Depth-Aware Shape from Shading

Linguang Zhang
Supervisor: Professor Chi-Keung Tang

Motivation


Figure 1

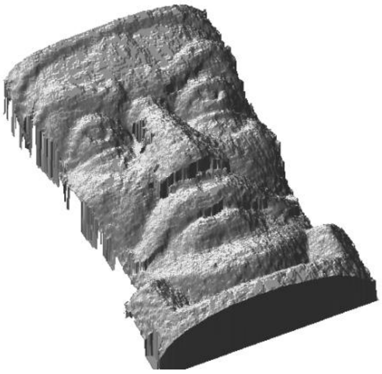


Figure 2


Figure 3

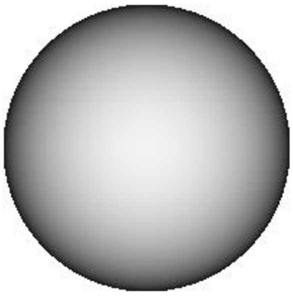


Figure 4

RGB-D image: traditional color images are usually composed of three channels – **R**ed, **G**reen, and **B**lue. However, people are no longer satisfied with 2D information only. Therefore Microsoft released Kinect (Figure 1), one of the most popular RGB-D cameras, where D means **D**epth. With the help of an infrared sensor, Kinect is able to tell the distance from the 3D scene point to the camera so that the 3D model of the current scene can be reconstructed. But the depth map captured by Kinect is often noisy and incomplete (Figure 2). As a result, most of the details cannot be preserved by Kinect.

Shape-from-Shading: shading is an important cue for human to recognize details. For example, although sculptures have no color information, people are still able to identify the 3D details by observing the shading (Figure 3). Based on this idea, computer vision people believe that it is possible to reconstruct the 3D model from a single RGB image only, and the classic Shape-from-Shading (SfS) problem is proposed. But one severe limitation that makes Shape-from-Shading an ill-posed problem is: different 3D models may correspond to the same shading, which is called the bas-relief ambiguity. For instance, the sphere shown in Figure 4 can be either concave or convex.

We found that Shape-from-Shading and RGB-D images can be highly complementary. Although Kinect is not able to capture the fine details, it can at least tell whether a surface is concave or convex. Therefore the bas-relief ambiguity of Shape-from-Shading is eliminated. We call this method “Depth-Aware Shape from Shading”.

Methodology

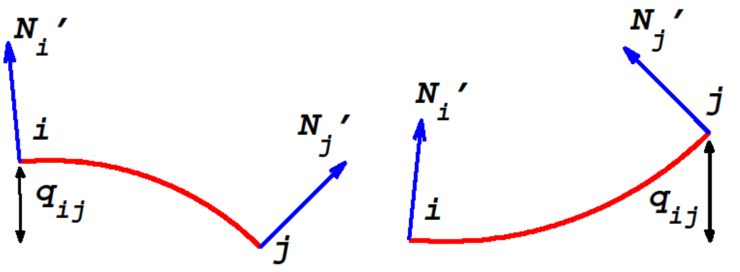


Figure 5

Light Source Estimation: our model assumes a single light source. But it can be easily proved that this model is also capable of handling the environment light. The Lambertian assumption is:

$$I = \rho \boldsymbol{N}^T \boldsymbol{L}$$

where I is the observed intensity value, $\boldsymbol{N}$ is the surface normal vector, and $\boldsymbol{L}$ is the light source direction. By trusting a small portion of normal vectors calculated from the depth map, we are able to estimate the light source direction $\boldsymbol{L}$.

Surface Normal Estimation: we are able to compute a rough normal map $\boldsymbol{N}_{rough}$ from the depth map given by Kinect. But since the quality of the depth map is low, $\boldsymbol{N}_{rough}$ can only be used as a reference, based on which we add the rough normal constraint. We also assume that the surface normal changes gradually, thus we add the smoothness constraint. Finally, the most important shape-from-shading constraint is added and the following weighted energy is generated:

$$E = E_{sfs} + \lambda_1 E_{smooth} + \lambda_2 E_{rough}$$

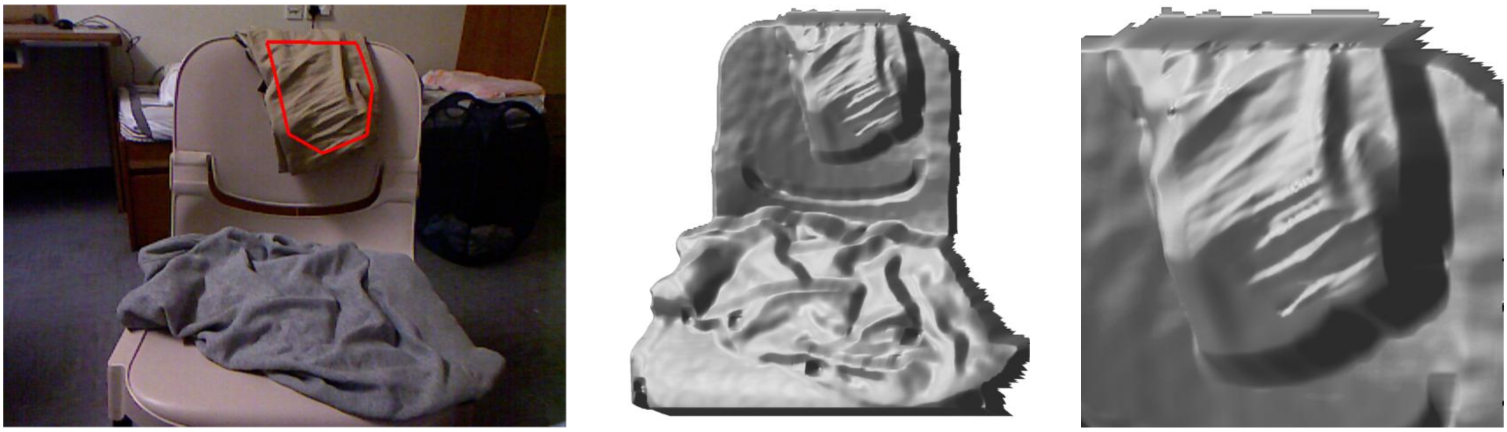
By minimizing the above energy function, the surface normal can be estimated.

Surface Reconstruction: the 3D surface can be reconstructed by interpolating a smooth connection between every two neighboring surface normal vectors. Figure 5 illustrates the basic idea of the interpolation.

Results

Interactive Refinement

Our method supports local optimization, thus users can specify the region to be processed.



Conclusion

We demonstrate SfS/depth synergy in achieving high-quality fully automatic surface reconstruction from single RGB-D images. Meanwhile, our approach can be formulated as a quadratic energy function that can be readily solved at an acceptable computational cost. Our experiments have also shown that when the scene is not adversely corrupted by shadows, the proposed method is robust against erroneous lighting estimation, inadequate resolution and quantization errors. Furthermore, to handle more complex situation, our method also supports local optimization where the user simply specifies a region of interest to process. Our RGB images used in the experiments are captured by Kinect, and the resolution is 640 × 480; a better RGB camera can significantly improve results on highly textured surfaces.