Next best view planning for active model improvement

Enrique Dunn
dunn@cs.unc.edu
Jan-Michael Frahm
jmf@cs.unc.edu

Department of Computer Science
University of North Carolina
Chapel Hill, USA

Abstract

We propose a novel approach to determining the Next Best View (NBV) for the task of efficiently building highly accurate 3D models from images. Our proposed method deploys a hierarchical uncertainty driven model refinement process designed to select vantage viewpoints based on the model's covariance structure and appearance, as well as the camera characteristics. The developed NBV planning system incrementally builds a sensing strategy by sequentially finding the single camera placement, which best reduces an existing model's 3D uncertainty. The generic nature of our system's design and internal data representation makes it well suited to be applied to a wide variety of 3D modeling algorithms. It can be used within active computer vision systems as well as for optimized view selection from the set of available views. Experimental results are presented to illustrate the effectiveness and versatility of our approach.

1 Introduction

One of the major areas of computer vision is the reconstruction of dense geometry from video. There now exists a variety of stereo methods to estimate the dense geometry [3],[21]. Recently, performance improvements lead to the ability of efficient reconstruction from large video data or photo-collections [16],[2]. In contrast to the traditional stereo approach, for these applications we face a view selection problem to deploy the most beneficial views for the multi-view stereo process to achieve the highest possible accuracy and with the smallest number of views. There are many instances where the input data or images are either insufficient for the algorithm to meet the specified coverage or accuracy requirements or are in fact redundant and inefficient to process, even sometimes leading to performance degradation. Sensor planning systems strive to determine the pose and settings of a vision sensor to undertake a vision task usually requiring several views [1]. The next best view problem (NBV) seeks a single additional sensor placement in order to improve an existing scene reconstruction derived from the current imaging configuration. Depending on the context, NBV can be seen as an incremental approach to building a camera network, designing a sensing strategy for autonomous exploration or selecting an input image dataset for multiview reconstruction.

Active 3D model improvement is considered in this work to be the sequential process of systematically increasing the precision and completeness of the estimated 3D model. Our

proposed NBV planning approach uses adaptive planar patches as the basic element of structure representation and the covariance matrices of these patches as the representation for the 3D reconstruction uncertainty. We combine these elements along with the object's texture properties to propose a novel viewpoint selection criterion seeking a balance between the reduction of geometric uncertainty and the attainment of reliable image measurements. We procure such balance by developing our criterion around three inherent object and camera properties involved in optical 3D reconstruction: 1) structure estimation uncertainty, 2) the projection properties of the object in the current view and 3) the surface texture appearance. We strive to develop a general NBV planner by decoupling the internal data representations used for our viewpoint discrimination criteria from the representations used in the deployed 3D reconstruction algorithm. Moreover, the developed approach is aimed at dense 3D reconstructions, which commonly output millions of surface points even for simple scenes. Accordingly, scalability and efficiency are major concerns when developing a viewpoint selection algorithm in this context. We propose a data parallel hierarchical approach that can efficiently deploy commodity parallel architectures like GPUs or multi-core processors.

The remainder of the paper is organized as follows. First, in Section 2 we review related work. Next, we present a system overview in Section 3. The model update process and adaptive structure representation utilized by our approach are described in Section 4. This is followed by the development of our camera selection criterion in Section 5. Afterwards, Section 6 includes experimental results.

2 Related work

The challenge of automatic viewpoint selection has been widely studied in robotics, computer vision and photogrammetry. Surveys that span from early approaches in this field to recent advances, were published by Newman *et al*. [12], Tarabanis *et al*. [17] and Scott *et al*. [15]. Recently, Chen *et al*. [1] provided a broad coverage of multiple research areas within sensor planning. We restrict our discussion to a subset of approaches, which are closely related to our proposed active 3D model improvement task.

