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

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*. [2], Tarabanis *et al*. [4] and Scott *et al*. [3]. Recently, Chen *et al*. [1] provided a broad coverage of multiple research areas within sensor planning. We propose a novel approach to determining the Next Best View (NBV) for the task of efficiently building highly accurate 3D models from images. The developed NBV system incrementally builds a sensing strategy by sequentially finding the single camera placement that best reduces an existing model's 3D uncertainty. In our approach, scene structure is locally approximated through adaptive planar patches, P_i, parameterized in terms of their position, orientation, surface appearance and 3D reconstruction uncertainty. These elements are used 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. We associate these properties with different reconstruction goals to be modeled into our criterion.

The first goal addressed is to achieve an adequate incidence with respect to a patch 3D uncertainty. We leverage the 3D uncertainty information contained in the scaled eigenvector matrix Ψ obtained from each patch positional covariance structure. Local scene structure and uncertainty are estimated through an extended Kalman filter framework based on optical triangulation. 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, $\mathbf{v}$, best reduces the 3D uncertainty contained in Ψ_i. We propose to find the viewing ray minimizing

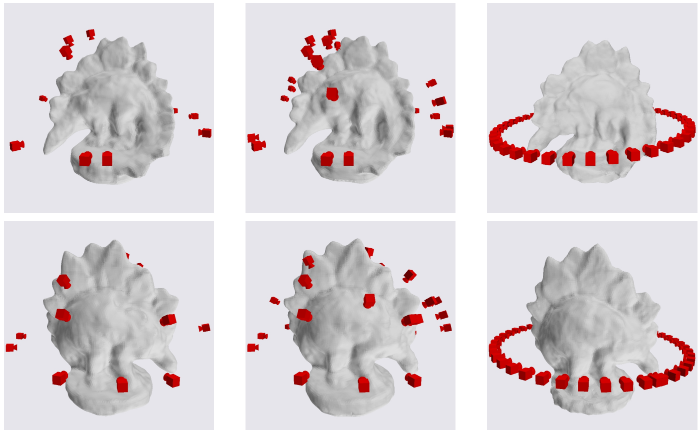
$$f(P,v) = \left|\left| \left(\frac{\mathbf{X}_i - \mathbf{x}_j}{\|\mathbf{X}_i - \mathbf{x}_j\|_2} \right)^T \left[\lambda_1 \mathbf{e}_1^i \; \lambda_2 \mathbf{e}_2^i \; \lambda_3 \mathbf{e}_3^i \right] \right|\right|_2 = \|\mathbf{v}^T \Psi_i\|_2. \quad (1)$$

The minimization of (1) considers 3D reconstruction as a merely geometric task, not taking into account practical aspects such as robustness and accuracy of image measurements and matching saliency.

The second goal factor under consideration is to obtain a favorable imaging resolution for the projected texture of a given 3D patch. This is motivated by observing that better accuracy in image measurements can be obtained as the imaging resolution increases, provided that the object texture is having sufficient visual richness. Given some fixed internal camera parameters, the main factors in determining the imaging resolution of an observed planar patch are viewpoint proximity and viewing ray incidence angle. We model the trade-off between these two factors by measuring a single quantity: the area of projection of a 3D surface onto the image plane. Moreover, it is straightforward to compute this quantity analytically for simple geometric primitives. Alternatively, it can be computed with high efficiency deploying a GPU for the required rendering.

The third goal incorporated into our criterion is to condition the relevance of imaging resolution on the texture of the observed surface. In this way, oblique views are favored for observing textureless surface regions where feature detection and extraction is unattainable. This relevance factor is modeled by a continuous step function with transition at a given texture threshold. We model this behavior by the error function. 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. 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 $a_i \in [-1,1]$. The matrix values are normalized and used to evaluate Shannon entropy. Hence, we obtain a texture saliency function

$$h(P_i) = 1 - erf\left(- \sum_{a_i \in A(\mathbf{S}_i)} p(a_i) \log p(a_i) \right). \quad (2)$$



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.

The aforementioned reconstruction factors are related by the formula

$$C(v,P) = \frac{g(v,P)^{h(P)}}{f(v,P)}, \quad (3)$$

where $g(v,P)$ denotes the viewpoint dependent projection area of the 3D primitive, while $h(P)$ and $f(v,P)$ are defined in Eqs. (2) and (1) respectively. Equation (3) is evaluated for each patch P_i and combined through a weighted sum to define our aggregate NBV criterion

$$F(v) = \sum_{i=0}^{N} w_i C(v,P_i). \quad (4)$$

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 priority in the viewpoint search process based on a simplex nonlinear optimization module.

The developed system can be used within active computer vision systems as well as for optimized view selection from a set of available views. In both of these scenarios, our results illustrate the effectiveness of the proposed planner at systematically reducing reconstruction uncertainty as well as increasing object surface coverage. As illustrated by the benchmarks in Figure 1, reductions of 20% in reconstruction error can be achieved by selecting viewpoints using our approach, while incurring in 33% less image acquisitions when compared against a regular viewing configuration. Moreover, we achieve planning performance that enables the integration of our planner into the online operation of active vision systems.

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

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

[3] 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.

[4] 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.