INCORPORATION OF A-PRIORI INFORMATION IN PLANNING THE NEXT BEST VIEW

Christoph Munkelt[1], Peter Kühmstedt[1] and Joachim Denzler[2]

[1]) Fraunhofer IOF,
Albert-Einstein-Str. 7, 07745 Jena, Germany
Email: {christoph.munkelt|peter.kuehmstedt}@iof.fraunhofer.de

[2]) Chair for Computer Vision,
Computer Science Department,
Friedrich-Schiller-University Jena
07743 Jena, Germany
Email: denzler@informatik.uni-jena.de

KEY WORDS: Optical, Three-dimensional, Measurement, Close Range, Acquisition, Planning, Automation

ABSTRACT:

The widespread use of optical 3D measurement in inspection and prototyping continues to demand improved strategies for planning the next best view (NBV). During the scanning process, better viewpoints can be chosen and fewer views need to be taken, if a NBV algorithm is able to use available a-priori information about an object to-be-scanned. The proposed extension to a voxel space based NBV algorithm initialises the planner with a low detail geometric model of the object. Iterative updates of this model, according to the data of the scanning process, enable the adaptation to possible inconsistencies. First promising results from tests with fringe projection ray-tracing simulations are presented.

1. INTRODUCTION

To sample all visible surfaces of nontrivial objects requires multiple range measurements from different viewpoints. In an automated surface acquisition system a NBV algorithm is used to determine the next viewpoint. The object/scanner is moved and a new measurement starts. Incrementally, surface data sampled from those positions is added to the partial object model. After sufficient surface data has been acquired, the algorithm should terminate the scanning process.

Different strategies have been employed for the different vision tasks at hand. If the object's geometries and poses are known (e.g. in model-based object recognition or in the "art gallery problem"), optimal sensor positions can be determined (see e.g. (Roberts and Marshall, 1998)). On the other hand, in situations with very limited object information, e.g. reverse engineering, NBV algorithms used simpler approaches to determine areas to scan. Nearly all researchers, according to (Pito, 1999), address this problem by identifying range discontinuities. Principles used herein include, but are not limited to, occlusions as a guide for planning the next view (Maver and Bajcsy, 1993) and volumetrically representations of the viewing volume (Massios and Fisher, 1998), (Banta, Wong, Dumont, and Abidi, 2000). See (Tarabanis, Allen, and Tsai, 1995) for a survey of sensor planning in computer vision.

One of the goals of this work is to make use of some of the available object knowledge while maintaining the intuitiveness and simplicity of the volumetric approaches. This targets at scenarios, where a prototype CAD model of the object (e.g. in inspection) at hand or a rough scan or some other sort of abstraction gives clues about the object geometry. Two aspects should benefit from such an improvement: the early stage of the scan ("begin game" according to (Hall-Holt, 1996)) completes with less views. Secondly, a better initial quality of measured points is achieved due to better chosen viewpoints in respect to object surface normals.

The rest of this paper is organized as follows. Section 2. discusses assumptions and terms used by the remainder of this article. Section 3. outlines available optimality criteria and develops the furthermore utilised criterion. Thereafter, the proposed NBV algorithm is introduced, which is based on these criteria. Section 4. describes an implementation and the results of scanning in a fringe projection simulation framework. Finally this paper is summarized by Section 5., which also depicts future work.

2. ASSUMPTIONS

In order to concentrate on the question, how a-priori information about the geometry of the to-be-scanned object can improve the NBV planning procedure, some assumptions have been made.

First, when speaking of *the sensor*, we refer implicitly to at least a fringe projector and one camera in a fixed setup. Similarly, *sensor position* stands for the positions of both - the exact positions of camera and projector can easily be calculated. This basic setup can be compared to a stereo camera system with a fixed base and orientation. However, to limit the effects of the resulting stereo visibility constraints, a rather small triangulation angle has been chosen.

To reduce the computational burden of the NBV algorithm, potential sensor positions are chosen from a circular path with the object in its centre. The sensor is always targeting the point of origin, in which the object is located (see Fig. 1 for illustration).

The actual intensity images used as input to the given fringe projection coordinate calculation kernel were created from a ray-tracing simulation using POV-Ray[1]. The object surface is assumed to be lambertian. This ensures easy, controlled, and reproducible sensor positioning.