The task of designing a viewing configuration for precise 3D reconstruction is known in photogrammetry as the photogrammetric network design (PND) problem. Fraser [2] early on identified the analytical difficulties of designing an optimal imaging geometry in the context of rigorous photogrammetric 3D measurements. His work identified the high non-linearity and multi-modality, which makes the PND problem ill-suited for canonical optimization methods. Mason [10] adopted an expert systems approach based on generic networks to achieve strong viewing configurations for model based PND. The developed system used a CAD model as input and followed a series of predetermined rules for each CAD element in order to design an imaging geometry. Olague & Mohr [13] addressed the PND problem by developing a criterion based on forward covariance propagation of image measurement uncertainty. The goal is to minimize the maximum element along the diagonal of the reconstruction's covariance matrix. To this end, the authors used a genetic algorithm to search for an optimum in the space of possible multi-camera viewing configurations. It is important to note that the aforementioned PND systems were designed to generate sets of multiple viewpoints used for highly precise 3D reconstruction tasks carried out in well controlled and customized environments (i.e. fiducial markers, high accuracy calibration patterns, etc.). Moreover, they mainly address the geometric aspects of 3D reconstruction omitting considerations on the role of texture saliency in the image measurement process.

Robot vision researchers have studied the sensor placement in which a controlled camera can be best used to achieve accurate 3D reconstructions. Whaite and Ferrie [20] developed a 3D modeling system that used parametric modeling of scene elements and used the internal model uncertainty to determine sensing actions. Marchand and Chaumette [9] developed a system for structured scene reconstruction by developing optimal strategies for surveying a set of volumetric primitives. We note that, while these systems successfully achieved autonomous operation, they used simple parametric models to represent scene elements, whereas our approach does not place any restrictions on the observed geometry and additionally considers the appearance of the object's surface. Accordingly, our approach is better suited to perform over a large variety of scenes.

In the computer vision community camera placement and configuration has recently received renewed interest [13],[14]. Wenhardt *et al*. [19] proposed a 3D reconstruction based on a probabilistic state estimation framework and the next best view is determined by a metric of the state estimation's uncertainty. The authors propose the use of three different metrics, which correspond to the concepts of D-, E- and T-Optimality found in the optimal experimental design literature. Results indicate variability of the attained configurations with respect to the optimality criterion proposed. Hornung *et al*. [6] propose an image selection scheme for multi-view stereo, which selects images in order to improve the coverage of a voxel based proxy. Their approach strives to achieve sufficient sampling of the entire object's surface while identifying regions with poor photo-consistency for additional redundant sampling. While these approaches address pertinent aspects of the 3D reconstruction process they are strongly coupled with the reconstruction algorithm being deployed. Our proposal strives for algorithm independence by decoupling the internal representation used for planning from that of the reconstruction algorithm. Accordingly, our approach is better suited to be integrated to a larger variety of reconstruction approaches.

3 System overview and notation

The operation of our NBV planner can be described as an iterative process consisting of the following sequential stages:

1. A partial input 3D model with uncertainty information is transformed into a 3D mesh.
2. Adaptive planar patches are extracted from the 3D mesh.
3. Candidate viewpoints are evaluated using patch 3D uncertainty and surface texture.
4. Image acquisition is performed according to the NBV selection.
5. An independent 3D reconstruction algorithm is deployed.
6. 3D mesh structure and uncertainty are updated after each sensing action.
7. Repeat until target accuracy is reached.

