EE5731 Assignment 2

Tan Pin Quan
A0067152J

1 Part 1: Noise Removal

In this task, the noise from the image is removed using MRF and Graphcuts. The two labels are 0 for foreground and 1 for background and their transformations to RGB values are as follows:

$$T(x_i) = \begin{cases} (0,0,255) & if\ x_i = 0 \\ (245,210,210) & if\ x_i = 1 \end{cases}$$

MRF and Graphcuts are used to find the set of labels for each pixel optimal labels $\{x^*\}$ where

$$\{x^*\} = \underset{\{x\}}{\operatorname{argmin}} \sum_i f_d(x_i, d_i) + \lambda \sum_{j \in N_i} fp(x_i, x_j)$$

The prior term is a smoothness constraint where for two points x_i and x_j:

$$f_P(x_i, x_j) = |x_i - x_j|$$

The data term is the average difference in RGB values from the predefined foreground values

$$f_d(x_i, d_i) = |(x_i) - d_i|$$

1.1 Results and Discussion

The file 'Part1.m' performs the noise removal. The matlab wrapper to perform the graphcuts is used with for λ = 1, 10, 100, 1000. The results are shown in Figure 1. When λ = 1, much of the noise is still not removed as the impact of the smoothness constraint is very small. As λ increases to 10, more noise is removed, and at $\lambda = 10$, all the noise is removed. However, the gap in the letter 'e' becomes closed. As $\lambda = 1000$, the smoothness constraint causes the letters to merge into one another.

This shows that an appropriate selection for λ is critical. Too small and the result will be noisy, too large and the result will be over-smoothed.

Figure 1: Noise reduction for lambda = 1, 10, 100 and 1000

2 Part 2: Depth from Rectified Stereo Images

In this task, the depth map is obtained from a pair of rectified images using MRF and Graphcuts. From observing the images, it can be observed that the maximum disparity is less than 60. While the disparity from image 1 to 2 is always positive, the disparity from image 2 to 1 is always negative. Therefore, both disparity directions are used. Therefore, the range of disparity labels is given as

$$\{D\} = \{-60, -59, \dots, 58, 60\}.$$

The optimal disparity map D^* is given as

$$D^* = \underset{\{D\}}{\operatorname{argmin}} \sum_x \left(f_d(x, D(x), I, I') + \lambda \sum_{y \in N_x} f_p(D(x), D(y)) \right)$$

The prior term is a smoothness constraint where for two points x and its neighbours $y \epsilon N_x$:

$$f_p(D(x), D(y)) = \min\left(6.5,\ \left(D(x) - D(y)\right)^2\right)$$

Therefore, the smoothness constraint is limited to a maximum value of 6.5.

In rectified images, for a point x in image I, the epipolar line is the corresponding horizontal row in image I'. The data term is the difference in RGB pixel values along the epipolar line.

$$f_d(x, D(x), I, I') = \left(I(x) - I'\left(x + D(x)\right)\right)^2$$

2.1 Results and Discussion

The file 'Part2.m' obtains the depth map. The data term in shown in Figure 2 and it there is a large amount of noise. The depth map for image 1 and image 2 is shown Figure 3 for a λ value of 5. The effect of smoothing has significantly reduced the noise from the data term. It can also be seen that the depth for the extreme left of image 1 and extreme right of image 2 is not accurate. This is because those sections cannot be found in both images and the data term is inaccurate.

Other small sections where the depth is inaccurate occur when the regions are the same colour. This causes the data term to be inaccurate due to ambiguities in the photo-consistency property

