3D RECONSTRUCTION MODEL

3D stereo reconstruction is based on the concepts of epipolar geometry, shown in Fig.1. 
P(x,y,z) is a point in 3D space, and the two points denoted E_L and E_R represent the epipoles. An epipole is the point of intersection of the line across the optical centers, i.e. the baseline, with the image plane. The points P, O_L and O_R form a plane called the epipolar plane. The line P-O_L is seen by the left camera as a point X_L because it is directly in line with the camera’s center of projection O_L. In the right image plane, that line is represented as E_R - X_R and is called the epipolar line. For each point observed in one image, the same point must be observed in the corresponding epipolar line on the other image. This is known as the epipolar constraint, which reduces the correspondence problem to a search along conjugate epipolar lines [14].

Given pixel coordinates of P_L(x_L,y_L) in the left image, and the corresponding pixel coordinates P_R(x_R,y_R) in the right image, the 3D world coordinates P(X,Y,Z) are computed as:

\[ x = \frac{x_L}{f}, \quad y = \frac{y_L}{f}, \quad Z = \frac{b}{f} \]  

where \( f \) is the focal length and \( b \) is the baseline (shown in Fig. 2.). Disparity represents the relative difference in the location between common objects (i.e. the location of the object in the right image, relative to the location of that object in the left image) on the stereo pair. The depth map is simply the reciprocal of the disparity map.

If we assume that a set of stereo images come from calibrated cameras (with known focal length, and fixed baseline distance), the first step in 3D reconstruction is the rectification of the stereo images. Rectification is the process that transforms each image into a common image plane, aligning the pairs of conjugate epipolar lines to a common image axis (usually the horizontal). Rectification allows the computation of stereo correspondence, which involves a 2D search in unrectified images, to be reduced to 1-D search along the epipolar line [1].

Rectification produces a transformation function used to project one of two images in the other’s common plane; the transformation function can remain the same if the stereo camera system remains calibrated, otherwise, a new transformation function is necessary [1]. In this work, we assume a steady and calibrated camera system, so we do not focus on the rectification; instead we take advantage of modern FPGA technology that includes embedded processors, utilizing an embedded processor to transform both images on a common plane.

An overview of a stereo 3D reconstruction system is shown in Fig. 3. The algorithm receives a pair of stereo images (left and right image) as an input, and outputs a disparity map (or the depth map). The 3D reconstruction algorithm must solve two basic problems: correspondence, which deals with finding an object in the left image that corresponds to an object in the right image, and reconstruction, which deals with finding the depth (i.e. the distance from the cameras which capture the stereo images) and structure of the corresponding point of interest. The correspondence problem is the most demanding in terms of computational complexity, and involves searching and matching techniques (to locate a common object in both images), the robustness of which determines quality and precision of reconstructed 3D data [1][14].

In order to locate a common point of reference in the two images, a small window (called correlation window) from both images is being evaluated by comparing the window from the left image to the window from the right image, via the sum of absolute differences (SAD) method [1]. The search is constrained along a horizontal scan line, as the images are rectified. Additionally, an object that appears in both stereo images will be found within a maximum horizontal bound, which depends on the object’s distance from the camera. Henceforth, a search limit can be imposed, known as disparity range, which constrains the search along a bounded horizontal scan line. The size of the search window, and the disparity range, impact...
the reconstruction algorithm significantly, both in terms of performance, as well as quality of results.

While rectification reduces the search space, further improvement can be obtained, by applying an edge detection process over the input images. Edge detection detects locations in the image where intensities change over a certain threshold through the x and y directions, indicating the presence of an edge [1]. These locations are described by the edge points, which determine the outline (i.e. perimeter) of an object, and distinguish it from the background and other objects [1][13]. Consequently, the matching procedure can concentrate only on the edge points of the original images, reducing thus the processing time. Moreover, in hardware implementations, edge images that contain only the edge points, can be represented in binary form (i.e. black or white) and encoded using only one bit per pixel, reducing drastically the computation space and complexity of the search and match operations.

When an object is located in both images, its pixel coordinates are then used to derive the disparity map, and by using triangulation, the 3D structure of the scene. Triangulation is dependent on the expected view of the object under different angles, and includes a projection of the depth map on a 3D space; this is however dependent on the host application and specific lighting conditions, and is left as part of future work [1].

![Figure 1. The epipolar geometry.](image1)

**Disparity between left and right tower.** \( X_R \) indicates the pixel coordinate of the object on the right image, and \( X_L \) the pixel coordinate in the left image.

\[
\text{Disparity} = | X_R - X_L |
\]

![Figure 2. Disparity Computation.](image2)

![Figure 3. Stereo 3D Reconstruction Process.](image3)