Active Visual Object Reconstruction using D-, E-, and T-Optimal Next Best Views

Stefan Wenhardt, Benjamin Deutsch[*], Elli Angelopoulou, Heinrich Niemann
Universität Erlangen-Nürnberg
Martenstrasse 3, 91074 Erlangen, Germany
wenhardt@informatik.uni-erlangen.de

Abstract

In visual 3-D reconstruction tasks with mobile cameras, one wishes to move the cameras so that they provide the views that lead to the best reconstruction result. When the camera motion is adapted during the reconstruction, the view of interest is the next best view for the current shape estimate. We present such a next best view planning approach for visual 3-D reconstruction. The reconstruction is based on a probabilistic state estimation with sensor actions. The next best view is determined by a metric of the state estimation's uncertainty. We compare three metrics: D-optimality, which is based on the entropy and corresponds to the (D)eterminant of the covariance matrix of a Gaussian distribution, E-optimality, and T-optimality, which are based on (E)igenvalues or on the (T)race of this matrix, respectively. We show the validity of our approach with a simulation as well as real-world experiments, and compare reconstruction accuracy and computation time for the optimality criteria.

1. Introduction

One fundamental task in computer vision is the reconstruction of a three-dimensional scene or object using only two-dimensional images. If only 2-D data is used, it is clear that one needs at least two images, if not more, to obtain a reconstruction. In many cases, the actual images are either predetermined, such as when reconstructing from a video, or taken from fixed views due to mechanical constraints.

In some cases, however, the reconstruction system is free to choose the views from which the images are generated. While it is typically true that more views can only provide more information, one is nevertheless interested in keeping the number of different views as low as possible to save time. For such constraints, it is desirable to find the sequence of views which allows optimal reconstruction, *i.e.* which minimizes the reconstruction error. The search for this sequence is the *next best view problem*.

Previous solutions to the next best view problem using 2-D images are quite rare in literature. Usually, these works are based on geometrical considerations to determine the next best view.

Kutulakos et. al. [6, 7, 8] use a structure from occluding contour approach. They analyze the first and second Gaussian fundamental forms of the surface of the object. Depending on the values of the fundamental forms, they can determine the object surface category (elliptic, hyperbolic, parabolic or planar). Further they derive a strategy to optimally estimate the values of the fundamental forms. A similar approach is used by Niem [11].

Marchand and Chaumette [1, 2, 10, 9] use a structure from controlled motion approach. They reconstruct geometrical primitives (points, lines, spheres, cylinders). These objects are described by their parameters (e. g. center and radius for a sphere), which are extracted from contour lines. They present some rules for how to reduce errors during contour line tracking and parameter estimation. The proposed algorithm is not limited to such primitives in general, but the calculation for complex objects is very expensive.

We present an approach to the next best view question, based on the works of [4] and more specifically [15]. This approach is not based on geometric clues, but instead considers each part of the scene to be reconstructed from a probabilistic viewpoint. The goal is to use the current information about the given scene to find the view which most reduces the uncertainty in this information. A similar method is used in [3], in which the next 2-D feature for object tracking is selected via mutual information between the state and the features, and between two features.

We express the next view problem in a more general probabilistic framework that allows the incorporation of different optimization metrics. We develop a modified D-optimality criterion, motivated by the information content

[*] This work was partially funded by the German Science Foundation (DFG) under grant SFB 603/TP B2. Only the authors are responsible for the content.

of the distribution and equivalent to the information theoretic measure of entropy. Its performance is tested against an E-optimality criterion, motivated by the desire to minimize the largest component in position error and a T-optimality criterion, which aims to minimize the trace of the covariance. These criteria are well known in optimal experimental design [12].

The main focus of this work is on the relative performance of these three optimality criteria. Their systematic analysis in a non-complex, controlled setup will show the limits and strengths of the different methodologies before further constraints are introduced to the system. Therefore, the important constraint of (self) occlusion is not yet handled in this paper, but is the topic of current, parallel research. Occlusions occur when, for a given view, part of the scene is hidden from the camera by other parts. Likewise, the exhaustive search used to find the next best view would not be used in a system with time constraints.

