Methods in 3D Depth Reconstruction from a Single Image

Wei Wen

Department of Electrical Engineering, BTH

Supervisor: Siamak Khatibi
Outline

• Introduction

• Methods and 3D Results

• Statistic Results

• Conclusion and Future work
Introduction

• Increasing need for geometric 3D models
  – Movie industry, games, virtual environments…

• Existing solutions are not fully satisfying
  – User-driven modeling: long and error-prone
  – 3D scanners: costly and cumbersome

• Alternative: analyzing image sequences
  – Cameras are cheap and lightweight
  – Cameras are precise (several megapixels)
Scenario

• A scene to reconstruct (unknown a priori)

• Several viewpoints
  – from 1 views up to several hundreds
  – 20~50 on average
Sample Image Sequence

How to retrieve the 3D shape?
First Step: Camera Calibration

• Associate a pixel to a ray in space
  – camera position, orientation, focal length…

• Complex problem
  – solutions exist
  – toolboxes on the web
  – commercial software available
General Strategy: Triangulation

Matching a feature in at least 2 views

↓

3D position
This process to recover the 3D information of an object from 2D images is called as 3D reconstruction.

An example: One of the 3D results.
• It is difficult to get 3D information from a single image

Usually, 3D reconstruction is from two or multiple images.

The reconstruction principle of a 3D structure is based on triangulation which is an approach to data analysis that synthesizes data from multiple sources.
Epipolar Geometry
For example the parallel lines of the track are intersecting at one point which is impossible in the reality.
Some clues can help to reconstruct 3D
• vanishing points and so on
• vanishing lines
• shape from shading and texture
Methods and 3D Results

- The four methods chosen for evaluation
- Each method is based on different clues and models.
1. 3D Depth Reconstruction from a Single Still Image

- Call it Make3D for short
- Both the 3-d location and the orientation of the small planar regions in the image using a MRF were inferred
Procedure of the method

• Creating Superpixels
• Creating Features and multiple segmentations (base on the superpixels)
• Calculating superpixel-shape features
• Inference (build 3D)
Original image

Superpixels map
one of 3d results from Make3D
2. Automatic photo pop-up

- Call it Popup for short
- What is the meaning of Popup?
- Build a coarse, scaled 3D model by classifying each pixel as ground, vertical or sky and estimating the horizon position.
- Color, texture, image location, and geometric features are all useful cues for determining these labels
Procedure

• 1. Image $\rightarrow$ superpixels
• 2. Superpixels $\rightarrow$ multiple constellations
• 3. Multiple constellations $\rightarrow$ superpixel labels
• 4. Superpixel labels $\rightarrow$ 3D model
one of 3d results from Popup

- Call it Auto3D for short
- This method is limited for urban scenes.
- The goal is to improve the efficiency of 3D reconstruction
3D model structure

3-d model is composed of a number of vertical walls and a ground plane
one of 3d results from Auto3D
4. Stereo image displaying based on both physiological and psychological stereoscopy from single image

• Call it Psych for short
• Introduce the binocular disparity to create the second image
• Make every pixel on the image shift on the horizontal direction, no shift on the vertical direction.
The image pair

Original image

New image
Materials for the evaluation

- Database includes 80 images from natural urban and indoor scene
- Some Codes
  Download from Internet
- Statistic test
  20 volunteers are recruited to mark for the 3D results
Rules for statistic test
If

1. The right angle and relative position between each two objects achieves
2. The right ground plane and the right horizon are found.
3. Clarity of depth perception for different objects in a scene is achieved
4. The shape of the object is fine, and is not distorted
5. It is possible to find as many details as there are in the scene
the statistic result
### Processing Time of Each Algorithm

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Processing Time (unit: second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make3d</td>
<td>60-120</td>
</tr>
<tr>
<td>Auto3d</td>
<td>1-4</td>
</tr>
<tr>
<td>Popup</td>
<td>20-40</td>
</tr>
<tr>
<td>Psychology</td>
<td>1-3</td>
</tr>
</tbody>
</table>
### Theory comparison

<table>
<thead>
<tr>
<th></th>
<th>Modeling by</th>
<th>Image type</th>
<th>3D equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make3d</td>
<td>MRF</td>
<td>All</td>
<td>No</td>
</tr>
<tr>
<td>Popup</td>
<td>Pixel classification</td>
<td>Outdoor</td>
<td>No</td>
</tr>
<tr>
<td>Auto3D</td>
<td>CRF ?</td>
<td>Urban</td>
<td>No</td>
</tr>
<tr>
<td>Psychology</td>
<td>From brain perception modeling</td>
<td>All</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Conclusion and Future work

• Base on the statistic result
  Psych is the best one
  Make3d is better for natural images,
  Auto3d is better for urban scenes.

• Future work
  combine the advantages of each method to make a better one
Thank you