We represent scene structure by an adaptive 3D triangular mesh M_i locally approximated through a planar patch considering the current reconstruction uncertainties. Hence, our model can represent general scene geometry by using patches as small as a planar surface of the size of a pixel at the scene distance, while efficiently representing larger planes through a single model patch. Each patch is parameterized by

$$P_i = [\mathbf{X}_i, \mathbf{n}_i, \Sigma_i, \mathbf{S}_{ij}] : \{\mathbf{X}_i, \mathbf{n}_i, \in \mathbf{R}^3, \Sigma_i \in \mathbf{R}^{3\times 3}, \mathbf{S}_{ij} \in \mathbf{R}^{p\times p}\}, \tag{1}$$

where $\mathbf{X}_i$ is the 3D position of the patch, $\mathbf{n}_i$ is the patch normal vector, Σ_i is the 3D covariance matrix and $\mathbf{S}_{ij}$ is the square set of $p \times p$ (p is a user defined integer value) neighboring

image pixels to the projection of P_i onto a reference image j. Viewpoint configurations are parameterized in terms of sensor position and orientation angles,

$$v_j = [\mathbf{x}_j, \theta_j] : \{\mathbf{x}_j \in \mathbf{R}^3, \theta_j \in SO(3)\}, \tag{2}$$

where $SO(3)$ is the special orthogonal group in three dimensions. The geometric structure of the 3D uncertainty of the input model is captured by the eigenvectors and the eigenvalues of the patch covariance matrix Σ. Namely, in Euclidian 3D space the eigenvectors $\mathbf{e}_k | k = 1 \dots 3$ convey the orientation of the 3D uncertainty, while the eigenvalues λ_k specify the magnitude in each direction. We define Ψ to be the matrix of eigenvectors scaled by their corresponding eigenvalue,

$$\Psi = [\lambda_1 \mathbf{e}_1 \; \lambda_2 \mathbf{e}_2 \; \lambda_3 \mathbf{e}_3]. \tag{3}$$

The planning approach presented here jointly uses the information contained in Ψ and the surface texture saliency as the guide for viewpoint selection.

4 Mesh update and 3D patch adaptation

Our triangular mesh representation can be readily extracted from typical 3D outputs such as depth maps, volumetric grids or point clouds. We assume mesh vertices to represent actual scene measurements and rely on the availability of vertex wise 3D uncertainty estimates for the input 3D mesh. Our NBV planner is independent from the procedure by which vertex-wise covariance estimates are obtained. In this work, we update covariance estimates for each patch with an extended Kalman filter based on triangulation [5] and the viewing geometry used during modeling. A recursive estimation approach is adopted in order to achieve scalable performance with respect to the number of input images. Covariance initialization for each vertex is obtained by estimating the 3D uncertainty of vertex triangulation of the first two viewpoints from which the vertex was sampled. Image measurements are assumed to be perturbed with Gaussian noise. Subsequent individual measurements of the same vertex are processed under an EKF framework in order to update structure and covariance estimates.

Image capture is to be performed at the selected NBV (see details of the selection in Section 5) and the acquired data is processed by an independent reconstruction algorithm. Once the reconstruction algorithm has been deployed, we obtain a new 3D model M_i^+, which must be reconciled with our current structure representation M_i. Regions of mesh intersection are identified efficiently on the GPU through detection of image based overlap of the existing and the novel mesh. The correspondence between vertices in the overlapping regions is performed by the 3D Mahalanobis distance proximity defined by the existing vertex covariance matrix Σ. Once all vertices in the intersection region of the novel mesh are associated, their positions and uncertainties (e.g. $\mathbf{X}$ and Σ) are fused using an EKF framework, which takes into account the displacement among vertices and the camera parameters used for sensing. Regions of the novel mesh M_i^+ not previously reconstructed are appended directly to the updated triangle mesh and their uncertainties are estimated incorporating all cameras observing each triangle.

After the uncertainty estimation, the 3D mesh is transformed into a set of 3D patches, which are deployed to determine the NBV. For each vertex we define a circular plane parallel to the best fit plane of the neighboring set of vertices and a radius equal to their average distance. Hence, our approach provides a planar proxy of the surface at each vertex, which is augmented with a 3D uncertainty estimate used in the evaluation of our NBV criterion.

