Matthew Jennings

Assignment 2 – Stereo Vision

Introduction

The purpose of this report is to examine how altering the parameters of the stereo process affects the result across a few different image types. I will be looking at using block matching and semi-global matching and how the changed values alter their product. To do this I will be examining them both separately and then comparing them to each other.

The goal is to generate the "best" disparity map possible for each of the three image sets. Defining best is a challenging exercise as there will likely arise a trade-off between accuracy and smoothness so the target should be a map with minimal noise but without losing too much information from smoothing.

The images used are at a resolution of 1920x1080 however the disparity maps will be generated and presented using these images scaled by a factor of 4.

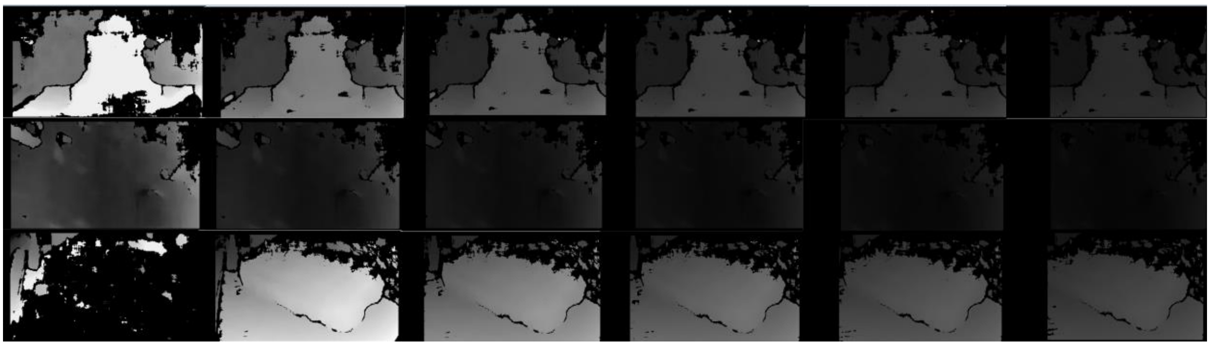
Block Matching

There are two parameters that can be altered in block matching; the max disparity which defines the range in which disparities are described; from 0 to the given value where this maximum value determines how close an object can be to the camera, and the block size which defines how many pixels to look at around the relevant pixel in order to perform matching.

My expectation going in is that a lower max disparity will result in more compressed information with less granularity resulting in a smoother less noisy disparity map but with less accuracy. It will likely also result in less overall brightness in the final image, but this is more a result of the way in which the maps are viewed, relative brightness within the image should be more meaningful than absolute. Regarding block size, I expect that a lower value will result in a more detailed map but with more room for wrong matches to arise causing more noise, with higher values I expect the opposite; a smoother image with less detail but also less noise arising from inaccuracies.

The first parameter considered will be the max disparity, while the block size is fixed at 21. There is an issue with this methodology in that the "best" max disparity might not be the same at all block sizes and trying to optimize two variables at the same time is an imperfect process. The results of this might be different if instead block size was optimized first and then disparity.

Above are image sets used, with the left on top and the right under.