Figure 2: Data term for both images

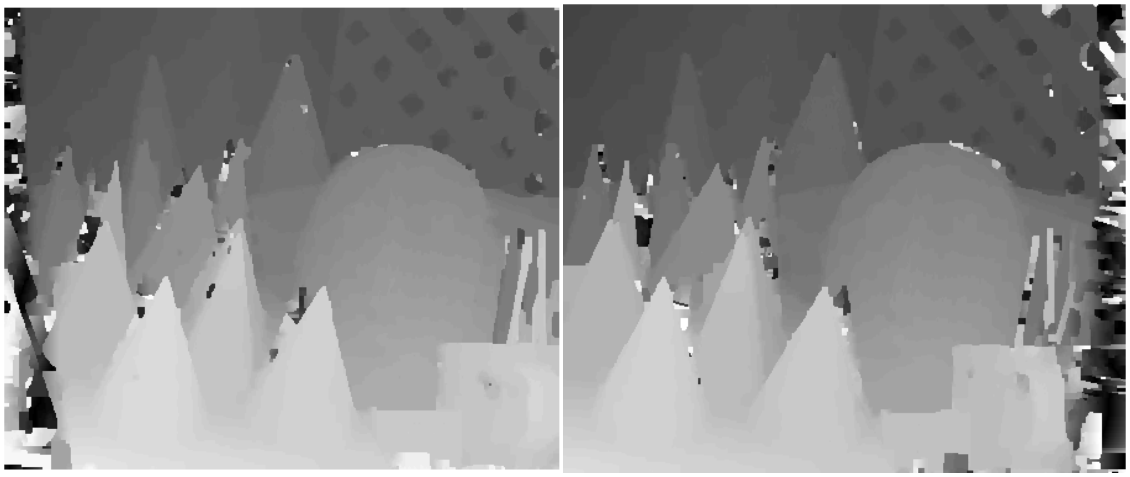

Figure 3: Depth map for both images

3 Part 3: Depth from Stereo

In this task the depth map is obtained from a pair of unrectified stereo images using the camera matrices. The process is similar from part 2 expect that the epipolar line l is now given as

$$\boldsymbol{x'} = \boldsymbol{K'R'^{T}RK^{-1}x} + d_x\boldsymbol{K'R'^{T}(T - T')}$$

where d_x is the disparity.

The prior term is a smoothness constraint where for two points x and its neighbours $y \epsilon N_x$:

$$f_p(D(x), D(y)) = \min\left(2e^{-8}, \left(D(x) - D(y)\right)^2\right)$$

Upon observation it can be seen that the optimal value of d_x lies between 0 and 0.0099. and a step size of 0.001 results in about 1 to 2 pixel resolution. This is shown in Figure 4 where the epipolar line for a point with one of the largest disparities in the image is shown. Therefore, values of d_x chosen are between $\{0, 0.001, 0.002, \ldots, 0.099\}$

Figure 4: A point on image 1 (left and the corresponding epipolar line where d = {0, 0.001, ... 0.0099}

3.1 Results and Discussion

The file 'Part3.m' obtains the depth map. The data term for both images is shown in Figure 5. The depth map for both images obtained is shown in Figure 6 using a λ value of 2e7. A larger smoothing value is used as the data terms are very noisy.

The results are not as accurate as desired. For example, due to occlusion from the stop sign, the region to the left of the sign in image 1 is being occluded by the sign in image 2. Therefore, the data term obtained is inaccurate and a smearing can be seen to the left of the stop sign in the depth map. Also, the sections on the right of image 1 cannot be found in image 2 and vice versa, once more resulting in poor depth in those regions.

Figure 5: Data term for image 1 and image 2

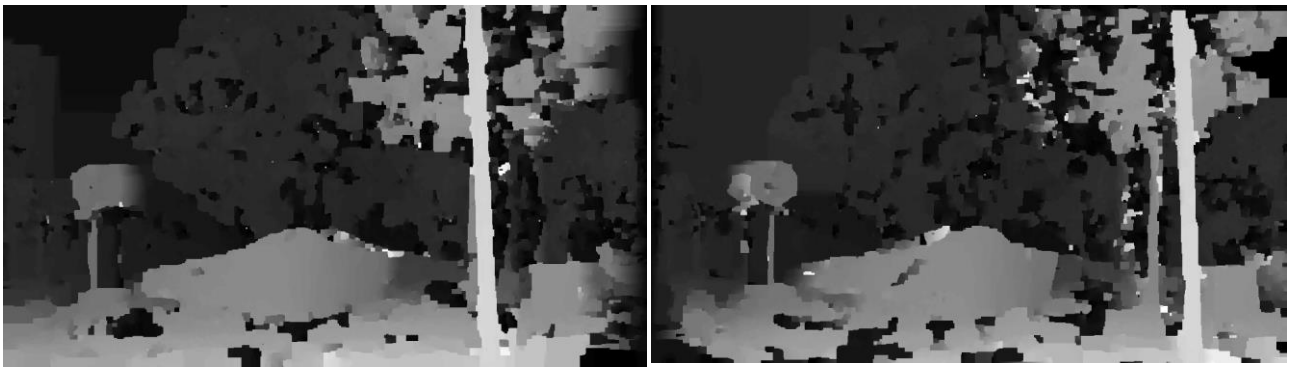

Figure 6: Depth map for image 1 (left) and image 2 (right)

Improvement using Bundle Adjustment

To improve the result, bundle adjustment is also performed using the method described in Part 4. This is implemented in the file 'Part3a.m'. The parameters used were $\lambda = 1250$, $\sigma = 1$, $\sigma_d^2 = 50$. The results of after initialization and bundle adjustment for both images are shown in Figure 7 and Figure 8. Performing bundle adjustment allows for a relative lower smoothness term, which increases the resolution.

