Figure 3. Scores for different number of points (left and right) with different distributions (top and bottom) in the image for $L = 3$.

late from panoramic image pairs to avoid degenerate points and thereby improve robustness of triangulation and subsequent image registrations.

4.2. Next Best View Selection

Next best view planning has been studied in the fields of computer vision, photogrammetry, and robotics [12]. Choosing the next best view in robust SfM aims to minimize the reconstruction error [17, 24]. Here, we propose an efficient next best view strategy following an uncertainty-driven approach that maximizes reconstruction robustness.

Choosing the next best view is critical, as every decision impacts the remaining reconstruction. A single bad decision may lead to a cascade of camera mis-registrations and faulty triangulations. In addition, choosing the next best view greatly impacts both the quality of pose estimates and the completeness and accuracy of triangulation. An accurate pose estimate is essential for robust SfM, as point triangulations may fail if the pose is inaccurate. The decision about choosing the next best view is challenging, since for Internet photo collections there is usually no a priori information about scene coverage and camera parameters, and therefore the decision is based entirely on information derived from appearance [17], two-view correspondences, and the incrementally reconstructed scene [53, 24].

A popular strategy is to choose the image that sees most triangulated points [52] with the aim of minimizing the uncertainty in camera resection. Haner *et al.* [24] propose an uncertainty-driven approach that minimizes the reconstruction error. Usually, the camera that sees the largest number of triangulated points is chosen, except when the configuration of observations is not well-conditioned. To this end, Lepetit *et al.* [34] experimentally show that the accuracy of the camera pose using PnP depends on the number of observations and their distribution in the image. For Internet photos, the standard PnP problem is extended to the estimation of intrinsic parameters in the case of missing or inaccurate prior calibration. A large number of 2D-3D correspondences provides this estimation with redundancy [34], while a uniform distribution of points avoids bad configurations and enables reliable estimation of intrinsics [41].

The candidates for the next best view are not the yet registered images that see at least $N_t > 0$ triangulated points. Keeping track of this statistic can be efficiently implemented using a graph of feature tracks. For Internet datasets, this graph can be very dense, since many images may see the same structure. Hence, there are many candidate views to choose from at each step in the reconstruction. Exhaustive covariance propagation as proposed by Haner *et al.* is not feasible, as the covariance would need to be computed and analyzed for each candidate at each step. Our proposed method approximates their uncertainty-driven approach using an efficient multi-resolution analysis.

We must simultaneously keep track of the number of visible points and their distribution in each candidate image. More visible points and a more uniform distribution of these points should result in a higher score $\mathcal{S}$ [31], such that images with a better-conditioned configuration of visible points are registered first. To achieve this goal, we discretize the image into a fixed-size grid with K_l bins in both dimensions. Each cell takes two different states: *empty* and *full*. Whenever a point within an *empty* cell becomes visible during the reconstruction, the cell's state changes to *full* and the score $\mathcal{S}_i$ of the image is increased by a weight w_l. With this scheme, we quantify the number of visible points. Since cells only contribute to the overall score once, we favor a more uniform distribution over the case when the points are clustered in one part of the image (*i.e.* only a few cells contain all visible points). However, if the number of visible points is $N_t \ll K_l^2$, this scheme may not capture the distribution of points well as every point is likely to fall into a separate cell. Consequently, we extend the previously described approach to a multi-resolution pyramid with $l = 1...L$ levels by partitioning the image using higher resolutions $K_l = 2^l$ at each successive level. The score is accumulated over all levels with a resolution-dependent weight $w_l = K_l^2$. This data structure and its score can be efficiently updated online. Fig. 3 shows scores for different configurations, and Sec. 5 demonstrates improved reconstruction robustness and accuracy using this strategy.

4.3. Robust and Efficient Triangulation

Especially for sparsely matched image collections, exploiting transitive correspondences boosts triangulation completeness and accuracy, and hence improves subsequent image registrations. Approximate matching techniques usually favor image pairs similar in appearance, and as a result two-view correspondences often stem from image pairs with a small baseline. Leveraging transitivity establishes correspondences between images with larger baselines and thus enables more accurate triangulation. Hence, we form feature tracks by concatenating two-view correspondences.

A variety of approaches have been proposed for multi-view triangulation from noisy image observations [27, 40, 5]. While some of the proposed methods are robust to a certain degree of outlier contamination [25, 35, 3, 44, 32],