As camera model we deal with the ideal pinhole model. Since the projector can be viewed as an inverse camera, this also holds for the projector. In the rendering stage of the NBV algorithm, the intrinsic parameters of the cameras are set to the values used by the ray-tracing.

[1]Persistence of Vision ray-tracer: http://www.povray.org

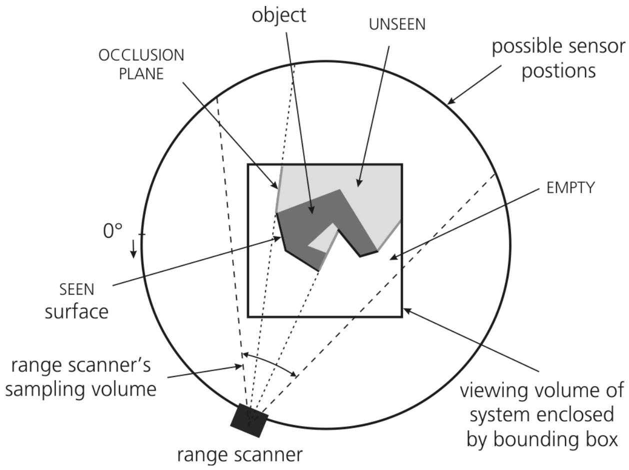


Figure 1. Simulated measuring setup and corresponding viewing volume. The resulting classification into EMPTY, SEEN, UNSEEN and OCCLUSION PLANE voxel of a measurement is shown.

2.1 Definitions

The whole measurement volume is divided into cubes with equal side length. Those cubes, commonly called *voxels* (short for volumetric pixel), are internally arranged in a 3-D array. In contrast to an occupancy grid, as seen in (Banta, Wong, Dumont, and Abidi, 2000), each voxel can store more information associated to the measurement points it represents. For instance, we assign both a quality and a voxel status to each voxel.

Local voxel space denotes the "scratch pad". Here basic marking is applied to the data coming from individual measurements. *Global voxel space* on the other hand accumulates all the information from the views taken so far.

Each voxel is associated with a set of attributes. The voxel's normal vector (*norm*) describes the surface normal of the patch of object surface within the voxel's volume. It is approximated via a plane fit over the voxel's points. The quality estimate ($0 \leq qual \leq 1$) should describe the confidence in the measurement of the points in this voxel. A typical measure of confidence for triangulation-based range scanners at a measured point is the cosine of the angle between the surface normal and the viewing ray at a particular point (Pito, 1999). Here we approximate it as the dot product between the viewing direction $\hat{v}$ of the sensor and this voxel's surface normal: $qual = norm \cdot \hat{v}$.

To allow for viewpoint evaluation, all voxels in a voxel space are classified as follows (according to (Massios and Fisher, 1998), see also Fig. 1):

EMPTY: (E) The volume of the voxel is free of measured points.

SEEN: (S) Measured points lie within this voxel's volume. Attributes like *qual* and *norm* are associated to this voxel.

UNSEEN: (US) The volume of this voxel has not been seen yet. Other voxels might occlude it.

OCCLUSION PLANE: (OP) The area of space of this voxel lies on an occlusion plane (see Fig. 1 and Fig. 2c). These are voxels on the border between unseen and empty voxels.

To aid the planning process in the first phase of measurement, it would be beneficial to rely on supplementary information in addition to the points scanned so far. Possible occlusions could be avoided and better viewpoints be chosen, if the planner would have access to even very approximate model information (see Fig. 2a and b for an example).

To consistently extend the given framework to use this kind of a-priori information, voxels which include only points from such a prototype model are treated as SEEN voxels with quality set to 0. We will refer to these voxels as HINT voxels.

3. CRITERIA OF OPTIMALITY

There are many different goals a NBV algorithm can strive for. Among the most popular is to achieve the "optimal" result after each particular scan. "Optimal" often means the highest score in some weighting function still satisfying a set of constraints. For instance, constraints can be the sheer possibility of a object surface to be scanned, with regards to the characteristics of the sensor positioning and the sensor itself. The weighting function usually guarantees the knowledge increase in terms of scanning so far unseen areas or enhance confidence in previously taken measurements. The weighting function $f_2(\hat{v})$ presented in this paper tries to maximise the information gain solely for the next view.