This paper is organized as follows: The next section gives an overview of state estimation with a changing observation function, using the extended Kalman filter. Section 3 applies this active state estimation system to the task of 3-D reconstruction, and discusses the compared optimality criteria. Section 4 shows the validity of our approach via experimental evaluations. Finally, section 5 summarizes this approach, and provides an outlook for potential future work.

2. Active state estimation

For our reconstruction system, we use a Kalman filter to fuse the 2-D information from several images into a comprehensive 3-D scene representation. Since the observation functions are non-linear, due to the perspective projection in the cameras, we use the *extended* Kalman filter. As we control the view for each observation, we use a Kalman filter with *sensor actions*.

2.1. The Kalman filter

The Kalman filter is briefly described here to introduce our notation, and show the extension of the filter with sensor actions. Interested readers are referred to [5, 14].

The Kalman filter estimates the state $\boldsymbol{q}_t \in \mathbb{R}^n$ of a discrete-time dynamic system at time step t. At each time step, an observation $\boldsymbol{o}_t \in \mathbb{R}^m$ is generated from the state by the observation step:

$$\boldsymbol{o}_t = \boldsymbol{g}(\boldsymbol{q}_t, \boldsymbol{a}_t) + \boldsymbol{r}_t \tag{1}$$

where $\boldsymbol{g}(\cdot,\cdot)$ is the known observation function, $\boldsymbol{a}_t$the sensor action and $\boldsymbol{r}$ a Gaussian additive noise process with $\boldsymbol{r}_t \sim \mathcal{N}(0, \boldsymbol{R})$. The action $\boldsymbol{a}_t$ is an abstract parameterization which influences the observation function, and which is known to or chosen by the estimation system.

The Kalman filter maintains the current estimate of the state in the form of a Gaussian probability distribution $\boldsymbol{q}_t \sim \mathcal{N}(\widehat{\boldsymbol{q}}_t, \boldsymbol{P}_t)$. Note that in contrast to the typical Kalman filter, we do not have a state propagation step. The *a priori* state estimate is therefore equal to the previous time step's *a posteriori* state estimate.

The Kalman filter incorporates new observations into the state estimate with the so-called *update step*:

$$\boldsymbol{K}_t = \boldsymbol{P}_{t-1}\boldsymbol{G}_t^{\mathrm{T}}(\boldsymbol{a}_t)(\boldsymbol{G}_t(\boldsymbol{a}_t)\boldsymbol{P}_{t-1}\boldsymbol{G}_t^{\mathrm{T}}(\boldsymbol{a}_t) + \boldsymbol{R})^{-1} \tag{2}$$

$$\widehat{\boldsymbol{q}}_t = \widehat{\boldsymbol{q}}_{t-1} + \boldsymbol{K}_t(\boldsymbol{o}_t - \boldsymbol{g}(\widehat{\boldsymbol{q}}_{t-1}, \boldsymbol{a}_t)) \tag{3}$$

$$\boldsymbol{P}_t = (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{G}_t(\boldsymbol{a}_t))\boldsymbol{P}_{t-1} \tag{4}$$

where $\boldsymbol{K}_t$ is the *Kalman gain matrix* and $\boldsymbol{G}_t(\boldsymbol{a}_t)$ is the Jacobian matrix of $\boldsymbol{g}(\cdot,\cdot)$, derived at $\widehat{\boldsymbol{q}}_{t-1}$ and using the sensor action $\boldsymbol{a}_t$. The linearization with the Jacobi matrix is necessary due to the non-linear observation function.

