Computers & Graphics 53 (2015) 210–223

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section

Single image-based data-driven indoor scene modeling

Yan Zhang [a,b,1], Zicheng Liu [a,b], Zheng Miao [a,b], Wentao Wu [a,b], Kai Liu [a,b], Zhengxing Sun [a,b]

[a] State Key Lab for Novel Software Technology, Nanjing University, China
[b] Department of Computer Science and Technology, Nanjing University, China

ARTICLE INFO

Article history:
Received 11 August 2015
Received in revised form
22 September 2015
Accepted 7 October 2015
Available online 20 October 2015

Keywords:
Data-driven
Indoor scenes modeling
Single image
Image analysis
Model retrieval
Relevance feedback

ABSTRACT

With a single input indoor image (including sofa, tea table, etc.), a 3D scene can be reconstructed from a single image using an existing model library in three stages: image analysis, model retrieval and relevance feedback. In the image analysis stage, we obtain the object information from the input image using geometric reasoning technology combined with an image segmentation method. In the model retrieval stage, line drawings are extracted from 2D objects and 3D models by using different line rendering methods. We exploit various tokens to represent local features and then organize them together as a star-graph to show a global description. By comparing similarity among the encoded line drawings, models are retrieved from the model library. Also, for a better user experience, we add a relevance feedback stage following the retrieval stage. The Support Vector Machine method is used to conduct the feedback operation. After this stage, the retrieved models are in conformance with the image semantic. The 3D scene is then reconstructed. Experimental results show that, driven by the given model library, indoor scenes modeling from a single image could be achieved automatically and efficiently.

1. Introduction

With the development of 3D acquisition and modeling technology, sharable 3D models are rapidly increasing. Repositories like Trimble/Google 3D warehouse [1] and TurboSquid [2] contain millions of digital models. On the basis of previous work, people deform these models to make them conform with new design requirements using the existing digital geometric content reasonably and repeatedly. It not only greatly reduces the rigmarole work, but also largely shortens the period of design and modeling. All of these factors have important value for the development of modeling technologies [3–5].

Since 3D model reconstruction from a single image has always been an ill-posed problem, it is difficult to directly construct the scene from images. However, many existing libraries contain a large number of useful surface models, from which similar models can be retrieved to reconstruct the virtual scene. From this point of view, we present a new approach for model-driven indoor scene modeling from single image. Since color images contain many geometric cues, in the first image analysis stage, geometric reasoning is combined with the image segmentation method to obtain the main object information from indoor scene images. The extracted object information can be used in the following retrieval stage. In the model retrieval stage, the idea from sketch-based retrieval [6,7] is borrowed. In our method, since 2D objects and 3D models can all be represented by line drawings, similar models can easily be obtained from the model library by comparing their similarity. In line drawing representation, we propose a novel feature encoding method, which hierarchically combines global and local features together to represent a line drawing. In this way, similar 3D models can be retrieved reasonably. In addition, retrieved models might not be semantically reasonable. We, therefore, amend the retrieval result using relevance feedback for better results. This stage brings user into the modeling procedure, and makes the reconstructed scene be more similar to the original image semantic.

Contribution: Our main contribution includes the following:

(1) With the support of existing model libraries, an approach for 3D indoor scenes reconstruction from a single image is presented. This method reconstructs objects from the scene automatically and efficiently.
(2) We propose a novel feature encoding method for line drawings. In this encoding method, not only local features but also structured global features are represented. It converts the clutter of pixel data into an ordered expression in line with the human visual perception, and the retrieval precision is improved accordingly.

2. Related works

Image-based 3D reconstruction: Image-based modeling is an ideal way for constructing a 3D world in people's mind. Traditional

E-mail address: zhangyannju@nju.edu.cn (Y. Zhang).
[1] Tel./fax: +86 25 89686577.

http://dx.doi.org/10.1016/j.cag.2015.10.004
0097-8493/

image-based modeling technologies require one to take photos from different viewpoints for one scene and then use technologies such as stereo vision to construct 3D scenes [8,9]. However, the inconvenience of taking several photos for a scene leads to limitations of these methods. Therefore, methods for single-image-based modeling technology would be more widely accepted, since a single scene image is usually available.

In the computer vision field, traditional single-image-based modeling methods construct a 3D scene using cues like lightness [10], texture [11], focal length [12], etc. However, these methods always have strict restrictions for object properties, such as shapes and light reflection in a scene. They are only available for certain scenes. In subsequent studies, some researchers attempt to add interactions to simplify the reconstruction problem [13,14,2]. Many of these methods construct the scene through manually setting vanishing points and geometric invariant for an image. However, they also have limitations in that only certain geometries and basic planes can be constructed. In our method, based on the existing model library, we could reconstruct a more reasonable scene from a single image.