Neighboring patches are merged based on their uncertainties and their plane parameters. A sphere is defined around a given *seed* patch with a radius $r = 3\sqrt{\lambda_1}$ proportional to the largest eigenvalue of the covariance matrix corresponding to the patch. All patches whose center lies within the sphere are merged into a single aggregate patch. The patch parameters (2) assigned to the merged patch are those of the *seed* patch. Accordingly, the extent of a merged patch approximates the 3D reconstruction uncertainty of the surface element it represents. The process of merging 3D patches is performed in a greedy fashion. Once the uncertainty of a patch is reduced through additional sensing actions, the geometry is represented on a finer level by multiple disjoint patches. The partition is performed by disassociating merged patches and applying the merging procedure with the updated mesh uncertainty and position estimates. Using adaptive patches, the current surface topology and uncertainty are represented by a dynamic structure that evolves during the sequential reconstruction process. Moreover, it is possible to manipulate the process of patch merging and partitioning in order to accommodate different computational requirements by limiting the number of patches considered by the NBV planner.

5 A novel next best view criterion

A novel viewpoint sought by our planner should achieve a balance between the reduction of geometric uncertainty and the attainment of reliable image measurements. We leverage the 3D uncertainty information contained in the scaled eigenvector matrix Ψ (3) obtained from each patch positional covariance structure as well as the patch projected image texture. Instead of optimizing an experimental design criterion as in [19], we define a criterion based on three reconstruction goals, which contribute to systematically reducing the uncertainty of the estimated 3D patch. The first reconstruction goal addressed is to achieve an adequate incidence with respect to a patch 3D uncertainty. The second reconstruction goal under consideration is to obtain a favorable imaging resolution for the projected texture of a given 3D patch. The third goal incorporated into our criterion is to condition the relevance of imaging resolution on the texture of the observed surface.

5.1 Incorporating the uncertainty of 3D reconstruction estimates

Let $\mathbf{X}_i$ denote the estimated 3D position of a primitive P_i. The goal is to find the viewpoint v_j with camera center $\mathbf{x}_j$ such that the unit length viewing direction

$$\mathbf{v} = \frac{\mathbf{X}_i - \mathbf{x}_j}{\|\mathbf{X}_i - \mathbf{x}_j\|_2} \tag{4}$$

best reduces the 3D uncertainty contained in Ψ_i. Given an estimate of a 3D point with non isotropic 3D uncertainty, the viewing rays $\mathbf{v}$ for minimizing triangulation uncertainty are the ones orthogonal to the main uncertainty direction vector. Accordingly, for 3D estimates where the majority of the uncertainty is found along a single direction $\mathbf{e}_1$ (e.g. the eigenvector with the largest associated eigenvalue), a viewing ray orthogonal to this vector corresponds to a solution of the product equation $\mathbf{v}^T\mathbf{e}_1 = 0$. However, a more general criterion is desired for arbitrary covariance structures. We propose to find the viewing ray, which minimizes

$$f(P, v) = \|\mathbf{v}^T \left[\lambda_1\mathbf{e}_1^i \; \lambda_2\mathbf{e}_2^i \; \lambda_3\mathbf{e}_3^i\right] \|_2 \; = \; \|\mathbf{v}^T\Psi_i\|_2. \tag{5}$$

The minimization of (5) considers 3D reconstruction as a merely geometric task, not taking into account practical aspects such as robustness of image measurements and matching saliency. We incorporate these aspects into our approach by also considering the effects of varying a viewpoint's incidence and proximity with respect to a given 3D primitive.

5.2 Incorporating camera projection properties

3D reconstruction estimates the position of the points located on a 3D supporting surface. It is typically the visual appearance of this supporting surface, which allows the identification and measurement of the projection of a given set of 3D points in multiple image planes. Better accuracy in image measurements can be obtained as the imaging resolution increases. Given knowledge of the camera's intrinsic parameters, the main factors in determining a surface's projection are the viewing angle and the distance from a given 3D surface. We combine both incidence and proximity by measuring a single quantity: the area of projection of a 3D surface onto the image plane. It is straightforward to compute this quantity analytically for simple geometric primitives like our planar patches. We measure the projected area with high efficiency in a GPU by using rendering. The benefits of using a GPU computation are that resolution, field of view and occlusions are handled by the graphics engine.