16 32 48 64 80 96

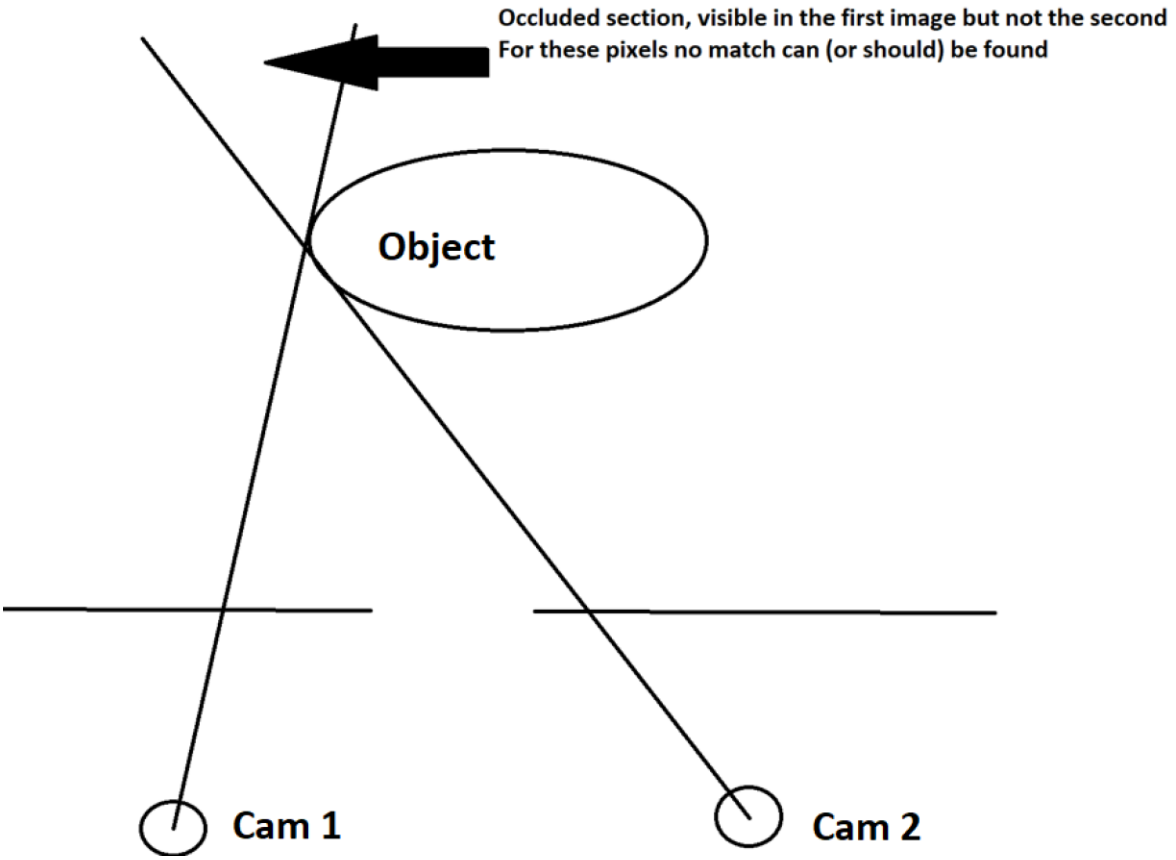
As we can see from these disparity maps resulting at different max disparities there is no universally optimal max disparity to use, and it varies wildly based on both the total depth within a scene and the subject of the scene.

For the bell, which is a relatively close object with a smooth surface, we get meaningful results at all max disparities, although with variations in the amount of noise. For this image based on an entirely subjective visual examination 32 seems like the correct value. At higher values we see significantly diminishing returns on noise reduced.

The archway scene is less clear, but it appears that as the subjects of the scene are further from the camera a shorter distance between matched pixels is to be expected so a lower max disparity yields better results, in this case 16, the minimum maximum, appears to yield the best result.

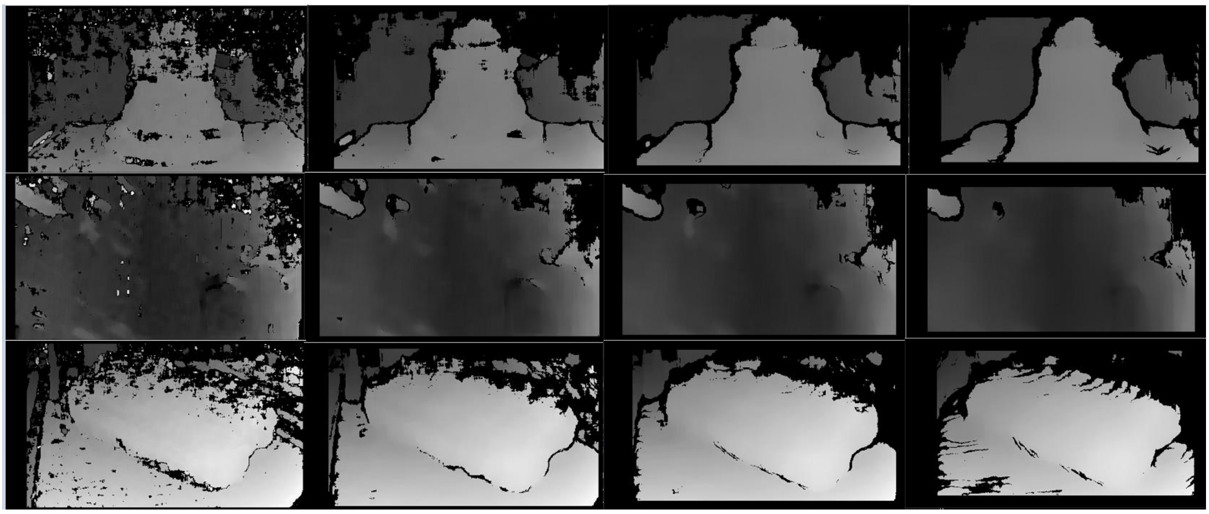
The rock is a rough surface but in none of these maps (at least at this block size) is that evident, the results are very similar to the bell just by the fact that they are a similar distance from the camera, this also seems like a case where 32 is best although an argument could be made for 48 or even 64.