Other goals could include an optimal result after *n* scans, which takes time and memory considerations into account. Graceful abortability, which warrants optimality after *n* more scans, can be seen as a subclass of this goal.

One important constraint not handled in our approach is the overlap constrain. If there is sufficient overlap between a scan of so far unseen and already seen areas of the object, overall quality of the resulting aggregated data set increases. This is due to the fact that many algorithms for registration and integration of range data perform best when the range data overlaps (Pito, 1999). However, since within our ray-tracing simulation environment exact sensor positions are known, no special registration / integration steps are needed.

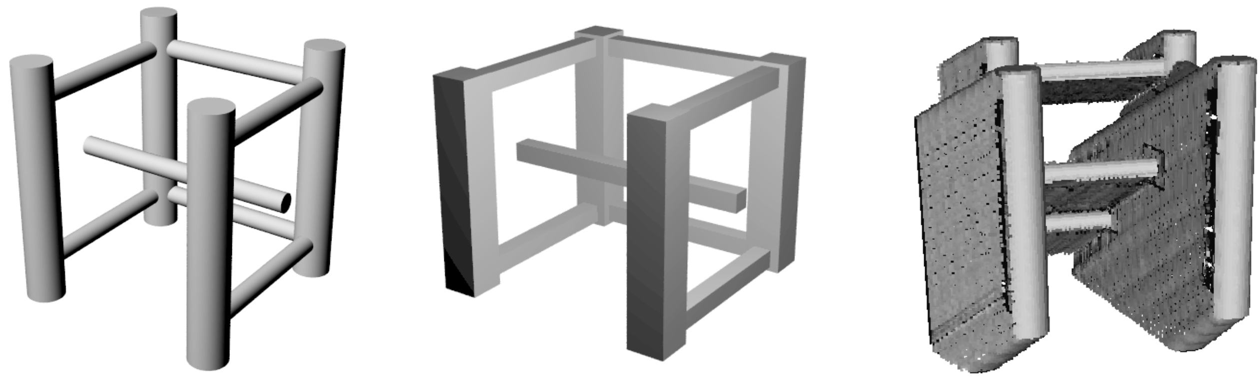
3.1 Evaluation criteria for the NBV

The proposed NBV algorithm is an extension to the algorithm described in (Massios and Fisher, 1998). As such, it maximises the amount of visible OCCLUSION PLANE voxels in order to scan still unseen areas. Furthermore it tries to rescan voxels with low estimated quality.

However, to aid the NBV algorithm in finding good views during the start phase, when large portions of the viewing volume are still unseen, the concept is enhanced as follows. Often knowing the prototype geometry and pose of the object to be scanned, one can create a point cover[2] of this object. This can either be done by a coarse scan using uniformly distributed sensor positions (if no CAD model is available), or by creating one using the CAD model.

We used this abstraction of the object, called the *hint model*, to initialize the global voxel space. As long as a sufficient amount of HINT voxels are still visible, the planner uses only the hint criterion (see Sec. 3.2) to compute the next view. During fusion of new views by repeatedly merging their local into global voxel space, HINT voxels are gradually replaced by EMPTY or SEEN voxels or are occluded and therefore not visible anymore. If the

[2]The point cover function creates a point model based on an existing polygon or CAD model.



(a) the ideal pipe model

(b) approximation of the pipe model for use as HINT data

(c) seen and occlusion plane voxels

Figure 2. The pipe model (Fig. a) and its abstraction (Fig. b) for use as a hint during early planning stages. Fig. c) shows SEEN and resulting OCCLUSION PLANE voxels from a range measurement.

maximal amount of visible HINT voxels in any evaluated sensor position drops below a certain threshold, the evaluation uses the conventional approach for the remainder of views.

To summarise, our NBV algorithm maximises the following in order to determine where to look next:

- during start phase: number and quality of visible HINT voxels with sufficient quality
- during model refinement: number of visible OCCLUSION PLANE voxels as well as number of SEEN voxels with low quality

3.2 Weighting functions

The projection of the global voxel space onto potential sensor positions allows for their evaluation. The visibility criterion $f_v(\widehat{v})$ introduced in (Massios and Fisher, 1998) maximises the number of visible OCCLUSION PLANE voxels from position $\widehat{v}$:

$$f_v(\widehat{v}) = sizeof\left(OP \cap V(\widehat{v})\right), \tag{1}$$

where $V(\widehat{v})$ is the set of visible voxels from sensor position $\widehat{v}$.