At this point we have defined geometric relationships, which favor the suitable observation of a generic surface. The motivation behind such definitions is to attain reliable image measurements. The next subsection incorporates texture saliency to model the ability to make reliable image measurements.

5.3 Incorporating surface texture appearance

Visual saliency of a scene surface is a requirement for accurate matching across images in feature based reconstruction. Accordingly, textureless scene regions present a major difficulty in the application of these algorithms. On the other hand, contour based approaches do not rely on texture saliency to estimate bounding volumes, but instead favor tangent views of the scene surface. In our approach, we strive for oblique views only for textureless regions. Note that the motivation behind measuring the projected area of a 3D primitive is to consider jointly a viewpoint's incidence and proximity. Taking into account that the projected area of a perpendicularly observed planar surface is null, the relevance of the projected area of a primitive is conditioned on its texture. This relevance factor is modeled by a continuous step function with transition at a given texture threshold. We propose to use a modification of the *Gauss error function* (encountered by integrating the normal distribution) of the form

$$erf(x) = \frac{1}{\gamma\pi}\int_0^x e^{\frac{(t-\tau)^2}{\gamma}}dt + \frac{1}{2} \tag{6}$$

where x is the texture measure estimated for a given primitive, τ is the texture threshold value and γ is a decay factor controlling the slope of the transition in the step function. The error function may be substituted by a logistic function. Using Eq. (6) we can obtain a value in the range $[0,1]$ to describe the relevance of the projected area of a given primitive. The measure used to quantify texture saliency is described next.

Let $\mathbf{S}_i$ denote the image region corresponding to the projection of patch P_i in a reference view. We describe texture saliency by measuring the entropy of the autocorrelation function $A(\mathbf{S}_i)$ for patch P_i. This is motivated by the fact that homogeneous texture regions will display "flat" profiles for $A(\mathbf{S}_i)$, leading to high entropy. On the other hand, surfaces with

salient texture will provide a well localized "peak" on the autocorrelation function landscape, leading to low entropy. For an image region of size $p \times p$, the $A(\mathbf{S}_i)$ will output a $2p-1 \times 2p-1$ matrix with values a_i in the range $[-1,1]$. The matrix values are normalized and used to evaluate Shannon entropy. We empirically determined a texture threshold value τ for our step function (6), as well as the decay value γ. Hence, we obtain a function $h(\cdot)$ of the form

$$h(P_i) = 1 - erf\left(-\sum_{a_i \in A(\mathbf{S}_i)} p(a_i)\log p(a_i)\right). \tag{7}$$

The proposed function (7) measures the quality of a correlation match given the local appearance. It does not include currently any correction for the texture frequency reduction due to potential projective distortions. This can easily be integrated into the computation of the auto-correlation. We found that in practice it does not change the results significantly, but the simpler measurement (7) is more efficient to compute since it is a constant for a given model and does not depend on the location of the new camera.

5.4 The aggregate criterion

In developing our criterion we seek to attain a trade-off between two (typically conflicting) objectives involved in depth estimation. These objective are: 1) maximizing the visibility of a given patch on the novel image, 2) aligning the camera viewing direction to the direction of smallest uncertainty for the considered 3D primitive P_i. We propose the following function to evaluate the contribution of a viewpoint v for a single 3D primitive P:

$$C(v,P) = \frac{g(v,P)^{h(P)}}{f(v,P)} \tag{8}$$

where $g(v,P)$ denotes the computed projection area of the 3D primitive, while $h(P)$ and $f(v,P)$ are defined in Eqs. (7) and (5) respectively. The function (8) is evaluated for each patch P_i and combined through a weighted sum to define our NBV criterion

$$F(v) = \sum_{i=0}^{N} w_i C(v,P_i). \tag{9}$$

We define the weight $w_i = \lambda_1^i$ of a patch to be equal to the largest eigenvalue associated with its covariance matrix. Accordingly, patches with larger uncertainty are given more attention in the viewpoint search process. It is important to note that the weight value could alternatively be defined in terms of more specific experimental design criteria (as presented in [19]) in order to favor the reduction of a particular uncertainty metric.