Half occlusions, the black void moving across from the left increases in size consistently across all images as the max disparity increases. This space represents pixels for which no match in the first image was found in the second. These occlusions are also evidenced by the black outlines most evident in the bell and the rock, which are much closer to the camera and so when attempting to find a match for those pixels in the second image they do not exist as they are now blocked by the movement of the object within the scene.



This suggests that the algorithm does not attempt to find matches from the first image in the second image within the range of the max disparity. As we have specified the max disparity, we have specified the amount we expect the object we are mapping to move within the image, therefore there cannot be a match within the first (maxDisparity) pixels of the scan lines in the second image. This to me makes this method seem more suited to mapping a single object with relatively consistent disparities than an entire scene with potentially wildly different disparities within it.

From here we use these semi-optimized disparity values for each scene to attempt to determine optimal values for the block size.



11 21 31 41

The effect here is clear, a higher block size creates a smoother less detailed image with reduced noise. It is also clear how the pixels that are not able to have a block generated around them are handled, they are ignored creating a larger black border around the disparity map as the block size increases.

There is a balance to be had here between accuracy and smoothness and which setting is best is not only entirely subjective but very dependent on what the goal is.

For the bell 31 strikes a fair balance but going to 41 does further reduce noise but at the expense of losing some of the clear definition of the bell. In the other direction more definition of the bell is clear, but more noise is introduced. We can also see that the boundary between the bell and the background becomes less accurate to reality and in some places, we get bleeding of the background into the bell and vice versa. This is a product of the increased window size leading to more ambiguity at the edges of the object where the depth dramatically shifts and so the partial occlusion created by the bell on the background shifts between the images resulting in less clear matches as the content of the background that we are now examining more of changes and the values within the bell itself dominate the disparity calculation as they are still examined as the background is mapped.

The archway scene is not overly meaningful, block matching is just not designed to deal with a scene of this type. The problem is that this method of matching operates on the assumption that the scene is not only mostly planar, but also co-planar with the camera projection. This is most evidently not the case in the archway scene with the walls being at angles and very little consistency of “blocks” between the images. When a block is searched for from one image to the next it is not so much translated and accurately matched as slanted at a different angle resulting in it being very difficult to find an accurate match with this method as the pixels within the window are very dissimilar despite being of the same object. 21 seems to strike the best balance between noise and smoothness.

