Rapid Interactive 3D Reconstruction from a Single Still Image

Ashutosh Saxena, Nuwan Senaratna, Savil Srivastava, and Andrew Y. Ng
Computer Science Department, Stanford University.
e-mail: {asaxena,nuwans,savil,ang}@cs.stanford.edu

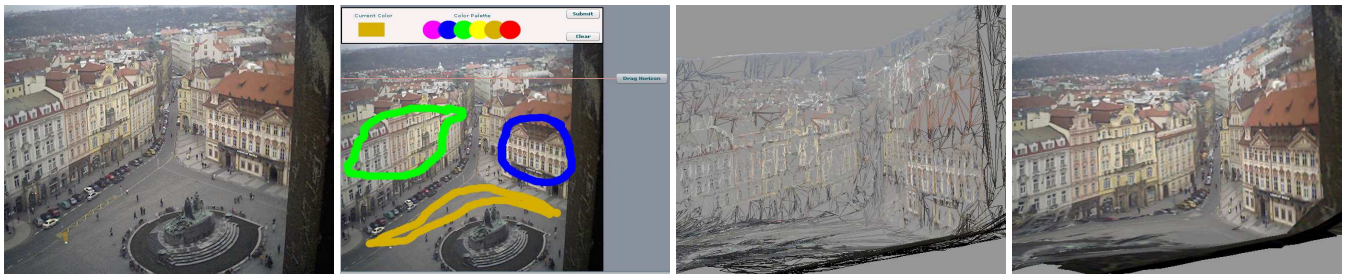

Figure 1: *(a) An original image. (b) Interactive tool: user draws over the image in 5-10 seconds. (c) The 3-d model predicted by the algorithm. (d) A screenshot of the textured 3-d model. (Also see attached video.)*

Abstract

We present a method to create photorealistic 3D models from a single still image. The user provides input in the form of "scribbles"—within a few seconds of user input, our algorithm generates a photorealistic 3D model and a 3D flythrough from the image. Our approach uses a machine learning algorithm that learns the relation between (single) image features and depth. In our preliminary experiments, 1238 users created good 3D models for 84% of the 3775 images they uploaded.

Keywords: Image-based Modeling, Single Image 3D reconstruction, Machine Learning

1 Summary

We consider the problem of creating a 3D model from a single still image. This problem, while interesting to both graphics and vision communities, still remains an extremely challenging one. In a narrow mathematical sense, it is indeed impossible to recover 3D depth from a single still image, since we can never know if it is a picture of a painting (in which case the depth is flat) or if it is a picture of an actual 3D environment.

Recently, [Saxena et al. 2005; Saxena et al. 2007a; Saxena et al. 2007b] proposed machine learning algorithms that learn the relation between the image features and depth, and therefore were able to create 3D models from a single image. Although these methods are completely automatic, they are successful only on a fraction of images (48.1% in our experiments) and often there are flaws in parts of the model. On the other extreme, one can also create 3D models completely manually using 3D editing software tools such as Google Sketchup, Autodesk, etc.; however, these demand significant skills and a large amount of time to create a model.

In this work, our goal was to create an algorithm that allows even inexperienced users to reliably create 3D models from an image in just a few seconds. The idea is to use our learning algorithm to infer the 3D models, but rely on user input where the learning algorithm fails. For such failure cases, we designed our algorithm to model the 3D structure of the scene by also using the additional user input, such as an approximate location of the horizon and some major planes in the scene. This helped our algorithm in creating much better 3D models.

We designed our user interface to be simple and intuitive to use. It comprises of: (a) a "scribbling" tool, where users can draw over the 2D image in different colors for identifying some major planes in the image, and (b) a line indicating the horizon, which can be dragged into a position by the user (Fig. 1). In our studies, we found that requiring more involved input from the user (e.g., the labels for the objects in the image) made the interface cumbersome to use, but gave only a marginal improvement in the performance.

The overall system runs in an internet browser as follows: (a) user uploads his photograph, (b) learning algorithm creates a 3D model, (c) if the 3D model is not good enough, then the user spends a few seconds (typically 5-10 sec) in the interactive tool to help the algorithm. In preliminary experiments, 1238 users (many of them were inexperienced, i.e., they had not used a 3d editing software before) created 3D models for 3775 images using the interactive tool, and were able to create good 3D models in 84% of the cases. More details and the online demo are available at:

http://make3d.stanford.edu

2 Related Work

[Hengel et al. 2007] provide an interactive method to create a 3D model from a video. Delage et al. [Delage et al. 2005] and Hoiem et al. [Hoiem et al. 2007] assumed that the scene is made of a horizontal ground and vertical walls (and possibly sky) to create "popup" type models. However, these methods work only on a small fraction of images that satisfy this assumption. [Bai and Sapiro 2007] used user input to segment the images.

References

BAI, X., AND SAPIRO, G. 2007. Distancecut: Interactive segmentation and matting of images and videos. In *ICIP*.

DELAGE, E., LEE, H., AND NG, A. 2005. Automatic single-image 3d reconstructions of indoor manhattan world scenes. In *ISRR*.

HENGEL, A., DICK, A., THORMAHLEN, T., WARD, B., AND TORR, P. H. 2007. Videotrace: Rapid interactive scene modelling from video. In *ACM SIGGRAPH*.

HOIEM, D., EFROS, A., AND HEBERT, M. 2007. Recovering surface layout from an image. *IJCV 75*.

SAXENA, A., CHUNG, S. H., AND NG, A. Y. 2005. Learning depth from single monocular images. In *NIPS 18*.

SAXENA, A., CHUNG, S. H., AND NG, A. Y. 2007. 3-D depth reconstruction from a single still image. *IJCV*.

SAXENA, A., SUN, M., AND NG, A. Y. 2007. Learning 3-d scene structure from a single still image. In *ICCV 3D Representation for Recognition (3dRR-07)*.