The quality criterion $f_q(\widehat{v})$ maximises the number of already SEEN voxels with low quality, which could be rescanned from the currently evaluated position:

$$f_q(\widehat{v}) = \frac{\sum_{i=1}^{sizeof(S\cap V(\widehat{v}))} \left(1 - qual(S[i])\right)\left(|\widehat{v} \cdot norm(S[i])|\right)}{sizeof\left(S \cap V(\widehat{v})\right)}. \tag{2}$$

This criterion favours sensor positions, which rescans voxels with low quality estimate vertically (i.e. where $|\widehat{v} \cdot norm(S[i])| \sim 1$).

To combine both criteria according to (Massios and Fisher, 1998), the weighted sum $f_1(\widehat{v})$ is calculated as follows:

$$f_1(\widehat{v}) = w_v * f_v(\widehat{v}) + w_q * f_q(\widehat{v}). \tag{3}$$

The weights were empirically set to $w_v = 1$ and $w_q = \max(f_v)$ in order to ensure comparability between the two criteria.

As an extension to these principles from (Massios and Fisher, 1998), we use given a-priori information to choose better viewpoints during the first measurements. The evaluation criterion is similiar to the quality criterion, but takes only HINT voxels into account. To prefer sensor positions, which not only view many so far UNSEEN voxels, but also favour good viewing angles[3], only HINT voxels which are seen under smaller viewing angles than θ_b are considered. The breakdown angle θ_b, according to (Pito, 1999), must be smaller than 90° and is, in most scanners, around 60°. The equations 4 and 5 express this condition:

$$sp_{min}(\widehat{v}, \widehat{n}) = \begin{cases} |\widehat{v} \cdot \widehat{n}| & \text{if } |\widehat{v} \cdot \widehat{n}| > \cos(\theta_b) \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$spc_{min}(\widehat{v}, \widehat{n}) = \begin{cases} 1 & \text{if } |\widehat{v} \cdot \widehat{n}| > \cos(\theta_b) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$f_h(\widehat{v}) = \sum_{i=1}^{sizeof(H\cap V(\widehat{v}))} spc_{min}\left(\widehat{v}, norm(H[i])\right) * \frac{\sum_{i=1}^{sizeof(H\cap V(\widehat{v}))} sp_{min}\left(\widehat{v}, norm(H[i])\right)}{sizeof\left(H \cap V(\widehat{v})\right)}. \tag{6}$$

The hint criterion according to Eq. 6 scales the amount of visible HINT voxels, which meet this breakdown angle condition, with the quality and number of all visible HINT voxels. This consciously penalises viewpoints which scan large object areas under questionable conditions.

If a-priori information is available, our NBV algorithm starts viewpoint evaluation with the hint quality criterion $f_h(\widehat{v})$. If the maximum amount of visible HINT voxels over all evaluated possible viewpoints drops below threshold t_{vh}, viewpoint evaluation switches to criterion $f_1(\widehat{v})$:

$$f_2(\widehat{v}) = \begin{cases} f_1(\widehat{v}) & \text{if } \max\left(sizeof\left(H \cap V(\widehat{v})\right)\right) < t_{vh} \\ f_h(\widehat{v}) & \text{otherwise} \end{cases} \tag{7}$$

with t_{vh} empirically set to 10% of the number of maximum visible HINT voxels in “view 0” (before the start of NBV determination).

[3] A voxel is scanned under a good viewing angle, if it is sampled nearly vertically, i.e. $|\widehat{v} \cdot norm(S[i])| \sim 1$

3.3 Termination criteria

Some examples of possible termination criteria are:

1. An already visited position is chosen as next sensor position (Massios and Fisher, 1998).
2. Realisation of the required completeness and minimal qualities in at least x% of the seen voxels.
3. Low estimated information gain during the next view.