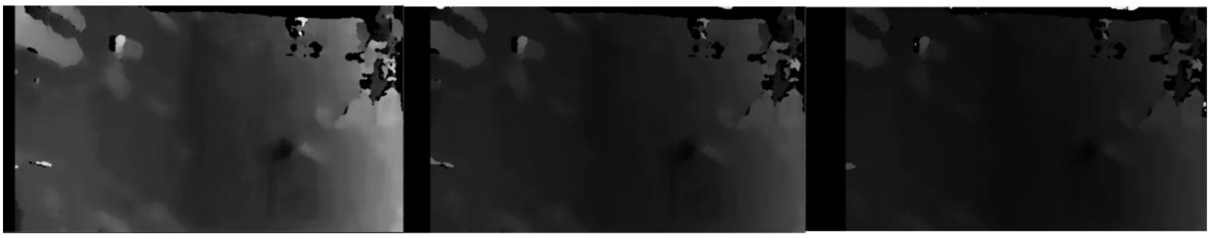
The rock is very similar to the bell again but with a few key differences. As we apply larger block sizes, we start to introduce more artifacts into the map, visible at the top of the rock in the form of what looks like tiger stripes. I believe this is on account of half occlusions being introduced by the rock surface itself, the visible surface of the rock shifts as the perspective shifts resulting in areas for which no match can be found when the block size gets large. 21 seems most balanced here.

Semi Global Block Matching

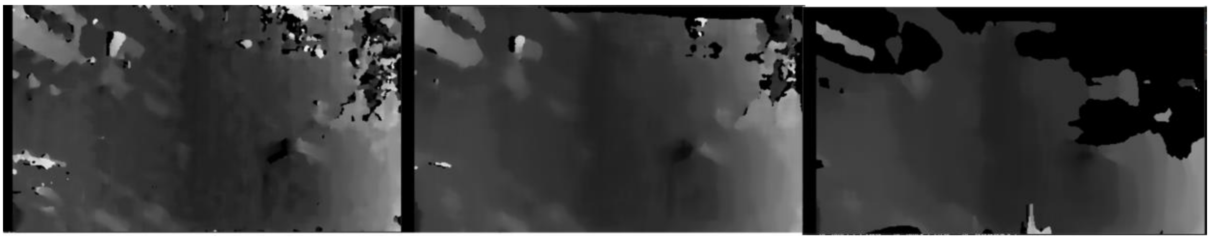
My expectation with semi-global matching is that it will not be massively different regarding max disparity and block size, fundamentally it is performing the same process but with an additional step of 1-d line optimizations across the image. So, for the sake of brevity, I will be testing this assumption by taking one “step” in each direction from the ‘optimized’ results from block matching. A step for max disparity is 16 and a step for block size is 10. Block size can be altered more granularly however change of 10 gives a clear visual result for evaluation.

Block size fixed at 31 and max disparity set at 16, 32 and 48.

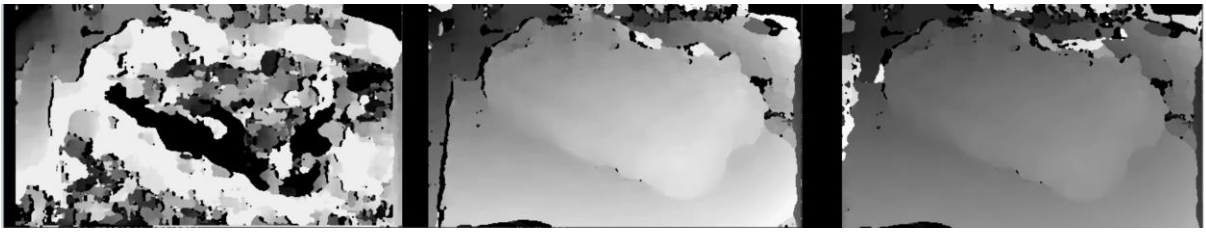
Disparity fixed at 32 and block size set to 21,31 and 41.



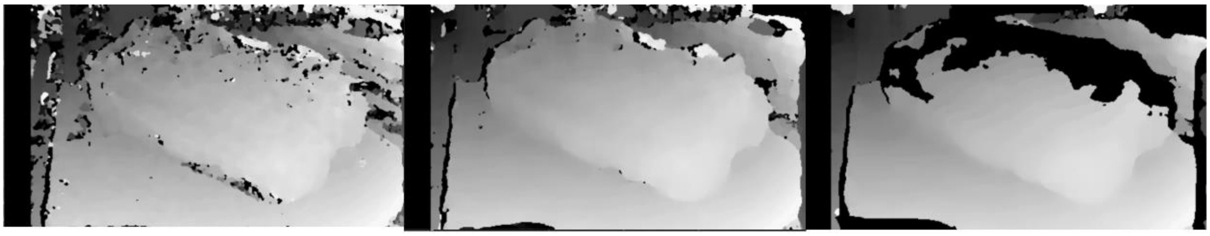
Block size fixed at 21 with max disparities 16, 32 and 48.