2.2. Sensor action evaluation

One important thing to note is that $\boldsymbol{P}_t$, the covariance matrix of the estimate, does *not* depend on the observation $\boldsymbol{o}_t$. In the extended Kalman filter, past observations do influence the covariance through linearization about the current *a priori* state estimate $\widehat{\boldsymbol{q}}_{t-1}$. However, this state estimate does not yet incorporate the *next* observation $\boldsymbol{o}_t$, and therefore, following (4), the expected covariance $\boldsymbol{P}_t$ is independent of $\boldsymbol{o}_t$. This allows $\boldsymbol{P}_t$ to be calculated *before* the view is adopted and the observation $\boldsymbol{o}_t$ is made. For each sensor action $\boldsymbol{a}_t$ of interest, we can calculate the corresponding $\boldsymbol{P}_t$. The optimal action $\boldsymbol{a}_t^*$ is the action which minimizes some criterion on $\boldsymbol{P}_t$:

$$\boldsymbol{a}_t^* = \operatorname*{argmin}_{\boldsymbol{a}_t} \phi(\boldsymbol{P}_t) \tag{5}$$

with $\phi(\cdot)$ a user-defined metric on covariance matrices. Since $\boldsymbol{P}_t$ depends on $\boldsymbol{a}_t$, $\phi(\boldsymbol{P}_t)$ is used to rate each action.

2.3. Visibility

Despite the independence of $\boldsymbol{P}_t$ on $\boldsymbol{o}_t$, there is still the problem of visibility. The Kalman update step can only be performed if an observation is actually present. In many state estimation tasks, it may not be guaranteed that an observation is present at every time step. Thus, at time steps t without an observation, the update step is skipped and $\boldsymbol{P}_t$ is identical to $\boldsymbol{P}_{t-1}$.

In the context of sensor action evaluation, the expected rating of an action must be considered. If it is possible to determine the probability w of an observation being present after the next time step, the optimal action is determined by the expected metric of the covariance:

$$\boldsymbol{a}_t^* = \operatorname*{argmin}_{\boldsymbol{a}_t} w \cdot \phi(\boldsymbol{P}_t) + (1-w) \cdot \phi(\boldsymbol{P}_{t-1}) \quad . \tag{6}$$

We will see in the next section how w can be determined during the problem of 3-D-reconstruction.

3. Next best view planning for reconstruction

The previous section described how we can find the optimal sensor action for a general active state estimation system. This section shows how 3-D reconstruction can be formulated as such a problem, and which optimization criteria can be used.

3.1. Reconstruction as state estimation

We choose a surface point based reconstruction, *i.e.* the state $\boldsymbol{q}_t \in \mathrm{I\!R}^{3i}$ is the concatenation of i 3-D points on the surface of the scene or object to be reconstructed. The observation $\boldsymbol{o}_t \in \mathrm{I\!R}^{2i}$ is the concatenation of the projection of these points into a camera's image plane. These projections can be obtained, for example, by feature point tracking on the camera images. A starting estimate for $\widehat{\boldsymbol{q}}_0$ can be obtained by triangulation from two initial images. The covariance is initialized by $\boldsymbol{P}_0 = \mathrm{diag}(10, \ldots, 10)\mathrm{mm}^2$.

One optimization is the use of the *sequential Kalman filter*. Since the observation noise for each point is presumed to be independent, we can handle each 2-D point separately in i Kalman filters. As each 2-D point also affects only a single 3-D point, we can further split the state as well and transform the filtering problem from a $\mathrm{I\!R}^{2i} \rightarrow \mathrm{I\!R}^{3i}$ system to i instances of a $\mathrm{I\!R}^2 \rightarrow \mathrm{I\!R}^3$ one. All these instances use the same sensor action $\boldsymbol{a}_t$. To evaluate $\boldsymbol{a}_t$, the final $\boldsymbol{P}_t \in \mathrm{I\!R}^{3i \times 3i}$ can be calculated from the individual $\boldsymbol{P}_t^{(i)} \in \mathrm{I\!R}^{3\times 3}$ as components of a block diagonal matrix, or the metric $\phi(\cdot)$ may allow component-wise evaluation.

An important constraint is scene point visibility. As shown in section 2.3, an update can only occur for a given 3-D point if it creates a valid observation. This requires the projection of the 3-D point to lie inside the camera sensor.