However more images will allow for much better results which is shown in the next section.

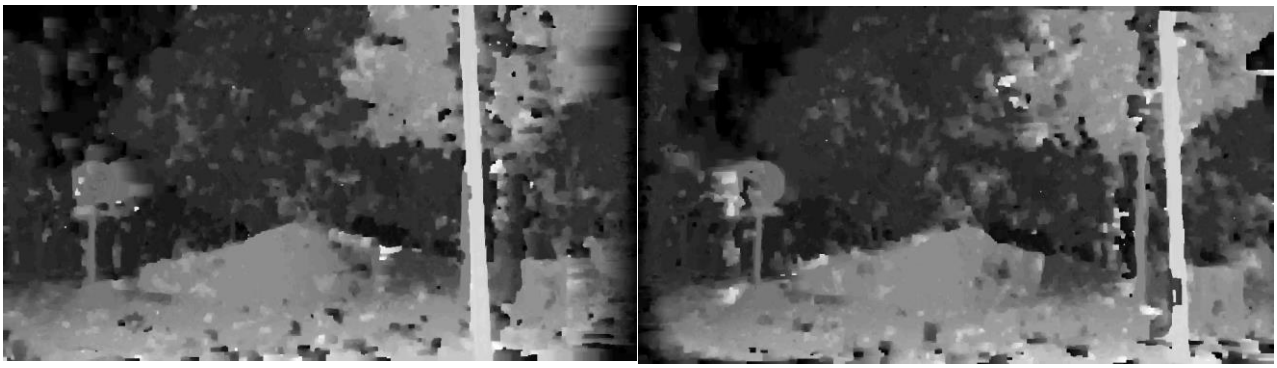

Figure 7: Initial depth maps for image 1 and image 2

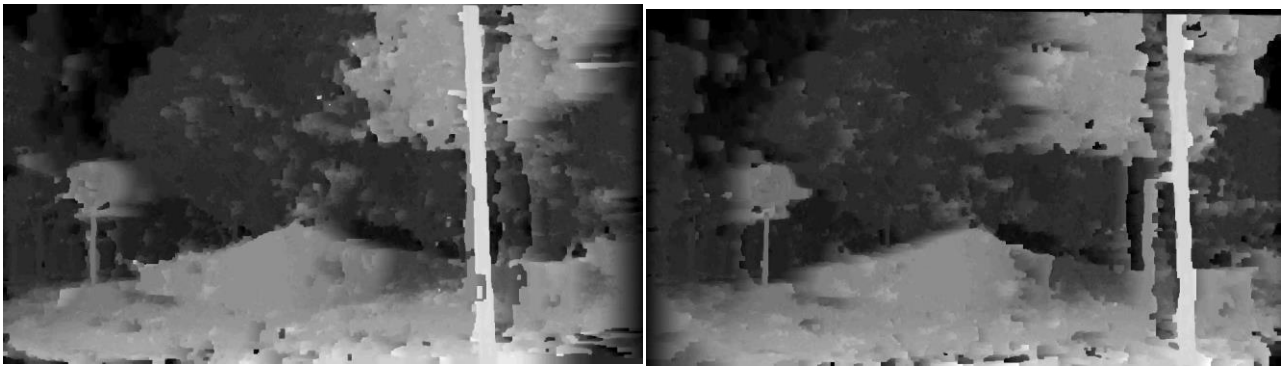

Figure 8: Adjusted depth maps for image 1 and image 2

4 Part 4: Depth from Video

In this task, the depth is obtained from 50 video frames beginning from frame 90 to frame 139. For each frame t, a window of frames t-10 to t+10 are used to obtain the data term and perform bundle adjustment.

