**Review for Learning 3-D Scene Structure from a Single Still Image**

**Mingyang Zhou**

**zmykevin**

**Problem Statement and Motivations**

The author of the paper tries to develop a method that can accurately infer the detailed 3-d structure, which includes the information about locations, orientations, and depth of different planes and objects in the image, without assuming constraints on the structure of the image. The algorithm will be able to create a 3-D model that can represent the information accurately, and have a pleasing outlook as well.

**Evaluation of the Algorithm**

The model is trained with images coming from custom-built-3D scanner applied to the surrounding environment of Stanford University to construct the MRF. It is then tested on 134 images from the 3D scanner and 588 images from google search following a list of key words. The images collected from google is done by a person not associated to the project in order to assure the non-constraint to the structure of the image data to be tested. The results are evaluated from 4 different metrics, which are log-10 scale average depth error, average relative depth error, correct percent of models qualitatively and the correct percent of major planes. With respect to the four metrics, the designed plane parameter MRF from the paper beats other models that are applied to extract the 3D information from the image. Besides using the four qualitative evaluation metrics, evaluation is also given by a human who is not related to the project.

**Main Idea of the Algorithm**

The main idea is to represent the pictures with ‘superpixels’, which is the homogeneous collection of a small number of pixels in the image that shares similar properties. Then it tries to train a MRF model with the superpixels that represent the images as the node to learn several important properties that can be applied to infer 3-d information, including image features and depth, connected structure, coplanar structure and co-linearity. The features that are computed for different superpixles include monocular image feature and features for the boundaries between different supperpixels, to better estimate the 3-D information of each superpixel. In order to accurately learn the 3-D information, object information is also incorporated to add some common constraints between different objects in the image.

**Creation in the Algorithm**

Compared to the previous related work, the creation of this algorithm lies on several aspect. First it uses one still image to infer 3D information, instead of multiple images as required by stereovision, and structure of motion. Compared to the method that reconstruct 3D information from a single image, the models that it setup learns much more detailed 3-D information from the images, and also incorporate object information, so that the reconstruction of 3D models are more accurate and visual pleasing.

**Weakness**

The evaluation with a person is not quite convincing, at least it should be evaluate by a good number of people to get a more reliable feed back. Texture rendering techniques can be improved to provide better texture. The number of test images is not as massive as to convince that the learned model can reconstruct the 3-D information without assumption on the structure. In terms of explanation of the method, many mathematics symbles and equations are quite confusing. Thus, it should include more explanations on those mathematics equations.

**Rate**

As far as I am concerned, this paper is a 7 point paper. Extracing 3D information from one single image is a challenging problem, but also a very practical techniques, which can be applied to autonomous cars and other vision techniques. The improvement in accuracy and the look of the model is exciting. While there is room to improve the evaluation of the model, and performance, the author can also explain their method more clearly with a more straight-forward explanations to those mathematics equations and representations.