s2p-hd: Gpu-Accelerated Binocular Stereo Pipeline for Large-Scale Same-Date Stereo

Supplementary Material

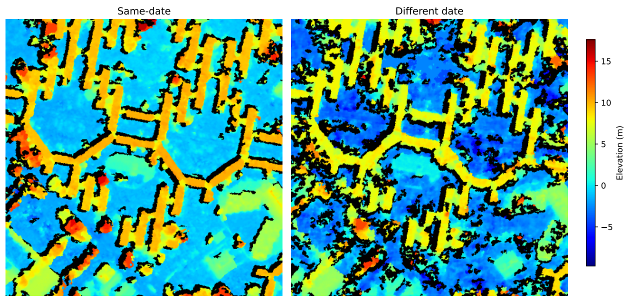

Figure 9. The two DSMs correspond to the same zone from the IARPA dataset. The one on the left was created with two same-date images, while the one on the right was constructed from images taken at different dates, the first one in December 2015 and the second one in November 2014.

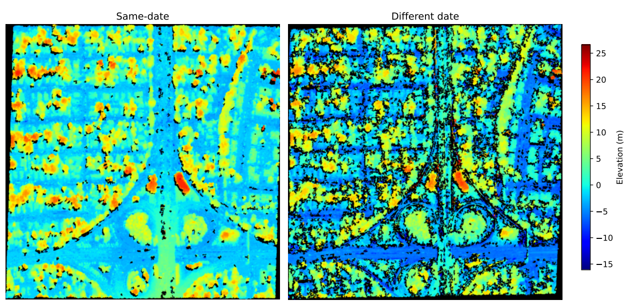

Figure 10. The DSMs both correspond to the *JAX_017* zone from the GRSS dataset. The one on the left was created with two same-date images, while the one on the right was constructed from images taken at different dates.

6. Method

6.1. Same-date and different-date images

We show more examples of the difference in the DSMs constructed from same-date images and images from different dates, on Figures 9, 10 and 11. The difference in shadows between images from different dates is the main factor for the lower quality of the DSM computed; a problem that is far less present with same-date images.

6.2. Input images for experiments

We display on Figure 12 the pair of images that was chosen to run the experiments for the IARPA challenge dataset, such that the same-date criterion was respected. These are two large images, of size 36,864x43,008 pixels.

Figure 13 shows two examples of image pairs selected

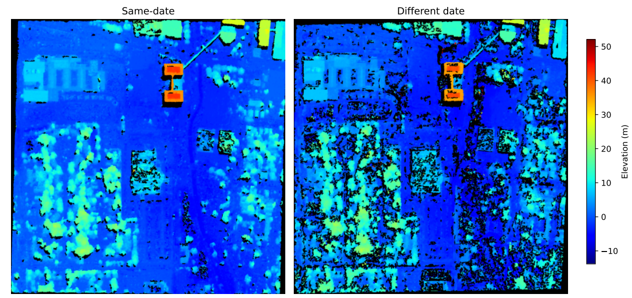

Figure 11. The DSMs both correspond to the *JAX_070* zone from the GRSS dataset. The one on the left was created with two same-date images, while the one on the right was constructed from images taken at different dates.

Figure 12. Image pair chosen from the IARPA challenge dataset. The two images were taken on the same date, December 18, 2015, and were taken 22 seconds apart. The DSMs of the different zones where ground truth is given will be computed with this image pair.

for the GRSS dataset. Each image is of size 2048x2048 pixels.

7. Experiments & results

7.1. Nuth & Kääb algorithm

The Nuth & Kääb method is widely used for aligning elevation models obtained from different sources, ensuring a fair error evaluation. The procedure consists of the following steps:

1. **Initial Vertical Offset Correction**
   - Compute the elevation difference $dZ = \mathrm{DSM}_{\mathrm{pred}} - \mathrm{DSM}_{\mathrm{GT}}$.
   - Estimate the mean (or median) vertical shift and subtract it from the predicted DSM.
2. **Horizontal Shift Estimation**
   - Compute the elevation difference as a function of slope and aspect.

Figure 13. Examples of selected pairs for 2 Jacksonville areas, namely *JAX_117* and *JAX_214*, from the GRSS dataset.