Disparity fixed at 16 with block sizes 11, 21 and 31.

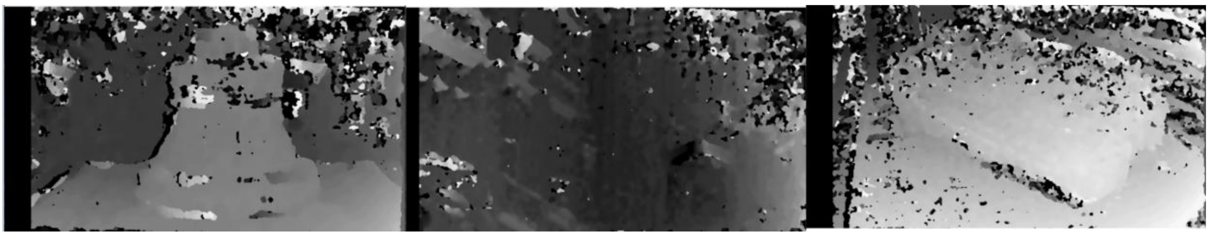


Block size fixed at 21 with max disparities 16, 32 and 48.



Disparity fixed at 32 with block sizes 11,21 and 31.

These images suggest that my original assumption was incorrect. Values that gave sensible results in block matching are giving much worse results using semi-global with the best results appearing when the block size is reduced. From these examples a lower block size is clearly preferred for semi global matching. This is because the additional optimization step across scan lines is improving the smoothness and reducing noise where in simple block matching, we were relying on large block sizes to denoise and smooth. This allows semi global matching to create better disparity maps as we are not sacrificing as much detail to achieve a result with acceptable amounts of noise. The max disparities however are similar where the visually best values for this in block matching are generally also the best for semi global block matching. This makes sense as this value relates to the actual depths present within the image, which does not change based on matching method.



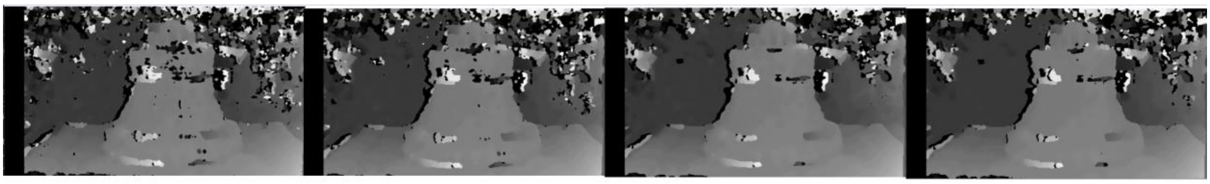
32/11 16/7 32/7

Above is one step further in block size reduction for all image sets, with maxDisparity/blocksize. This shows a significantly more accurate map especially in the case of the bell most visible by the outline of the bell being much truer to reality and with less distortion, this is due to this mapping in more granular detail. This does however clearly introduce far more noise in all cases with many more negative/zero values from matching failures showing up as black patches. The goal of the next section will be to attempt to further optimize the semi global matching parameters attempting to preserve this added detail and accuracy while denoising and smoothing as much as possible.

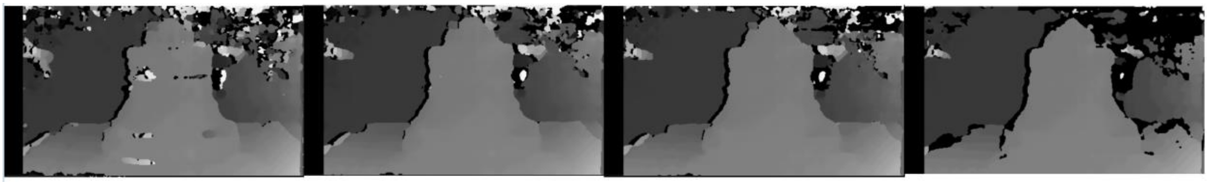
Optimization of Semi-Global Block Matching

In this section I will explore how modifications to the parameters of the semi global matching system affect the resulting disparity map. For this section I will only be examining a single image set, the bell images. In this case optimisation will be defined as generation of a visually correct map with minimized noise. The mode used will be MODE_HH.