5.5 Criterion optimization

The function defined by our criterion (9) is a multimodal search space for selecting the NBV. The main obstacle for attaining an analytical (or closed form) solution of our criterion optimization problem consists on modeling object self occlusions. For our online planning module, we seek to balance the number of function evaluations utilized to find the NBV and the quality of the obtained solution. Robust planning behavior was observed by using the Nelder-Mead simplex algorithm for continuous optimization [11]. The algorithm belongs

to the general class of direct search methods [8], since it uses only function values at some points in the search space and does not try to form an approximate gradient at any of these points. The algorithm operation defines a simplex of $n+1$ points in the $\mathbf{R}^n$ search space and deterministically follows a set of geometric transformations to the simplex, which directs it towards the optimum of the specified nonlinear unconstrained function. We incorporate viewpoint selection constraints implicitly in our viewpoint parametrization as well as by explicitly penalizing the evaluation of (9) for invalid viewpoints.

6 Experiments

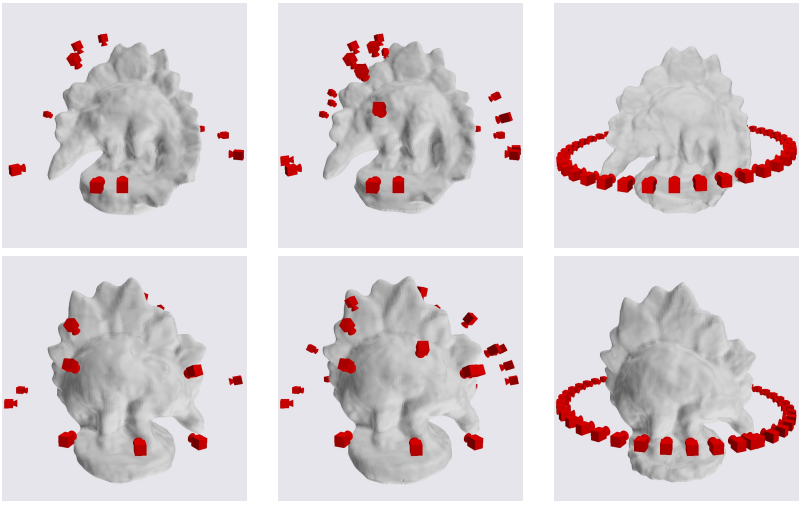
Experiments on real images were performed using the Middlebury *dino* dome dataset. Our method determines the NBV and selects the closest available camera on the dataset. The viewpoint parameterization (2) was modified to spherical coordinates with constant radius c, in order to enforce the camera placement restrictions imposed by the dome viewing configuration. In this way, for two rotation angles $\phi \in [0,\pi]$ and $\kappa \in [0,2\pi]$, we have

$$\nu_j = [\mathbf{x}_j(\phi,\kappa,c), \theta_j(\phi,\kappa)] : \{\mathbf{x}_j(\cdot) : \mathbf{R}^3 \rightarrow \mathbf{R}^3, \theta_j(\cdot) : \mathbf{R}^2 \rightarrow SO(3)\}. \tag{10}$$

