality point clouds but without semantic analysis of the objects appear in the scene. Recently, object classification is employed to assist modeling for massive indoor scenes that containing many instances of chairs, desks, etc. [Koppula et al. 2011] first introduce the learning algorithm to understand the RGBD data of an indoor scene. To further reconstruct the 3D model for a cluttered indoor scene, 3d model databases can be used as template by searching for similar 3D model and then fitting the template to the scanned data [Shao et al. 2012; Nan et al. 2012]. [Kim et al. 2012] do not manually collect 3d models to build the database. The template model is reconstructed by scanning the same object in different configuration. Each model has an additional presentation by geometric primitives. [Shao et al. 2012] trains the class model based on geometry and appearance features to segment and label the RGBD data captured under sparse views. By learned an initial model for each class of object in indoor environments from a pre-labelled database, the model are refined progressively with user-refined segmentation results. The 3D model can be generated by placing the most similar model in the database according to the RGBD data. If objects move in a scene, they can be detected and reposed by segmented and classified based on the learned model from previously reconstructed model [Liu et al. 2014]. Different with these techniques, we pay more attentions on analyzing the object behaviors from the dynamic range data.

Reconstruction from Dynamic Point clouds Many techniques have been proposed to reconstruct the object surfaces from the range data sequences. [Wand et al. 2007] uses a *statistical framework* to reconstruct the geometry from real-time range scanning. Each frame is divided into 3d pieces. A statistical model is used to iteratively merge adjacent frames by aligning pieces and optimizing their shapes. However, some geometric artifacts remain due to structured outliers and in some boundary regions. [Chang and Zwicker 2011] presents a global registration algorithm to reconstruct *articulated 3D models* from dynamic range scan sequences. The surface motion is modeled by a reduced deformable model. Joints and skinning weights are solved in the system to register point clouds in different poses. (xuejin: We may also consider the furniture objects in indoor environments as articulated models, whose shapes under different poses can be deformed through connectors like hinge, slide, and so on.)

[Bouaziz et al. 2013] propose a new formulation of the ICP algo-

rithm using sparse inducing norms. While it achieves superior registration result on the data with outliers and missing region, only rigid alignment is handled. [Yan et al. 2014] employ a proactive capturing by asking the user to move the objects to capture both interior and exterior of a scene. The correspondence between adjacent frames is built first then segmentation. (xuejin: However, the motion information worths more than just helping registration. It can be used for analysis of object movements and functions.)

(xuejin: 1. In comparison, the range data used in our system is captured with a large time spacing and the motion of the objects in the scene varies in a wide range. 2. Can we use image/texture data for more reliable correspondence?)

Data-Driven Furniture Layout The general way producing the layout of furniture objects is to model a set of design rules and then to optimize an energy function given constraints by individuals. [Merrell et al. 2011] formulates a group of layout guidelines in a density function according to professional manuals on furniture layout. When the user specifies the room shape and an initial arrangement of the set of furniture to be placed in the room, this system generates a number of layout suggestions by a hardware-accelerated Monte Carlo sampler. Instead of manually define the layout guidelines, the hierarchical and spatial relationships of the furniture objects can be learned from a set of examples [Yu et al. 2011]. Assembling these relationships and other ergonomic factors into a cost function, multiple arrangements can be yielded quickly by simulated annealing using a Metropolis-Hastings state search step. In these methods, manual labours are required in modeling the design rules and providing an initial layout. Fisher et al. [2012b] trains a probabilistic model for indoor scenes from a small number of examples. A variety of indoor scenes can be automatically synthesized from a few of user specified examples. Indoor scenes bring more difficulties for scene analysis because there are always many cluttered objects in different scales, shapes, and functions. A focal-driven analysis and organization framework is presented for heterogeneous collections of indoor scenes [Xu et al. 2014]. They develop a co-analysis algorithm which interleaves frequent patten mining and subspace clustering. The interrelations between objects play important role during furniture arrangement in these systems. However, the 3D scene models takes many efforts to collect for training. In comparison, our system provides an efficient framework to generate 3d model examples for many further applications.

