Figure 1. Visualization of produced co-planar points in MegaDepth Dataset. The red point is one anchor point \( p_m \), while the green dots are the co-planar pixels among \( K \) sampled candidates \( \{ q_{mn} | 1 \leq n \leq K \} \), computed following Eqn. 1.

Production of the indicator matrix \( O^+ \). In the main paper Eqn. 14, we utilize an indicator matrix \( O^+ \) to indicate the co-planar pairs between anchor and candidate points. Given the \( K \) anchor points \( \{ p_m | 1 \leq m \leq K \} \) and \( K \times K \) candidate points \( \{ q_{mn} | 1 \leq n \leq K \} \), the indicator matrix \( O^+ \) of size \( K \times K \) is computed as:

\[
O^+_{m,n} = 1 \quad \text{if} \quad \begin{cases} 1 - \arccos (\mathbf{n}_{p_m}^T \mathbf{n}_{q_{mn}}) < k_1 \\ \text{dist}(\mathbf{n}_{p_m}, \mathbf{p}_m, \mathbf{d}_{p_m}, \mathbf{d}_{q_{mn}}, \mathbf{K}_1) < k_2 \\ \| \text{proj}(\mathbf{H}_{p_m}^T \mathbf{p}_m, \mathbf{q}_{mn}) - \mathbf{q}_{mn} \|_2 < k_3. \end{cases}
\]

The function \( \text{dist}(\cdot) \) computes the point-to-plane distance in 3D space. The plane is spanned by the norm \( \mathbf{n}_{p_m} \) and the re-projected 3D point at \( p_m \). The function \( \text{proj}(\cdot) \) indicates the planar projection under pixel \( p_m \) homography matrix \( \mathbf{H}_{p_m} \) (see main paper equation Eqn. 13). The \( k_1, k_2 \) and \( k_3 \) are set to 0.002, 0.02, and 1 respectively. We visualize the produced groundtruth co-planar points in Fig. 1.

Additional Visualization. We include additional visualization of the paired MIM pretexting task (Fig. 2) and reconstruction visualization (Fig. 4 and Fig. 3).

References

Figure 3. Visual Quality of the Reconstruction on MegaDepth [1].
Figure 4. Visual Quality of the Reconstruction on ScanNet.