For our system we chose criterion 3. If $f_2(\hat{v})$ is below a threshold t_{sc}, we stop the measuring process. In our experiments we set t_{sc} to approximately 1% of the maximum $f_1(\hat{v})$ during evaluation of "view 1" (after the first measurement was integrated).

3.4 Priority of voxel marking

The integration of newly acquired views into the global voxel space is the key to planning success. To ensure consistency between the planning model and the real object, the following rules are applied when updating the global voxel space:

- SEEN voxels are allowed to update UNSEEN, EMPTY, HINT and OCCLUSION PLANE voxels
- EMPTY voxels can overwrite UNSEEN, HINT and OCCLUSION PLANE voxels
- OCCLUSION PLANE voxels may overwrite only UNSEEN voxels

Some conflicts, however, arise due to the discrete nature of the voxel space. Empty voxels shouldn't be updated to seen, for instance. Also, voxels with just very few points inside should perhaps be considered empty. Plans to deal with such situations are explained in Sec. 5..

3.5 Sequence of evaluation

After each measurement, the acquired data has to be transformed into the global coordinate system in order to make it available to the planning process. First, the local voxel space is initialized and every voxel's status is set to EMPTY.

The following basic marking is then applied while still in the local voxel space. Measured points within a voxel set its status to SEEN. Voxels on the ray from a SEEN voxel away from the sensor get marked UNSEEN. Voxels, which lie on the ray from a SEEN or bounding box voxel towards the sensor, will be marked as EMPTY.

After basic marking, the voxel status of the local system is integrated into the global system. This is done according to the status update rules (see Sec. 3.4). Then voxels on occlusion planes are determined in the resulting global model

Following this global integration the evaluation of possible next sensor positions takes place. First, a number of reachable and allowed positions are generated. Second, the voxel space is projected onto each of those positions. The projections, which include SEEN, HINT and OCCLUSION PLANE voxels, are then evaluated according to the weighting functions in Sec. 3.2.

Finally, the view pose with the maximum score in $f_2(\hat{v})$ is chosen as the next view. Its parameters are used to generate the next view with the ray-tracing simulation. This sequence is repeated until a termination criterion from Sec. 3.3 is met.

4. EXPERIMENTAL RESULTS

4.1 System integration

First test were conducted by implementing the core functionality in C/C++ and using OpenGL to perform the projections and calculate, for instance, dot products in image precision. The actual measurement process was simulated by creating ray-traced input pictures to our fringe projection system. This was done due to the better controllability of view positions in this early stage. Later, tests with a six degrees of freedom eye-in-hand robot positioning system will take place.

Voxels were drawn as actual cubes with several information encoded. The colour of a voxel of a projection of the global model onto a possible sensor position can represent its index in voxel space, approximated angle between its surface normal and the viewing direction or the distance to the sensor. Furthermore, visibility of a given voxel with respect to other already measured points and a chosen sensor position is determined automatically by the rendering process.

4.2 Results using the pipe model

In this section, we present first results from the NBV system. The resolution of the rendered intensity images used as input for 3D-calculation is 640×512, and the resolution of the voxel space is $160 \times 160 \times 160$. All tests were conducted using a pipe model (see Fig. 2a, similar to the one used in (Banta, Wong, Dumont, and Abidi, 2000)). This low complexity object consists of cylinders in a grid pattern, such that the model contains many self occlusions. Status update and evaluation of 36 possible sensor positions (equiangular distributed every 10°) took approximately 15 s each on an Intel Pentium IV 3.0 GHz with a low cost graphics accelerator.

To compare our approach against the basic strategy described by the $f_1(\hat{v})$ criterion (Eq. 3), we took two series of measurements. One with model initialisation through hint data (see Fig. 3 and 4), one without (see Fig. 5). Both terminated after seven single measurements with reasonable[4] and comparable completeness.

4.3 Problems

The basic challenge of our approach will be the alignment of the hint model with respect to the measured data. While this is no problem in our simulation environment or when using rough scans of the object as hint data, strategies for alignment of a CAD model need to be integrated as prerequisite for the usage of such data.

The handling of voxels inside the object also needs to be improved. This would also lead to proper determination of the surface normals of the hint model, which is a minor visual flaw.

