

3D Reconstruction by Stereo Imaging

Stereo vision is the process of extracting 3D depth information from multiple 2D images. Conventionally, two horizontally separated cameras are used to obtain two different perspectives on a scene. Because the cameras are separated, each feature in the scene appears at a different coordinate in both images. This difference between these coordinates is called the disparity and the depth of each point in the scene can be computed from its disparity. The difficult part is computing the disparity at each point.

Algorithms for computing features between images are generally complex, memory inefficient and require random access to large portions of memory. The state of the art Stereo Algorithm is based on *Semi Global Matching* (SGM), a global correspondence optimization. This algorithm performs very well in practice but is extremely memory inefficient.

However, SGM can be modified with some heuristics such that the amount of temporary memory it uses is proportional to the number of pixels in the image. This improved memory complexity makes *efficient-SGM* (eSGM) ideal for implementation on FPGAs.

Writing a complete stereo pipeline requires many diverse components. A full stereo pipeline based on eSGM requires rectified camera images, and rendering the output in point clouds and depth maps. Furthermore we will need to design a memory architecture that can maintain a high throughput and utilize our computation resources effectively.