Co-analysis of shape, functions in a large database of 3D object models With the growth of 3D shape databases on the Internet, many techniques have been proposed for co-analysis in a large shape collection of the same object category. A series of geometry processing tasks such as model segmentation, shape retrieval, and shape synthesis. Point-to-point networks are used to represent the shape correspondence between shapes [Rustamov et al. 2013]. To better explore the shape space, [Huang et al. 2014] propose a framework for computing consistent functional maps within heterogeneous shape collections. Cycle-consistency of the functional map network largely reduce the noise correspondences. Based on the continuous nature of functional maps, the proposed framework outperforms point-based representation in shape interpolation, shape retrieval and classifications for both man-made and organic shapes. Large 3d model collection can also help recovering the depth map for a single image [Su et al. 2014]. With a non-rigid registration formulation, the image is popped-up to minimize the distance between corresponding points in the image and similar 3d shapes in the database. However, all these techniques focus on the geometry characteristics within one object category. In a cluttered environment, the inter-connection between different types of objects has not been investigated yet.

3 Overview

Our system consists of four main steps, as Figure 1 shows. The four steps are iteratively performed until it converges.

1. *Registration/Labeling:* Given a sequence of RGBD data captured in different times and views, we first register the camera views to detect motions. The point clouds are divided into two categories: static objects and dynamic objects. Object detection is performed by combing the motion depth and appearance features to label and segment the different objects in the scene.

2. *Spatial Distribution:* After each round of registration and labeling of the input RGBD images, we can refine the spatial distribution map of each object. The map describes its position, poses and so on in the indoor environment. The interconnection between objects are then progressively refined.

3. *Object Modeling/Form Reconstruction:* Static objects are reconstructed based on geometry primitives/model database. Dynamic objects are typically small size, connected with a large static object. They can be reconstructed by combing the point clouds captured under different views.

4. *Function Analysis/Behavior Map:* We build a dynamic behavior map including all of the objects in the scene. Each node is an object/a part of an object. The edges in the graph describe the spatial relationship between objects/parts. The object behavior can be explored from its surrounding objects in the dynamic structure graph. With a dynamic behavior map, the RGBD data is re-segmented and re-labeled to obtain semantical consistency of the scene.

4 Registration and Labeling of RGBD images

Since the method used in [Shao et al. 2012] requires a certain amount of user interactions, we first register RGBD images taken in different views and different times using (xuejin: method for automatic registration.)

SIFT Matching SIFT feature is a method for feature detection and description robust to lighting, viewpoints, scale, rotation changes. For cluttered indoor scene, SIFT actually shows its ability to find correspondences between objects, especially for small objects with rich textures which will lead to discriminative features. As shown in Figure 2, the scene is quite cluttered with many small-scale objects on the desk. (xuejin: Very dense correspondences are obtained, most of which are on the small objects with free-form shapes or rich textures.) However, if the scene is clean, as Figure 3 shows, SIFT leads to sparse features and correspondences with a large portion of outliers. (xuejin: More dense viewpoints with enough overlaps between frames, will improve the accuracy and density of the feature matching, such as matches between 9-10.)

Registration of RGBD Images TBD.

References

BOUAZIZ, S., TAGLIASACCHI, A., AND PAULY, M. 2013. Sparse iterative closest point. *Computer Graphics Forum (Symposium on Geometry Processing) 32*, 5, 1–11.

CHANG, W., AND ZWICKER, M. 2011. Global registration of dynamic range scans for articulated model reconstruction. *ACM Transactions on Graphics 30*, 3.

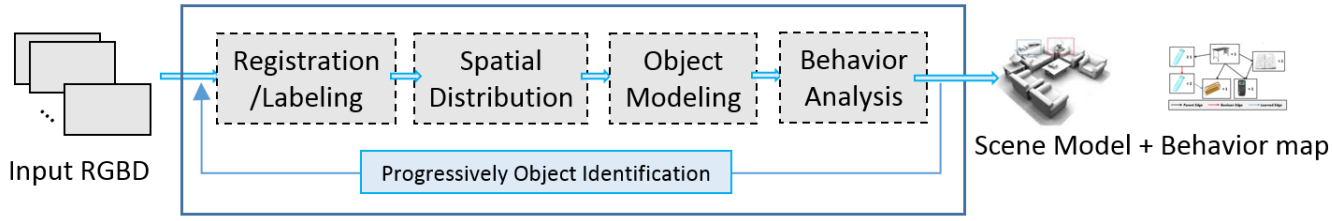

Figure 1: *System overview.*

Figure 2: *Feature detection and matching results using SIFT for a cluttered indoor scene.*

Figure 3: *Feature detection and matching results using SIFT for a* clean *indoor scene.*

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., 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.

