CS 532: Homework Assignment 4
Due: November 4, 6:00pm

Philippos Mordohai
Department of Computer Science
Stevens Institute of Technology
Philippos.Mordohai@stevens.edu

Collaboration Policy. Homeworks will be done individually: each student must hand in their own answers. It is acceptable for students to collaborate in understanding the material but not in solving the problems. Use of the Internet is allowed, but should not include searching for previous solutions or answers to the specific questions of the assignment. I will assume that, as participants in a graduate course, you will be taking the responsibility of making sure that you personally understand the solution to any work arising from collaboration.

Late Policy. No late submissions will be allowed without consent from the instructor. If urgent or unusual circumstances prohibit you from submitting a homework assignment in time, please e-mail me explaining the situation.

Submission Format. Electronic submission on Canvas is mandatory. Submit in a zip file, **a single pdf file containing**:

- the source code,
- brief explanations of what was done,
- images, disparity/depth maps and error maps.

Also include the source code separately.

Problem 1. Download the *cloth3* dataset from the course web page. The zip file contains 7 views taken from a camera that moves perfectly horizontally on a rail in 40mm increments. The baseline between consecutive views, view0 and view1 for example, is 40mm, while the baseline between view1 and view5 is 160mm. The zip file also contains ground truth disparity maps for view1 and view5. These were generated using structured light techniques. Views 1 and 5 were assumed to be the reference and target image for disp1 and their roles were reversed for disp5. The disparity values have been multiplied by 3 before being stored in the image files to enhance contrast.

The focal length of the camera remains constant at 1247 pixels and the principal point is at the center of all images. The disparity range between views 1 and 5 is from 0 to 85.

(a) Use the provided ground truth disparity maps for views 1 and 5 to generate the ground truth disparity map for view 3. Disparity values equal to 0 mean missing ground truth and should be ignored. Start by rendering the disparity map from view1 onto view3. In order to do that, you have to generate a 3D point in the camera coordinate system for each valid disparity value. Then, each 3D point has to be transformed to the global coordinate system and projected onto view 3. It is up to you to store a disparity or a Z (depth) value at the pixels of view 3. Repeat the same process using view 5 as the source and view 3 as the destination. When multiple 3D points project on the same pixel, pick the one nearest to the camera.

Show the resulting disparity or depth map in your report. Make sure that the range of the image is appropriate for the values being displayed.

(b) Without using the ground truth disparity any more, compute binocular depth maps for three views: view 1, view 3 and view 5 using views 2, 4 and 6 as target views respectively. Use your code form homework assignment 2 and SAD in 9×9 windows (or any other size of your choice). The disparity range for these stereo pairs is 0 to 22. **Why?**

Combine the three disparity/depth maps on view 3, by rendering all reconstructed 3D points onto it using two strategies. For each pixel, select:

1. the candidate that is nearest to the camera, or

2. the candidate with the smallest SAD cost.

Evaluate the two resulting disparity/depth maps using the ground truth generated in part (a) and report the error rates, i.e. the percent of pixels with disparity estimates more than 1 level off the ground truth. Also, show the obtained depth or disparity maps.

Hints

- Do not forget to divide the ground truth disparity values by 3 (and check the types of your variables).

- Disparity values in disp5 should have been negative following the conventional notation. They are positive for convenience and for ease of visualization.

- If you chose to use disparity maps throughout, think about the actual effects of the baseline in each step. Make sure your choices are consistent.

- Depth maps here and in most computer vision literature store Z values, not distances from the camera center.

Acknowledgement. The input images and disparity maps have been generated and made available by Daniel Scharstein and his students at Middlebury College.