First, I will look at how the input values P1 and P2. These values are used to control the smoothness of the map and apply penalties to disparity changes across neighbouring pixels.



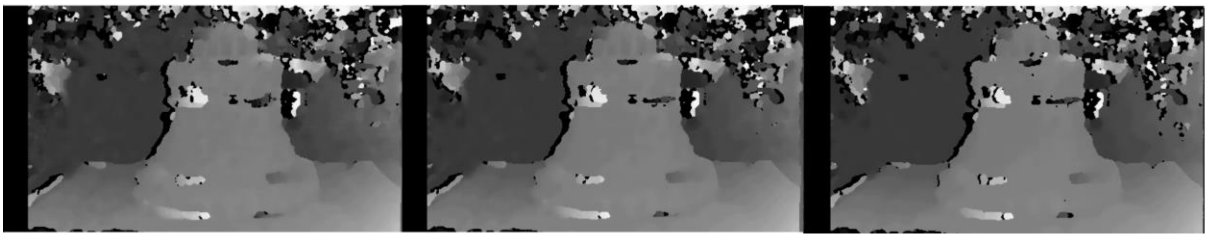
0/0 | 100/150 | 100/500 | 450/500



1000/1050 | 2000/3000 | 4000/4050 | 5000/7500

In general, the larger the values the smoother the maps are however this does not seem to be applied evenly throughout the map. As P1/P2 get in the thousands the bell itself starts to become more smoothed and lose noise at a faster rate than the background. This must be because the penalties applied by P1 and P2 are more dominant within the algorithm for closer objects with larger disparities.

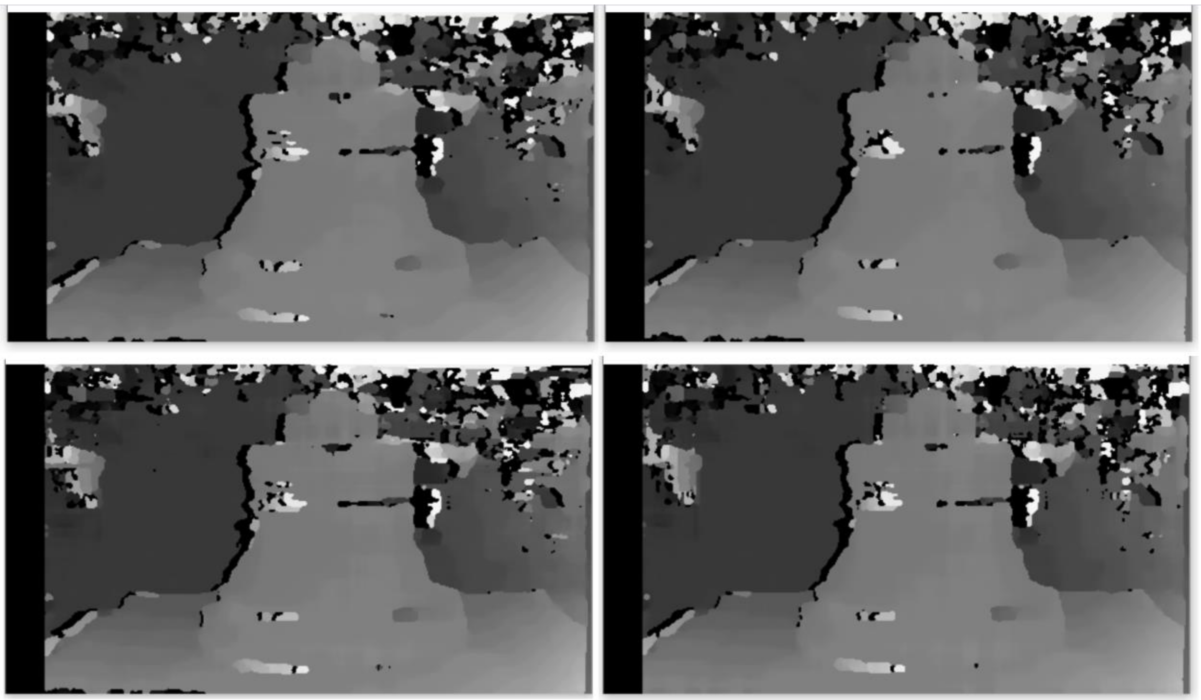
As the values get very high, we see a downside to this method, predisposing pixels to have disparities more similar to their neighbours not only smooths the correct sections but can amplify incorrect sections, this is clearest in the top right black patch that is increasingly more weighted to be larger, but it is not correct to reality. We also see a more prominent occlusion boundary around the bell as this boundary is also now weighted to be predisposed to continue being black.