5. CONCLUSIONS AND OUTLOOK

We presented an extension to voxel space based NBV algorithms, which enables the usage of available a-priori information. This geometric object information, which can be generated from a rough scan or CAD model, works well as planning base. However, no clear advantage above other NBV algorithms using volumetric representations could be achieved. Reasons probably lie

[4] Since the allowed sensor positions were on a circular arc above the model, the bottom side of the model remained unscanned.

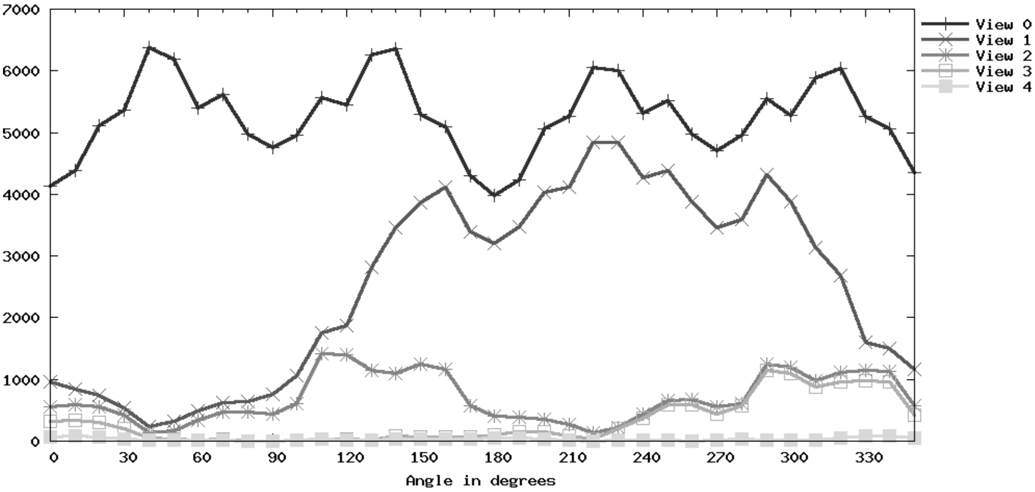
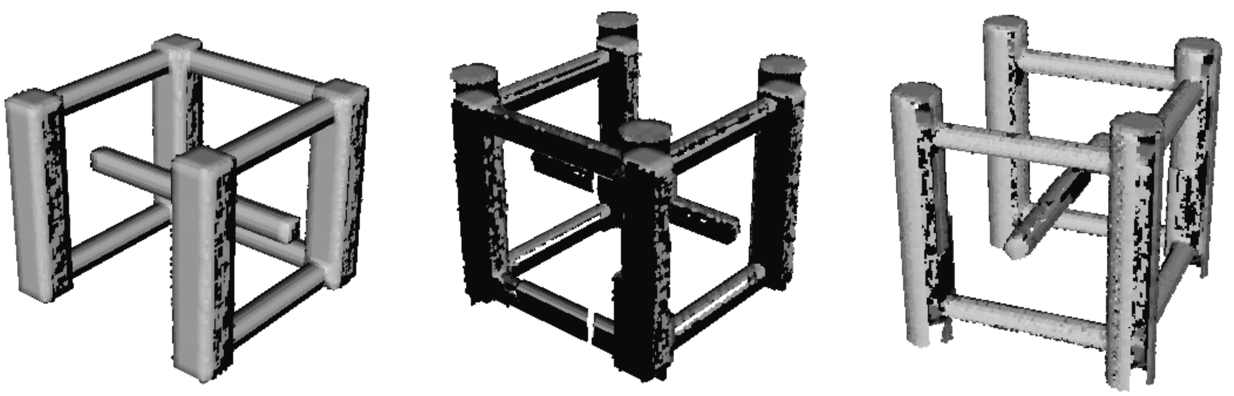


Figure 3. The hint criterion $f_h(\hat{v})$ (Eq. 6) during the evaluation of the pipe model (using HINT data from Fig. 2). View 0 denotes the determination of the first view. After four simulated scans (angles 40°, 220°, 110° and 290°) the number of visible HINT voxel drops below the threshold t_{vh}. Further scanning is conducted using the $f_1(\hat{v})$ criterion.



