We first present a novel criterion for Next-Best View (NBV) optimisation with varying numbers of stereo pairs. The final column shows the reference. The 3D reconstruction of scenes and objects from 2D images is an explanation. As shown in Figure 2 γ is a parameter that can encourage or discourage exploration. As shown in Figure 2  γ of 1 will give the highest cost to unobserved voxels, preferring to reduce the uncertainty of observed voxels, while 0 will give them the lowest, preferring exploratory behaviour.

![Figure 1: Results for Middlebury Dino (top) and Temple (bottom) Datasets, with varying numbers of stereo pairs. The final column shows the reference.](image1.png)

1 Next-Best View Optimisation

We first present a novel criterion for Next-Best View (NBV) optimisation based on a compromise between the competing objectives of coverage and accuracy. The coverage objective will drive the system to collect views of previously unobserved parts of the scene (e.g., due to restrictions on the field of view or occlusion), whereas the accuracy objective will drive the system to choose the next pose to reduce the point cloud’s uncertainty. These two criteria are optimised jointly, making use of an octree structure and a dense point cloud. The octree allows for quick and efficient calculations on scene geometry, while the dense cloud (and covariances) allow for more detailed calculations about scene noise and and viewing angle.

The NBV is calculated as follows: Given a Configuration Space (CS) of sensor poses, the cost of each pose can be estimated by casting a set $S_r$ of random rays from the camera centre through the image plane. Each ray will reach its end either after hitting an occupied ($V_o$) or unobserved ($V_u$) voxel, ignoring empty ($V_e$) voxels. When a ray $r \in S_r$ intersects with an occupied voxel $v \in V_o$, we can estimate a cost for each point $p \in P_r$ as $\psi(r, p) = e^{-|\lambda_p \cdot r|}$, where $\lambda_p$ and $e_p$ are the largest eigenvalue and eigenvector, respectively, of the covariance $\Sigma_p$. Consequently, the cost of a voxel is defined as the average point cost

$$\psi(r, v) = \frac{1}{|P_r|} \sum_{p \in P_r} \psi(r, p).$$

The NBV cost of a particular pose $x$ is defined as

$$C_x = \frac{1}{|S_r|} \sum_{r \in S_r} \begin{cases} \psi(r, v) & \text{if } v \in V_o \\ \gamma & \text{else } v \in V_u \end{cases} \quad \text{if } \gamma \in [0, 1].$$

In this equation, $\gamma$ is a parameter that can encourage or discourage exploration. As shown in Figure 2  γ of 1 will give the highest cost to unobserved voxels, preferring to reduce the uncertainty of observed voxels, while 0 will give them the lowest, preferring exploratory behaviour.

![Figure 2: Average Error (Left) and Average Coverage (Right) with different values of γ.](image2.png)

2 Next-Best Stereo Optimisation

When there are multiple collaborating sensors available, we can extend NBV to also optimise the stereo arrangement of the sensors. This can be achieved by selecting another view with respect to the NBV, to create the best possible stereo pair. Actively selecting stereo pairs allows sensors to be positioned to allow an optimal vergence and baseline, respective to the observed parts of the scene.

This implies several requirements: First, the baseline of the cameras must be scaled, depending on the distance to the observed geometry, and the vergence angle should be minimised to allow the dense matching to be performed with the least amount of error possible ($C_T, C_v$). Second, we must ensure robustness against rotation in the image ($C_R$). Finally, the distance between the vergence point and the nearest geometry should be minimised, to ensure that the sensors are trained on actual scene geometry and not empty space ($C_I$). We then find the pose that minimises

$$C = C_b + C_T + C_R + C_I.$$