In this experiment, the model update phase is performed using the algorithm developed by Furukawa & Ponce [3], which offers robust performance for multiview stereo reconstructions. The reconstruction algorithm of [3] was executed for each novel viewpoint selected by our planner. Similarly to [6], we compared the performance of our NBV planner against an evenly distributed set of cameras. A total of 48 evenly distributed cameras comprising the base of the hemisphere of the dataset were selected. The obtained 3D model was compared against models generated from configurations of 16 and 32 viewpoints designed by our NBV planner. The resulting accuracy and coverage can be seen in Figure 1. We used parameters comparable to [3] for all our experiments. The deviating performance numbers to [3] can be explained by the different procedure used to transform the oriented point cloud into a 3D mesh [7], as well as the choice of the algorithm operating parameters. The two configurations generated by our planning approach significantly outperform 48 evenly selected views in terms of coverage (93.5% and 95.4% against 91.1%) and reconstruction error (0.87mm. and 0.74mm. against 0.94mm.). It is noteworthy that the algorithm converges systematically towards full object coverage even though our criterion in Eq. (9) does not incorporate coverage. This is a result of the fact that the most uncertain regions are at the border of the mesh given the small number of observations of those points. This naturally drives the optimization to new views covering this area and hence extending the mesh at the border.

The next experiment considers NBV planning for an autonomous agent capable of performing arbitrary sensing actions. To achieve online performance, we adopt a SfM approach to markerless camera pose estimation/tracking and deploy an efficient GPU based global depth map fusion technique [21]. To estimate depth maps for the NBV, intermediate viewpoints between the NBV and the current sensing position are interpolated. We consider an object centered scene and adopt a viewing sphere camera placement. The goal of our NBV planner is to achieve a precision of 0.002% units with respect to the sphere viewing distance of 35 cm. Our termination criteria validates that 95% of all mesh vertices fulfill our precision requirements and that the $k = 3$ most recent NBV specifications have failed to either increase the coverage of the mesh or to reduce the uncertainty of the sensed regions.

All experiments were carried out on an Intel Centrino laptop equipped with 2GB RAM and an Nvidia 570M graphics card. While the generated 3D models were in the order of



16 Planned Views (0.87mm. , 93.5%) | 32 Planned Views (0.74mm. , 95.4%) | 48 Regular Views (0.94mm. , 91.1%)

Figure 1: Experimentation on the Middlebury dino dataset. Evaluation benchmarks describe 3D reconstruction error and object coverage.

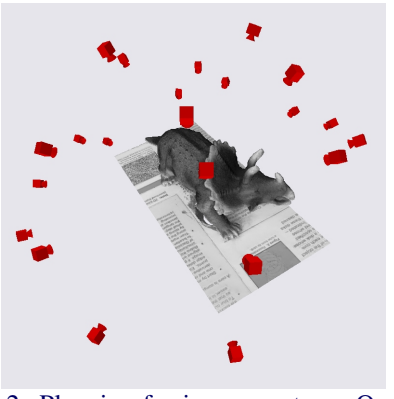
500K vertices, the adaptive patch merging and partition process was controlled to limit the maximum number of patches to 20K. This allows us to achieve a viewpoint evaluation performance of 60Hz. The simplex search for the NBV was limited to 100 iterations in order to obtain planning results within 2 seconds of completing the model update phase. Model update latency varies between 350ms to 2 seconds depending on the number of known vertices and the object region observed by the previous NBV.

7 Conclusions and future work

We presented a new NBV planning approach for active 3D model improvement. The proposed approach selects viewpoints to reduce the 3D uncertainty of an incrementally obtained scene model. In contrast to previously presented approaches, our novel approach considers the measurement uncertainties of the model and the appearance of the scene. This enables us to balance the contradicting paradigms of geometric uncertainty reduction and matchability in the image. Our experiments on different NBV planning scenarios illustrated the generic nature of our approach. Future research paths include experimentation on open scenes and the integration of our planner in an autonomous robotic exploration system.

Acknowledgements

We thank Christopher Zach and Rahul Raguram for their support in building some of the 3D models. Enrique Dunn was supported by CONACyT foreign posdoctoral visit award No. 82053.