The current estimate of each 3-D point is maintained in the form of a Gaussian probability distribution. A sample set of each point is drawn from the Gaussian distribution with the current mean and covariance for this point. The set can be projected to the image plane and we can count the number of points which lie in the image plane. By calculating the Laplacian relative frequency, we get an approximation of the probability w. So this is a classical Monte Carlo integration. Potential self occlusion can also influence w, but we do not handle self occlusions in this work.

3.2. Optimality criteria

Of particular importance to next best view planning is the optimality criterion, ϕ, a metric on the state estimation's covariance matrix. We compare several optimality criteria in this work: First is the information theoretic criterion of minimal entropy ϕ_D, which is shortly called D-optimality because it depends on the determinant (as described below). Next is the modified E-optimality ϕ_E, based on the maximum eigenvalue, and lastly T-optimality ϕ_T, derived from the trace.

The entropy $H(\boldsymbol{q})$ of a probability distribution $p(\boldsymbol{q})$ is defined as

$$H(\boldsymbol{q}) = \int p(\boldsymbol{q}) \log p(\boldsymbol{q}) d\boldsymbol{q} \tag{7}$$

For a n-dimensional Gaussian distribution $p(\boldsymbol{q}) = \mathcal{N}(\widehat{\boldsymbol{q}}, \boldsymbol{P})$, the entropy can be calculated in a closed form:

$$H(\boldsymbol{q}) = \frac{n}{2} + \frac{1}{2}\log(2\pi^n |\boldsymbol{P}|) \tag{8}$$

The entropy does not depend on the mean $\widehat{\boldsymbol{q}}$, only on the covariance $\boldsymbol{P}$. This allows us to use the entropy as an optimality criterion ϕ_D, as described in section 2.2. In addition, the term $\frac{n}{2}$ and the factors $\frac{1}{2}$ and $2\pi^n$ are constant with respect to action selection, and therefore do not affect the ranking of actions due to the strict monotony of the logarithm. Ignoring these terms and factors leaves us with the D-criterion

$$\phi_D(\boldsymbol{P}) := \log(|\boldsymbol{P}|) \quad . \tag{9}$$

Visibility can be integrated by treating each point separately, since ϕ_D can be separated. For example, let the scene contain $i = 2$ 3-D points. The state estimation $\widehat{\boldsymbol{q}}_t \in \mathrm{I\!R}^6$ is the concatenation of the two 3-D coordinates of these points, $\widehat{\boldsymbol{q}}_t^{(1)}, \widehat{\boldsymbol{q}}_t^{(2)} \in \mathrm{I\!R}^3$. The state covariance matrix $\boldsymbol{P}_t \in \mathrm{I\!R}^{6\times 6}$ is a block diagonal matrix with the two components $\boldsymbol{P}_t^{(1)}, \boldsymbol{P}_t^{(2)} \in \mathrm{I\!R}^{3\times 3}$

Let the probability of visibility w_1, w_2 of these two points be known. There are four combinations of visibility for these two points: both visible, neither visible, only the first visible and only the second visible, with probabilities $w_1 w_2$, $(1 - w_1)(1 - w_2)$, $w_1(1 - w_2)$, and $(1 - w_1)w_2$ respectively. The corresponding *a posteriori* covariance matrices for each of these two cases are $\mathrm{diag}(\boldsymbol{P}_t^{(1)}, \boldsymbol{P}_t^{(2)})$[1], $\mathrm{diag}(\boldsymbol{P}_{t-1}^{(1)}, \boldsymbol{P}_{t-1}^{(2)})$, $\mathrm{diag}(\boldsymbol{P}_t^{(1)}, \boldsymbol{P}_{t-1}^{(2)})$, and $\mathrm{diag}(\boldsymbol{P}_{t-1}^{(1)}, \boldsymbol{P}_t^{(2)})$.