HENRY, P., KRAININ, M., HERBST, E., REN, X., AND FOX, D. 2012. RGB-D mapping: Using kinect-style depth cameras for dense 3D modeling of indoor environments. *International Journal of Robotics Research (IJRR) 31*, 5 (April), 647–663.

HUANG, Q., WANG, F., AND GUIBAS, L. 2014. Functional map networks for analyzing and exploring large shape collections. *ACM Trans. Graph. 33*, 4 (July), 36:1–36:11.

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 *ACM Symposium on User Interface Software and Technology*.

KIM, Y. M., MITRA, N. J., YAN, D.-M., AND GUIBAS, L. 2012. Acquiring 3d indoor environments with variability and repetition. *ACM Transactions on Graphics 31*, 6, 138:1–138:11.

KOPPULA, H., ANAND, A., JOACHIMS, T., AND SAXENA, A. 2011. Semantic labeling of 3D point clouds for indoor scenes. In *Conference on Neural Information Processing Systems (NIPS)*.

LIU, Z., TANG, S., XU, W., BU, S., HAN, J., AND ZHOU, K. 2014. Automatic 3D indoor scene updating with rgbd cameras. *Computer Graphics Forum (Pacific Graphics) 33*, 7.

MERRELL, P., SCHKUFZA, E., LI, Z., AGRAWALA, M., AND KOLTUN, V. 2011. Interactive furniture layout using interior design guidelines. *ACM Trans. Graph. (Siggraph'11)*.

NAN, L., XIE, K., AND SHARF, A. 2012. A search-classify approach for cluttered indoor scene understanding. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2012) 31*, 6.

RUSTAMOV, R. M., OVSJANIKOV, M., AZENCOT, O., BEN-CHEN, M., CHAZAL, F., AND GUIBAS, L. 2013. Map-based exploration of intrinsic shape differences and variability. *ACM Trans. Graph. 32*, 4 (July), 72:1–72:12.

SHAO, T., XU, W., ZHOU, K., WANG, J., LI, D., AND GUO, B. 2012. An interactive approach to semantic modeling of indoor scenes with an rgbd camera. *ACM Trans. Graph.*, 136–136.

SU, H., HUANG, Q., MITRA, N. J., LI, Y., AND GUIBAS, L. 2014. Estimating image depth using shape collections. *ACM Trans. Graph. 33*, 4 (July), 37:1–37:11.

WAND, M., JENKE, P., HUANG, Q., BOKELOH, M., GUIBAS, L., AND SCHILLING, A. 2007. Reconstruction of deforming geometry from time-varying point clouds. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, SGP '07, 49–58.

XIAO, J., AND FURUKAWA, Y. 2012. Reconstructing the world's museums. In *Proceedings of the 12th European Conference on Computer Vision*, ECCV '12.

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 Transactions on Graphics 32*, 4, 123:1–123:12.

XU, K., MA, R., ZHANG, H., ZHU, C., SHAMIR, A., COHEN-OR, D., AND HUANG, H. 2014. Organizing heterogeneous scene collection through contextual focal points. *ACM Transactions on Graphics, (Proc. of SIGGRAPH 2014) 33*, 4, to appear.

YAN, F., SHARF, A., LIN, W., HUANG, H., AND CHEN, B. 2014. Proactive 3d scanning of inaccessible parts. *ACM Transactions on Graphics(Proc. of SIGGRAPH 2014) 33*, 4.

YU, L.-F., YEUNG, S. K., TANG, C.-K., TERZOPOULOS, D., CHAN, T. F., AND OSHER, S. 2011. Make it home: automatic optimization of furniture arrangement. *ACM Trans. Graph. 30*, 4, 86.