1/500 | 100/500 | 450/500

It appears that the higher of the two values have a greater effect than the sum of the values, looking at the maps resulting from 100/500 and 450/500 they are very similar despite vastly different values for P1. This suggests that the difference between P1 and P2 is not as significant as the highest value, which must always be P2. This is further evidenced by introducing the extreme case, where P1 is set to 1. These results suggest that P2 is the more important value and P1 allows for finer tuning.

One of the other main options available for this method is the mode. Selecting a starting point for this is tricky, from the previous images it appears that values in the 2000-4000 range are providing the best balancing of smoothness and noise reduction to amplification of incorrect information. However, in order to see the effects of the mode most clearly selecting a noisier map will I expect yield more interesting results. For this purpose, values of 1000/1050 will be used as a hopefully good balance.



Top left: MODE_SGBM. Top right MODE_HH. Bot left: MODE_SGBM_3WAY. Bot right: MODE_HH4

The effects on the resulting map are much more subtle than I was expecting, especially considering how much more resource intensive some of these algorithms are. There are clear changes though, in the bottom right in all modes except HH4 there are varying amounts of black area with no matches found. Overall, it seems like doing some form of semi global matching is massively beneficial but beyond that doing more passes and/or more directions have diminishing returns, MODE_HH does seem to have yielded the best result by a small margin, most evident in the black streak running into the bell around its mid-point being the least pronounced of the modes.

A sample size of one is not enough to say this is true for all cases and changing the mode may well yield more significant results on different image set types. In hindsight the rough surfaced rock would likely have been more interesting as it would have had more detail and complexity to the surface we are trying to map, potentially giving the algorithm more to work with.

Final Words

From these tests a few things are clear. Compared to simple block matching semi-global block matching prefers a smaller block size as it has more granular detail to work with, but the max disparity should be approximately the same regardless of method. This makes sense as increasing the block size and performing semi-global matching are in a way doing the same thing, smoothing a reducing noise. Increased block size does this by making a pixel more like its neighbour by having more overlapping data as the block size increases and semi global matching does this by adding an additional step of 1-d optimizations across the image. If both a large block size and semi-global matching are used you have the disadvantages of the increased block size, a loss of detail and accuracy, while not taking as much advantage of semi-global as possible, resulting in a lossy map with double smoothing.

There is a relationship between what the max disparity should be set to the actual depth within a scene, if the object you wish to map is far away then the distance between that object in the images will be small, and we specify the max disparity as an expectation of what this distance will be.

Rough surfaces are more prone to changes in angle due to many tiny partial occlusions within their surface, this causes both small and large block sizes to cause issues, small causes many wrong matches causing noise and large causes many failed matches causing large black voids.

Limitations

Three image sets are not enough to draw any truly definitive conclusions, they do encompass a wide variety of cases; close smooth object, close rough object and distant scene with less relative difference in disparities. I believe the case of attempting to map multiple objects at different depths would have been interesting and potentially yielded quite a different result.

There are also other aspects that would likely have been worth exploring, the time taken for semi global matching would likely be interesting allowing for a comparison between the methods and the relative gain in fidelity. This would go hand in hand with a larger sample size for testing the fine-tuning parameters of semi-global matching. Further examination of the effects of various scale factors on the images used would likely also have been interesting.