Since the determinant of a block diagonal matrix is the product of the determinants of the component matrices, eq. (6) can be extended to cover four cases of visibility:

$$\begin{aligned}
\boldsymbol{a}_t^* &= \underset{\boldsymbol{a}_t}{\mathrm{argmin}}\ w_1 w_2 \log(|\boldsymbol{P}_t^{(1)}| \cdot |\boldsymbol{P}_t^{(2)}|) \\
&\quad + w_1(1 - w_2)\log(|\boldsymbol{P}_t^{(1)}| \cdot |\boldsymbol{P}_{t-1}^{(2)}|) \\
&\quad + (1 - w_1)w_2 \log(|\boldsymbol{P}_{t-1}^{(1)}| \cdot |\boldsymbol{P}_t^{(2)}|) \\
&\quad + (1 - w_1)(1 - w_2)\log(|\boldsymbol{P}_{t-1}^{(1)}| \cdot |\boldsymbol{P}_{t-1}^{(2)}|) \\
&= \underset{\boldsymbol{a}_t}{\mathrm{argmin}} \sum_{i \in \{1,2\}} w_i \log(|\boldsymbol{P}_t^{(i)}|) + (1 - w_i)\log(|\boldsymbol{P}_{t-1}^{(i)}|) \\
&= \underset{\boldsymbol{a}_t}{\mathrm{argmin}} \sum_{i \in \{1,2\}} w_i \phi_D(\boldsymbol{P}_t^{(i)}) + (1 - w_i)\phi_D(\boldsymbol{P}_{t-1}^{(i)})
\end{aligned}$$

[1] $\mathrm{diag}(\boldsymbol{A}, \boldsymbol{B})$ with $\boldsymbol{A}, \boldsymbol{B} \in \mathrm{I\!R}^{3\times 3}$ is defined as $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{pmatrix}$

thereby treating each 3-D point independently. This can be extended to $i > 2$ by induction and is valid for the three criteria tested in this paper. The optimaziation equation is therefore

$$\boldsymbol{a}_t^* = \operatorname*{argmin}_{\boldsymbol{a}_t} \sum_{i=1}^{n} w_i \phi_c(\boldsymbol{P}_t^{(i)}) + (1 - w_i)\phi_c(\boldsymbol{P}_{t-1}^{(i)}), \quad (10)$$

where ϕ_c denotes the criteria, i. e. $c = \{D, E, T\}$-criteria, and n denotes the number of points.

The optimization problem (10) is solved by an exhaustive search over the discretely sampled action space.

The E-optimality criterion is based on the eigenvalues of $\boldsymbol{P}$. For $\boldsymbol{P} \in \mathbb{R}^{3\times3}$, the covariance matrix describes a standard deviation ellipsoid with its eigenvectors as main axes and its eigenvalues as the axis lengths. Minimizing the eigenvalues corresponds to reducing the expected state estimation error along an axis. For n points the sum of these expected errors should be minimized. Let $e_{\max}(\boldsymbol{P}_t^{(i)})$ denote the largest eigenvalue of $\boldsymbol{P}_t^{(i)}$. Thus, the E-optimality criterion ϕ_E is:

$$\phi_E(\boldsymbol{P}_t^{(i)}) := e_{\max}(\boldsymbol{P}_t^{(i)}) \quad (11)$$

The optimal action is then found as in equation (10).

Lastly, we consider the T-optimality criterion, based on the trace of $\boldsymbol{P}_t$, which is equivalent to the sum of the trace of each sub-covariance $\boldsymbol{P}_t^{(i)}$. This criterion ignores the off-diagonal elements of $\boldsymbol{P}_t$, the actual co-variance terms, and focuses on the variance terms found on the diagonal:

$$\phi_T(\boldsymbol{P}_t^{(i)}) := \operatorname{trace}(\boldsymbol{P}_t^{(i)}) \quad (12)$$

The optimal action is determined as in equation (10), again.

4. Experiments

In this section we present the experimental results for the different optimization criteria. We first describe the experimental setup and the test objects. After that, we discuss the experimental results.