Model-driven 3D shape modeling: In recent years, data-driven modeling based on the model library has become an emerging modeling technology [3,4,15,16], since the number of 3D models on the Internet is rapidly increasing. Many studies have developed modeling technologies for a single object. In these studies, there are three categories according to various input items.

In the first category, people transform models guided by images. Chaudhuri et al. [3] introduce data-driven suggestions for 3D modeling which allows the artist to benefit from customized examples that stimulate creativity. They describe shape and retrieval and shape correspondence techniques for creative prototyping of 3D models. Further research can be found in paper [4], where they introduce assembly-based modeling which is a promising approach to broadening the accessibility of 3D modeling. In assembly-based modeling, new models are assembled from shape components extracted from a database. A probabilistic reasoning approach is introduced to the problem of identification of relevant components to be presented to the user. Given a repository of shapes, they learn a probabilistic graphical model that encodes semantic and geometric relationships among shape components. A similar approach based on the photo is introduced by Xu et al [17]. They introduce an algorithm for 3D object modeling where the user draws creative inspiration from an object captured in a single photograph. They use the photo as the guidance to retrieve the model parts to deform models and finally to reconstruct a model.

Another category of methods starts from sketches. Some researchers have implemented sketch-based model retrieval technologies [7,6]. With the sketches as input, they retrieve similar models in the model library. However, these kinds of methods pay more attention to the increase of retrieval precision, rather than the modeling itself. Some people use sketches as the guidance to construct 3D models based on the model library. They provide a new direction for single object modeling. Fan et al. [18] present a novel sketch-based modeling system which allows novice users to create 3D custom models by assembling parts based on a database of pre-segmented 3D models. Different from previous systems, their system supports the user with visualized and meaningful shadow guidance under his strokes dynamically to guide the user to convey his design concept easily and quickly. Xie et al. [19] rethink model design as navigating through different possible combinations of part assemblies based on a large collection of pre-segmented 3D models and propose an interactive sketch-to-design system, where the user sketches prominent features of parts to combine.

In the third category, people use collected point cloud data or estimate image depth to help complete the reconstruction. With cloud data, Shen et al. [20] introduce a technique that allows quick conversion of acquired low-quality data from consumer-level scanning devices to high-quality 3D models with labeled semantic parts and their assembly reasonably close to the underlying geometry. However, only a single object could be reconstructed in this kind of methods.

In addition, apart from sketch-based retrieval technology, other methods have certain limitations on the model library. When objects in different categories have to be modeled, the library has to be built by categories. In the scene reconstruction, since there are several objects in one scene, it is difficult to exploit the above methods. Since the different objects in a scene are not marked, we do not know how to map them to the model categories.

Model-driven 3D scenes modeling: With the development of technology, some people proposed scene modeling approaches based on the model library which are classified into two categories according to different input items.

One category of methods is for sketches. Based on the existing model library, sketch retrieval technology is combined with sketch modeling technology to construct a scene in some approaches. Cases such as various objects in one scene can be handled. However, there are still certain requirement on the order and category for input sketches. Lee et al. [21] introduce a modeling system for parts composition with a sketching interface which needs considerable user interaction. In the paper [22], Shin et al. propose a system that takes simple 2D sketches of models in a scene as input for 3D scene construction. The system then automatically identifies corresponding models in a database and puts them in the appropriate location and posture so that their appearance matches the user's input sketches. However, their system needs extensive interaction when drawing sketches and retrieval algorithms also need to be improved. Further research has [23] leveraged more precise retrieval methods, but the limitations are similar.

In the other category of methods, with the help of model library, people use the collected point cloud data to construct a scene. Shao et al. [24] present an interactive approach to the semantic modeling of indoor scenes with a consumer-level RGBD camera. Meanwhile, Nan et al. [25] present an algorithm for recognition and reconstruction of scanned 3D indoor scenes. Chen et al. [26] also present a novel solution to automatic semantic modeling of indoor scenes from a sparse set of low-quality RGB-D images and Contextual relationships learned from the 3D database are used to constrain reconstruction, ensuring semantic compatibility between both object models and parts. In these methods, researchers usually segment the scene image with point cloud data. Each object in the image can then be located and a similar object is easily retrieved in the model library. When every object is fixed, the scene reconstruction is completed accordingly. Although these studies have achieved some success, there are special requirements for the input. In addition, this technology is inconvenient for most people, especially non-professional who users would find it difficult to draw the scene sketch and would be unable to collect point cloud data without professional devices. We note that single scene images are relatively easy obtained from the real world, therefore, if the scene reconstruction can be completed from a single image with the help of an existing model library, the method would be more valuable and widely available for common users.

We start with this motivation to carry out research about reconstruction from a single image. Based on this idea, a version of this paper has been published in Graphics Interface 2015 [27]. In that paper, we first analyze the image and separate furniture out from the image background. Then model retrieval is then conducted to get models from the library. The retrieved models are put in the room according to the relative position and finally reconstruction is completed. Compared with the conference