By keeping the window size fixed at 20 instead of using all 50 images, the computation time only increases linearly with the number of images processed

Disparity Initialisation

For disparity initialisation, we obtain the disparity maps D_t of all images

$$D_t^{init} = \underset{\{D\}}{\operatorname{argmin}} \sum_x \left(1 - u(x) f_d^{init}\big(x, D_t(x)\big) + \lambda \sum_{y \epsilon N_x} f_p\big(D_t(x), D_t(y)\big) \right)$$

The prior term is a smoothness constraint where for two points x and its neighbours $y \epsilon N_x$:

$$f_p(D(x), D(y)) = \min(0.004, \|D(x) - D(y)\|)$$

The data term is given as

$$f_d\big(x, D_t(x)\big) = \sum_{t\prime}^{N} \rho_c(x, d, I_t, T_{t'})$$

where the photo-consistency term is given by.

$$\rho_c(x, d, T, T_{t'}) = \frac{\sigma_c}{\sigma_c + \left\| I_t(x) - I_t'\left(l_{t,t'}(x, d)\right) \right\|}$$

MRF and graphcuts is used to obtain the initial disparity maps.

Bundle Optimisation

After disparity initialisation, bundle adjustment is performed to account for the geometric properties and reduce flickering. The data term is changed to account for the geometry:

$$f_d\big(x, D_t(x)\big) = \sum_{t\prime}^{N} \rho_c(x, d, I_t, T_{t'}) \rho_v(x, d, D_{t'})$$

where the geometric coherence term is given by

$$\rho_v(x, d, D_t{}') = \exp\left(- \frac{\left\| \bar{x} - l_{t',t}(\bar{x}', D_t(\bar{x}') \right\|^2}{2\sigma_d^2} \right)$$

4.1 Results

The file 'Part4.m' computes the depth map for 50 images. The parameters used were $\lambda = 500$, $\sigma = 1$, $\sigma_d^2 = 50$. Each image uses its neighbouring 20 images to perform the disparity initialization and bundle optimisation. The initial and adjusted depth maps of frame 12 and 20 are shown below. The full video of the initialised disparity maps and adjusted disparity maps can be found in the files *'dispartiy_ init.avi'* and *'disparity_ adjust.avi'*. The matlab script 'PlayResults.m' also plays the video for the initialised and adjusted depth maps.

The result after using 50 images is much improved from the result in Part 3. This shows how using multiple images improves the result of the depth map. The amount of flickering is also significantly reduced after performing the bundle adjustment as the geometric properties of each frame is taken into account. This is especially evident in the videos provided.