4.1. Test Objects and Setup

Since we do not model self occlusion, we restrict ourselves to either planar objects or polyhedral concave objects. The first test object is a simulated plane, which looks similar to a calibration pattern: its points are arranged in a square with 10 by 10 points. The 2-D observations are calculated by perspective projection of the simulated 3-D points to the image plane of a virtual camera and adding a Gaussian noise term. These are the observations for the Kalman filter.

The second object is a mouse pad (cf. Figure 2a) and the third is a “L”-shaped one (cf. Figure 2c). The 2-D observations are generated by a point tracker [13, 16]. Since the optimal view points are quite far from each other, the point tracker would provide very bad results, resulting in missing point correspondences. Therefore, we also use the images during the camera movement to track the feature points. But for 3-D reconstruction, we use only the images of the determined optimal view points.

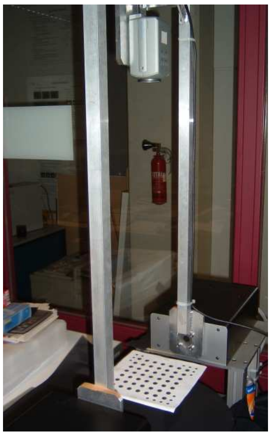
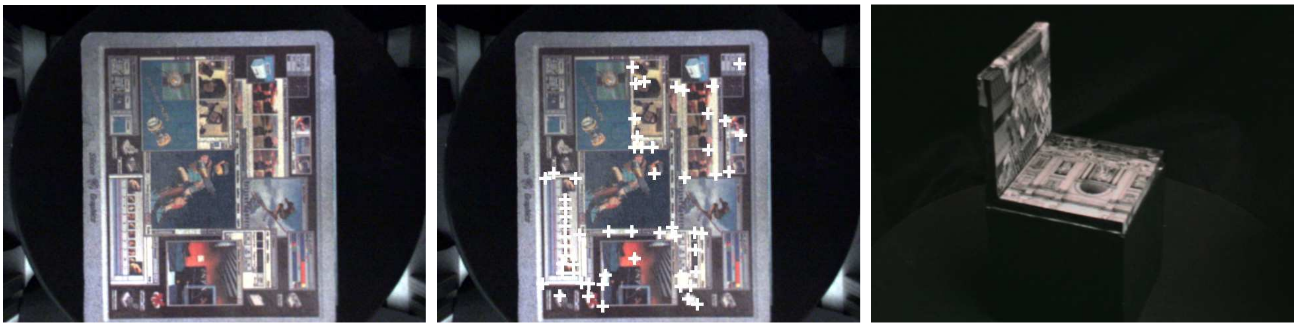


Figure 1: Turntable with calibration pattern

The camera movement is supplied by a turn table with tilting arm (cf. Figure 1). This allows the camera to be placed on the upper half of a sphere. The hemisphere is sampled discretely, saving an image of each viewpoint, so the experiments are done “offline”. Each object is positioned on the turn table so that its center of gravity is approximately the center of the hemisphere. In the simulation, we use similar assumptions: the virtual camera is also placed on a hemisphere over the object. In these cases, the camera action $\boldsymbol{a}_t$ consists of the two angles (table angle and angle of the tilting arm).

In the simulation, error measurement can be done directly: the error is the Euclidian distance of each reconstructed point to its corresponding ground truth. For the mouse pad, we have neither ground truth data, nor do we know distances between the points, since the tracked points are not regular on the mouse pad (cf. Figure 2b). So the only property which we know is that the points lie on a plane. Therefore, we calculate a regression plane to the reconstructed points and the reconstruction quality is measured by the root mean square error of the regression problem. For the “L”-shaped object, two planes are fitted to the 3-D points and the RMS is measured.

In addition to the reconstruction accuracy, the reconstruction time is also measured. We measured the time necessary for the complete reconstruction process, including triangulation to get the starting estimate, reconstruction by the Kalman filter, point tracking (if necessary), and planning the next view (approx. 3950 possible view positions are evaluated for the optimization) with the specific opti-



(a) Mouse pad

(b) Mouse pad with tracked feature points

(c) L-shaped object

Figure 2: The test objects mouse pad (left) and L-shaped object(right); middle: mouse pad with tracked feature points

mality criterion.