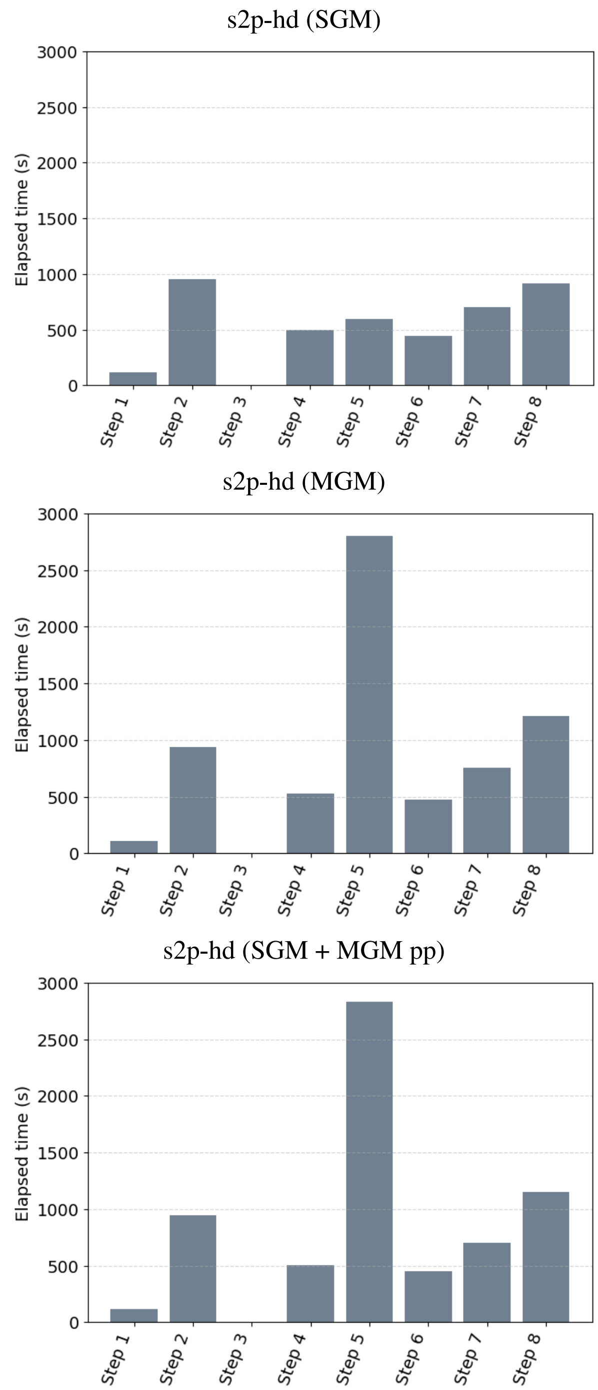

Figure 14. Runtime for each stage of the s2p-hd pipeline, with the different stereo matching strategies: SGM, MGM and SGM with MGM subpixel refinement post-processing.

   - Determine the systematic shift by analyzing elevation errors along different terrain orientations.
   - Estimate the optimal east-west (Δx) and north-south (Δy) shifts.
3. **Shift Correction and Re-Evaluation**
   - Apply the computed shifts to the predicted DSM.
   - Recompute the elevation error statistics to verify alignment.
   - Iterate until the residual shift is negligible.

7.2. Time distribution

We show on Figure 14 the distribution of runtime for each step of the s2p-hd pipeline, which are:

1. Image tiling and masked tiles discarding.
2. Local epipolar correction.
3. Global epipolar correction.
4. Tile rectification and disparity range estimation.
5. Stereo matching.
6. Tile triangulation: constructing the 3D point cloud from the matched disparities.
7. Local DSM rasterization: computing the DSM of each tile from the point clouds.
8. Global DSM rasterization: merging all the local DSMs into a global one.

We highlight the runtime advantage of the GPU-accelerated SGM integration in s2p-hd over other stereo matching methods. As shown on Figure 15, stereo matching in the standard s2p pipeline takes nearly five times longer than in s2p-hd with GPU-accelerated SGM. It is also about 1.5 times slower than the s2p-hd pipeline when using MGM.

7.3. San Diego experiments

We extended our evaluation to include additional regions around San Diego, which offer a broader range of land cover characteristics compared to the original urban-focused

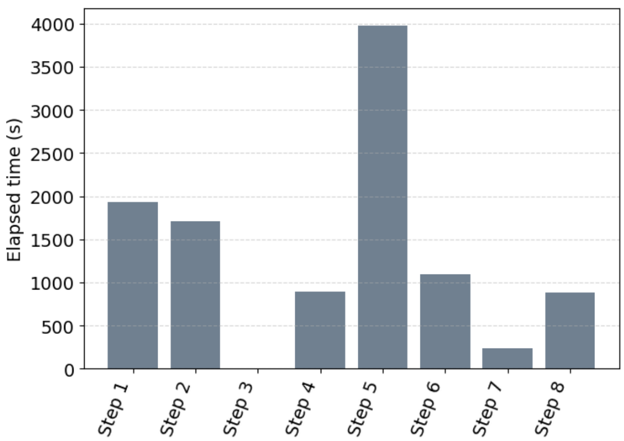

Figure 15. Runtime for each stage of the s2p pipeline, using MGM for stereo matching.

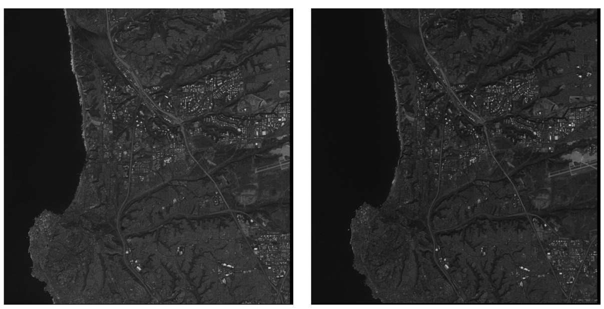

Figure 16. Image pair chosen from the San Diego dataset, acquired on October 27 and 28, 2014. The left part of the images corresponds to water bodies.

datasets. Unlike the dense urban areas in the IARPA and GRSS datasets, the San Diego scenes include significant natural terrain variability. The ground truth for this area was extracted [30] from a 2014 USGS lidar acquisition.

Figure 16 shows the image pair on which the experiments were conducted. These additional experiments allow us to better assess the robustness of our pipeline in non-urban, low-texture, and high-relief environments. Results are shown on Table 1 in the main paper. The results confirm that s2p-hd maintains competitive accuracy and runtime performance even under these more varied conditions, supporting its generalizability beyond structured urban scenes.