Viewpoint	Object Coverage	Average Uncertainty
0	33%	0.783860
1	34%	0.314736
2	36%	0.121641
4	63%	0.011238
6	71%	0.001754
8	83%	0.001193
10	99%	0.000970
14	99%	0.000766
18	99%	0.000707
20	99%	0.000673

Figure 2: Planning for image capture. On the left, planned viewing configuration used to reconstruct object. On the right, reconstruction statistics.

References

[1] S. Chen, Y.F. Li, J. Zhang, and W. Wang. *Active Sensor Planning for Multiview Vision Tasks*. Springer-Verlag, Berlin Heidelberg, 2008.

[2] C. S. Fraser. Network design considerations for non-topographic photogrammetry. *PERS*, 50(8):1115–1126, 1984.

[3] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. In *Proccedings of ICPR*, 2007.

[4] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S.M. Seitz. Multi-view stereo for community photo collections. pages 1–8, Oct. 2007. doi: 10.1109/ICCV.2007.4408933.

[5] R.I. Hartley and P.F. Sturm. Triangulation. *CVIU*, 68(2):146–157, November 1997.

[6] Alexander Hornung, Boyi Zeng, and Leif Kobbelt. Image selection for improved multi-view stereo. In *Proccedings of CVPR*, pages 1–8, 2008.

[7] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *SGP '06: Proceedings of the fourth Eurographics symposium on Geometry processing*, pages 61–70, Aire-la-Ville, Switzerland, Switzerland, 2006. Eurographics Association. ISBN 30905673-36-3.

[8] Tamara G. Kolda, Robert Michael Lewis, and Virginia Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45: 385–482, 2003.

[9] E. Marchand and F. Chaumette. Active vision for complete scene reconstruction and exploration. *PAMI*, 21(1):65–72, 1999.

[10] S. Mason. Heuristic reasoning strategy for automated sensor placement. *PERS*, 63(9): 1093–1102, 1997.

[11] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, January 1965. doi: 10.1093/comjnl/7.4.308. URL http://dx.doi.org/10.1093/comjnl/7.4.308.

[12] T. Newman and A. Jain. Survey of automated visual inspection. *Computer Vision and Image Understanding*, 61(2):231–262, 1995.

[13] G. Olague and R. Mohr. Optimal camera placement for accurate reconstructions. *Pattern Recognition*, 35(4):927–944, 2002.

[14] R. Sablatnig, S. Tosovic, and M. Kampel. Next view planning for a combination of passive and active acquisition techniques. In *Proceedings of 3DIM*, page 62Ű69, 2003.

[15] W. R. Scott, G. Roth, and J.-F. Rivest. View planning for automated three-dimensional object reconstruction and inspection. *ACM Comput. Surv.*, 35(1):64–96, 2003.

[16] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:519–528, 2006. ISSN 1063-6919. doi: http://doi.ieeecomputersociety.org/10.1109/CVPR.2006.19.

[17] K.A. Tarabanis, P.K. Allen, and R.Y. Stag. A survey of sensor planning in computer vision. *IEEE Trans. on Rob. and Automat.*, 11(1):86–104, 1995.

[18] S. Wenhardt, B. Deutsch, J. Hornegger, H. Niemann, and J. Denzler. An information theoretic approach for next best view planning in 3-d reconstruction. In *Proccedings of ICPR*, 2006.

[19] Stefan Wenhardt, Benjamin Deutsch, Elli Angelopoulou, and Heinrich Niemann. Active visual object reconstruction using d-, e-, and t-optimal next best views. In *Proccedings of CVPR*, 2007.

[20] P. Whaite and F.P. Ferrie. Autonomous exploration: driven by uncertainty. *PAMI*, 19 (3):193–205, 1997.

[21] C. Zach. Fast and high quality fusion of depth maps. In *International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, 2008.