Computer Vision Lab 4 Report

Jingyan Li

1. MODEL FITTING

In the section we implemented least-squares and RANSAC to fit the data. The result is shown in Table 1 and is visualized in Figure 1.

Table 1. The estimated parameters in model fitting

	k	**b**
Ground Truth	1	10
Least-squares	0.6160	8.9617
RANSAC	0.9987	9.9970

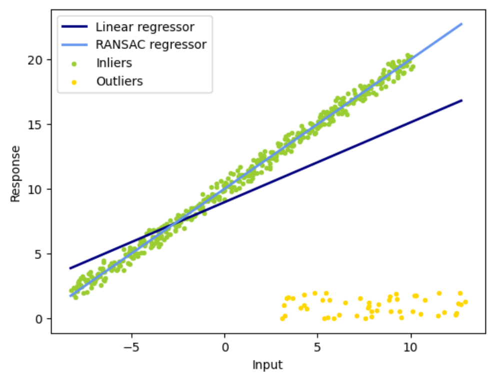

Figure 1. The ground truth and estimation by least-squares and RANSAC.

2. MULTI-VIEW STEREO

2.1 Differentiable warping

For pixel $\mathbf{p}$ in the reference feature and a depth value d_j, we can get the corresponding pixel $\mathbf{p}_{i,j} := \mathbf{p}_i(d_j)$ in the source view j according to (1).

$$\mathbf{p}_{i,j} = \mathbf{K}_i \left(\mathbf{R}_{0,i} \left(\mathbf{K}_0^{-1} \mathbf{p} d_{i,j} \right) + \mathbf{t}_{0,i} \right) \tag{1}$$

$\mathbf{K}_0$ and $\mathbf{K}_i$ represent the intrinsic matrix of reference view and source view i respectively. $\mathbf{R}_{0,i}$ and $\mathbf{t}_{0,i}$ are the rotation matrix and the translation vector from reference view to source view i. $d_{i,j}$ denotes the depth value of pixel $\mathbf{p}_i$ at depth index j.

2.2 Training

The training loss quickly converged after around 500 iterations in the first epoch. Due to the bug of tensorboard (Moodle Question), I cannot open it and show the loss curve. The loss is then stored in the log file in *log.txt*.

jingyli@ethz.ch

2.3 Test

We use geometric consistency filtering to filter out unreliable estimations. Given the camera parameters: $\mathbf{K}_0$, $[\mathbf{R}_0|\mathbf{T}_0]$ for reference view and $\mathbf{K}_i$, $[\mathbf{R}_i|\mathbf{T}_i]$ for source view i, we have the depth value estimation d_0,d_i for both reference view and source view i. In geometric consistency filtering, we firstly project point $\mathbf{p}_j$ from reference view to the source view i and denote as $\mathbf{p}_{j,i}$. Then we can get its estimated depth value $\hat{d}_{i,j}$ in source view i by interpolating d_i. Then we reproject $\mathbf{p}_{j,i}$ with depth $\hat{d}_{i,j}$ back to reference view and get the reprojected coordinates of $\mathbf{p}_j$, which is denoted as $\hat{\mathbf{p}}_j$. Meanwhile we also get the reprojected depth value $\hat{d}_0[i,j]$ at $\hat{\mathbf{p}}_j$ in the reference view.

Then we can calculate the divergence between original data and the reprojected data in the reference view. We can get the depth difference by $\Delta depth_{i,j} = \hat{d}_0[i,j] - d_0[j]$, which is the difference between estimated depth value and reprojected depth value in the reference view. Besides the distance deviation between original points and reprojected points in the reference view is calculated by $dist_{i,j} = ||\mathbf{p}_j - \mathbf{p}_{j,i}||$.

We finally can filter out unreliable points with large distance deviation and large depth difference.

The visualization of test scans are shown in Figure 2 and Figure 3.

Figure 2. The reconstructed point cloud for test dataset - scan 1.

Figure 3. The reconstructed point cloud for test dataset - scan 9.

2.4 Discussion

It is suitable for large-scale scenes. According to (2), the disparity d is inversely related to the depth Z. Here, f denotes focal length and b denotes the baseline.

$$Z = \frac{fb}{d} \tag{2}$$

Then we get (3).

$$\Delta d = fb(\frac{1}{Z_2} - \frac{1}{Z_1}) = fb\frac{\Delta Z}{Z_2 Z_1} \tag{3}$$

In the large-scale scenes, the depth value Z_2 and Z_1 is very large compared to the change of depth ΔZ between points. Then the uniformly sampled points in the reference view can be very close in the source view, namely small Δd. To enlarge the distance between sampled points in the source view, we can utilize sampling in the inverse range (i.e. $\frac{1}{\Delta Z}$).

In the method, we take average when integrating the matching similarity from source views. However, in multi-view stereo images, some parts of the scene cannot be seen by all stereos. Some views with such occlusions hold smaller similarity to the reference view. Simply taking average of multi source views makes similarity impacted by such occlusions and therefore the method may not achieve robustness in this way.