(a) 40° (b) 220° (c) 110°

Figure 4. The global voxel space (HINT and SEEN voxel) projected onto the chosen viewpoints from Fig. 3. The hint model is gradually overwritten (black voxel result from flipped surface normals).

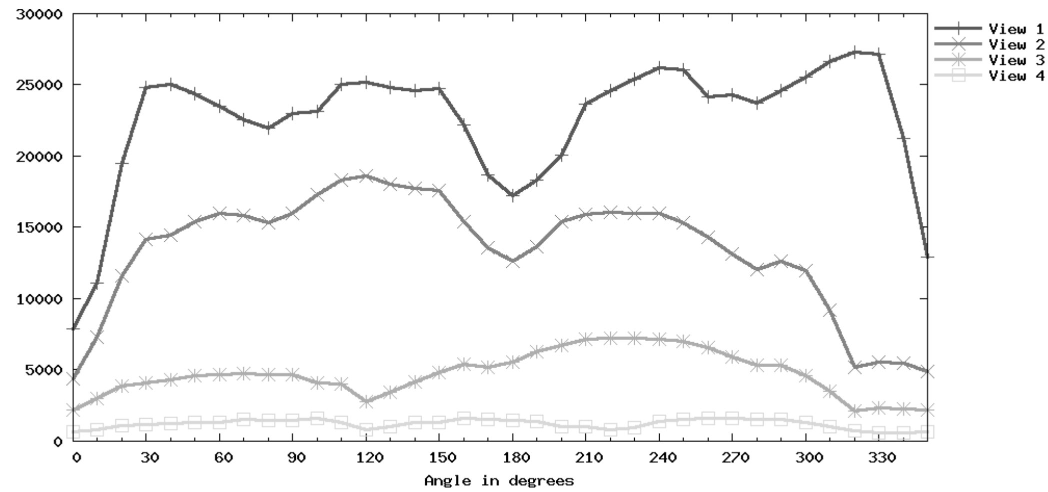


Figure 5. The $f_1(\hat{v})$ criterion (Eq. 3) during the evaluation of the pipe model (no HINT data used). The first scan was arbitrarily chosen to be 0°. The planner selected 320° as next view ("View 1"), followed by 120°, 220° and 270°.

in the hitherto existing inability to model the characteristics of the sensor consisting of camera and fringe projector (keyword: stereo visibility). Additionally, inaccuracies during voxel marking and the used discrete voxel states still limit the application to the first phase of measurement.

Future work will therefore concentrate on enhancing the discrete voxel states with floating point measures of confidence and voxel fill level. Furthermore, precision of the iterative update of the planning model can be increased, if hierarchical voxel structures (octrees) will be used. This enables also larger voxel spaces and faster operation. Eventually, planning of more than one view ahead, while minimizing the number of needed scans, can be realised using this approach. Another interesting possible field of application is the scanning of large measurement volumes, where the sensor's field of view is small compared to the entire volume.

REFERENCES

Banta, J., Wong, L., Dumont, C., and Abidi, M., 2000. A next-best-view system for autonomous 3-d object reconstruction. *IEEE Trans. Systems, Man and Cybernetics* 30(5), 589–597.

Hall-Holt, O., 1996. Scrutinizing Real-World Geometry: the Next Best View. `http://www-graphics.stanford.edu/~olaf/nbv.html`.

Massios, N. A. and Fisher, R. B., 1998. A Best Next View selection algorithm incorporating a qulity criterion. Mastersthesis, University of Edinburgh.

Maver, J. and Bajcsy, R., 1993. Occlusions as a Guide for Planning the Next View. *IEEE Trans. Pattern Analysis and Machine Intelligence* 15(5), 417–433.

Pito, R., 1999. A Solution to the Next Best View Problem for Automated Surface Acquisition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(10), 1016–1030.

Roberts, D. R. and Marshall, A. D., 1998. Viewpoint Selection for Complete Surface Coverage of Three Dimensional Objects. In *British Machine Vision Conference*, Department of Computer Science, Cardiff University, PO Box 916, Cardiff. CF2 3XF, U.K., pp. 740 – 750.

Tarabanis, K. A., Allen, P. K., and Tsai, R. Y., 1995. A survey of sensor planning in computer vision. In *IEEE transactions on Robotics and Automation*, pp. 86–104. 11(1):86-104.