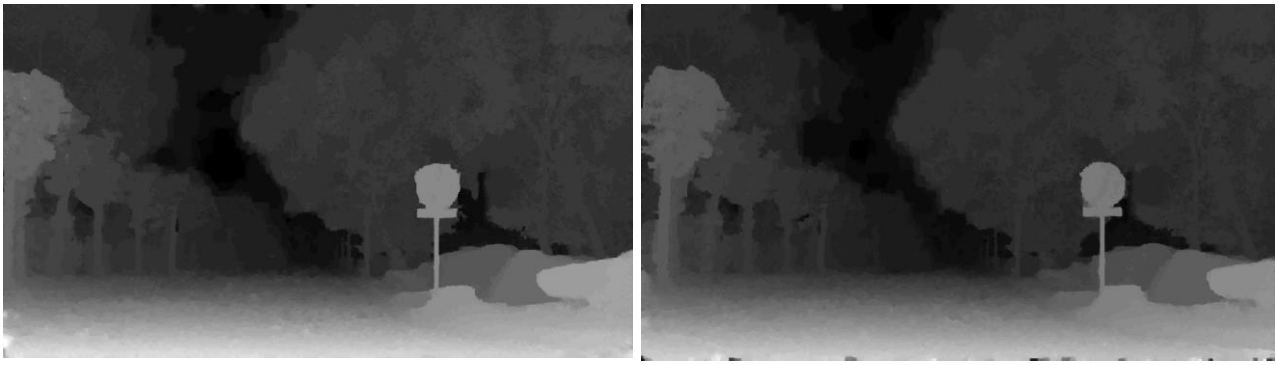

Figure 9: Initial depth map for frames 12 and 20

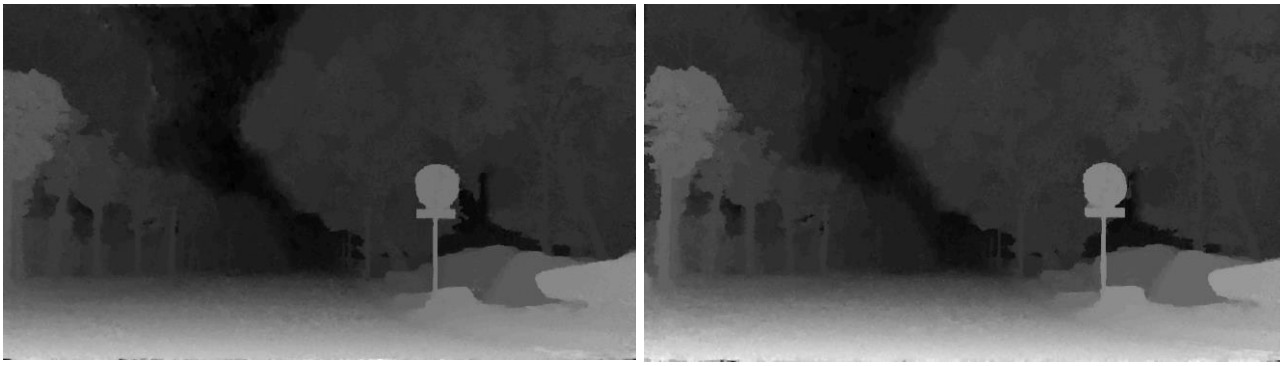

Figure 10: Adjusted depth map for frames 12 and 20

5 Part 5: Possible Improvement

One drawback of the implementation is that the extrinsic parameters **R**, **T** and c need to be known for each frame. Originally, only the fundamental matrix F was required to obtain the epipolar line using the equation $\boldsymbol{l'} = \mathbf{F}\boldsymbol{x}$. Note that the notation used follows the paper "Comments on Consistent Depth Maps Recovery from a Video Sequence"

One possible improvement is to obtain the extrinsic parameters **R** and **T** by using matching point pairs $\{\mathbf{x}, \mathbf{x}'\}$ from the image and the intrinsic parameters **K**. This is more practical as the intrinsic parameters are easier to obtain and in the case of a video, they are identical for all frames. In this case, the point pairs were simply manually obtained but this can be obtained using SIFT.

The steps are therefore as follows:

1. From the point pairs estimate the fundamental matrix **F**
2. Using **K** obtain the essential matrix using $\mathbf{E} = (\mathbf{K}')^{\mathrm{T}}\mathbf{F}\mathbf{K}$
3. From **E** obtain the extrinsic properties **R** and **T**
4. Obtain the depth map using the same procedure as part 4

These steps are performed on the images shown below in the file 'Part5.,'

Figure 11: Images used

5.1 Obtaining F

F is obtained using the 8-point algorithm where

$$[x'x \quad x'y \quad x' \quad y'x \quad y'y \quad y' \quad x \quad y \quad 1]\begin{bmatrix} f_{11} \\ f_{12} \\ \vdots \\ f_{33} \end{bmatrix} = 0$$

Using the F obtained, the epipolar line can be drawn as shown below

Figure 12: Epipolar lines from fundamental matrix

5.2 Obtaining E, R, T and C

The essential matrix E can be obtained from F K and K':

$$\mathbf{E} = (\mathbf{K}')^{\mathrm{T}}\mathbf{F}\mathbf{K}$$

The extrinsic parameters can then be obtained by performing SVD on E

$$\mathbf{E} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$$
$$\mathbf{R} = \mathbf{U}\mathbf{W}\mathbf{V}^{\mathrm{T}} \text{ or } \mathbf{U}\mathbf{W}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}$$
$$\mathbf{T} = \pm\mathbf{U}\mathbf{W}\mathbf{D}\mathbf{U}^{\mathrm{T}}$$

where $\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, and $\mathbf{Z} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

This gives four possible solutions for R and T and the correct solution is chosen. This then gives four possible vales c is then computed as

$$\mathbf{c} = -\mathbf{R}^{\mathbf{T}}\mathbf{t}$$

5.3 Depth Map

A similar method as part 4 is used to obtain the depth map. The epipolar line is then defined as

$$\mathbf{x}' = \mathbf{K}'\mathbf{R}'\mathbf{R}^{\mathbf{T}}\mathbf{K}^{-\mathbf{1}}\mathbf{x} + \mathbf{d}\mathbf{K}'\mathbf{R}'(\mathbf{c} - \mathbf{c}')$$

With 4 possible sets of {c, R}, for each pixel, there are 4 possible epipolar lines. In practice, only 2-3 lines appear within the image boundaries. The epipolar lines are shown below for multiple points. The depth can then be found using the same MRF and graphcuts similar to part 4. The only modification is to search though all the 4 possible epipolar lines instead of one.

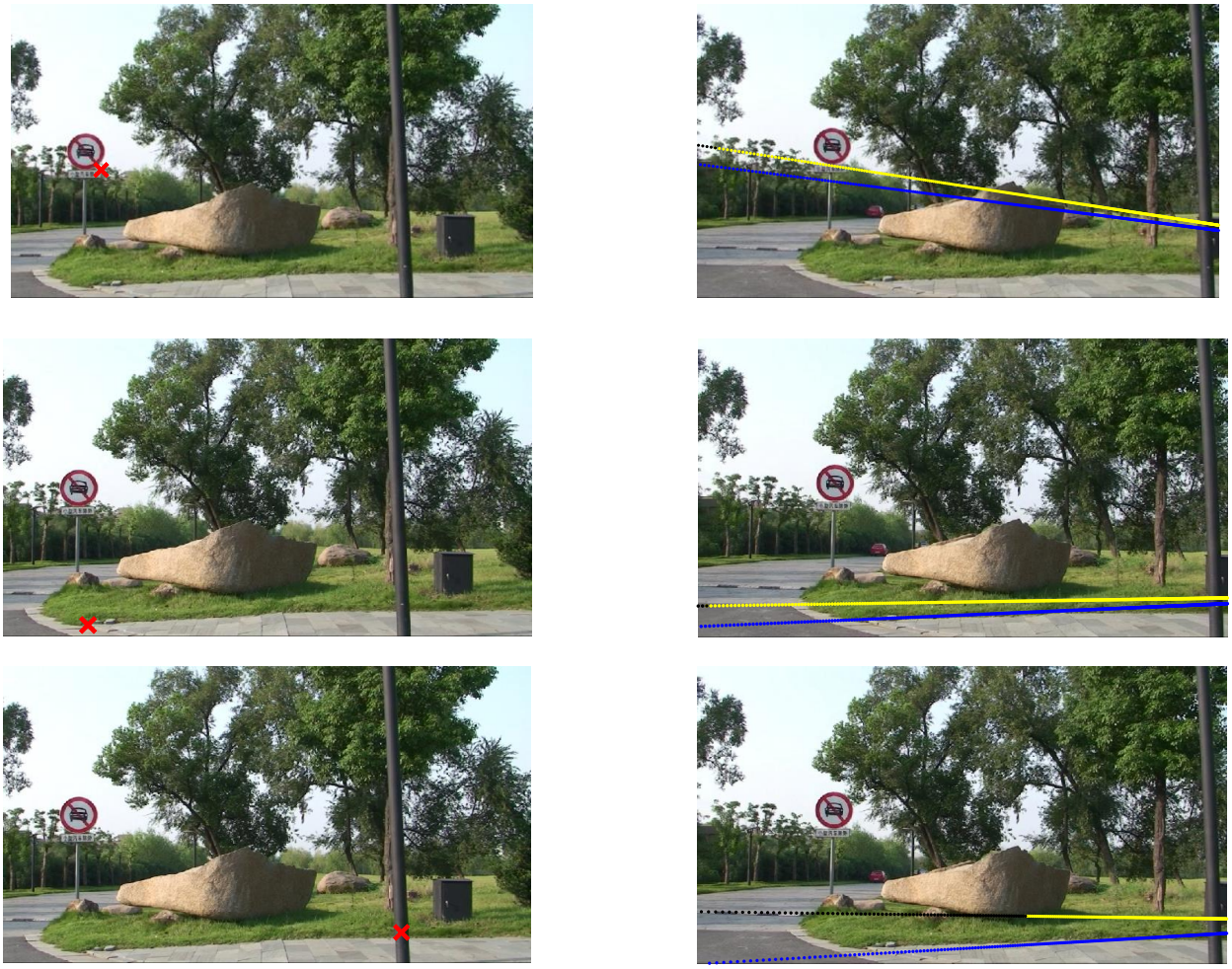

Figure 13: Epipolar lines obtained from estimated extrinsic parameters.

5.4 Discussion.

Unfortunately, due to lack of time, the depth portion was not implemented. However, the proposed improvements show that the depth can also be obtained without knowledge of extrinsic parameters R and T. The search space will be larger as there are multiple lines to search. In addition, the effect of noise may cause the estimated extrinsic parameters. There may also exist a way to determine the correct set of R and T to return to the same case as part 4.