We show the benefit of view planning by comparing the errors to regular sampled view points (the arm angle is constantly 45° and the table is rotated by 36° between every step) and to the mean of three sequences of randomly chosen ones.

4.2. Experimental Results

For the simulation, we get the results presented in Figure 4a. We can observe that there is no significant difference between the three criteria in reconstruction accuracy, although the view points which are determined as optimal are different (cf. Figure 3). But the error decreases faster in each time step compared to regular sampling or randomly chosen view points. There is a difference for computation time: the fastest one is the T-criterion with 5:08 minutes, then the D-criterion with 5:27 minutes and the slowest is the E-criterion, which took 7:12 minutes. Of course, the methods without planning are faster: the regularly sampled case took 33 seconds and the random one approx. 35 sec.

The mouse pad results are presented in Figure 4b for 60 tracked feature points. We observe no single method consistantly outperforming the others. For each of the three criteria, there is an iteration where it is better than the other ones. But all criteria give a better result than regular sampling or a random path, from the second iteration step on.

The comparison of the computation times gives: D-criteria 1:07:38h (511 frames tracked), E-criteria 1:10:47h (527 frames), T-criteria 1:06:22h (512 frames). The drastically higher computation times in comparison to the simulation can be explained by the time for tracking, which is not necessary in the simulation. Again, the computational time for non-optimized view points is lower: it took 9:43 min (134 frames) if regularly sampled and 28:51 min (472 frames) in the random case.

Though the "L"-shaped object is not flat, the results are similar to the ones of the other experiments (cf. Figure 4c). For computation times we get: D-criteria 40:22 min (376 frames), E-criteria 42:19 min (323 frames), T-critera 40:14 min (304 frames), 17:25 min (130 frames) for regularly sampled and an average of 35:03 min for random paths (274 frames). The lower computation time compared to the mouse pad can be explained by the quite small number of valid view points, because the camera may only be positioned in front of the object.

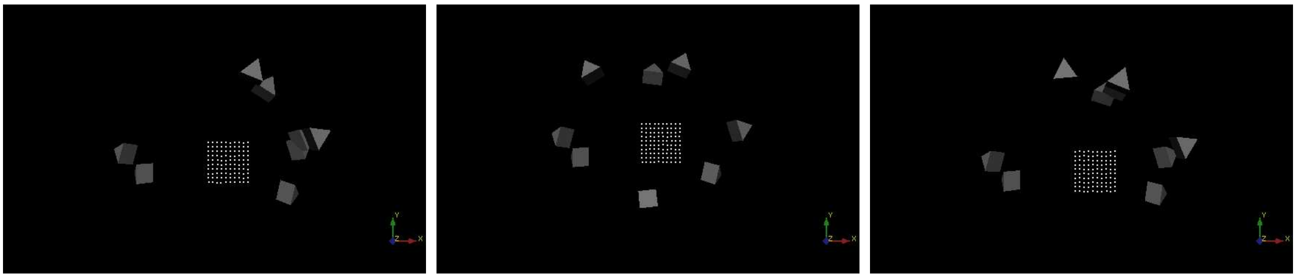
4.3. Discussion

The comparison of the three optimality criteria shows that there is no significant difference in the reconstruction accuracy, although the view positions determined as optimal are different (cf. Figure 3). So we conclude that the choice of the criterion is less important for these tested objects. But with all tested criterias the accuracy is higher than in unplanned views.

Regarding the computation time, we see that in the simulation the T-criterion is the fastest, followed by the D-criterion, and the E-criterion is slowest. This behavior can be easily explained: the computational complexity for calculating the trace is lower than for determinants, and this is again lower than for eigenvalues. But in the real world experiments, the computation time is similar for all criteria. This is because tracking is also time consuming, and so the different computational complexity of the various criteria becomes less important. Furthermore, the time for regular sampling in the real world experiments is shortest, since there is no large tracking between selected view points.

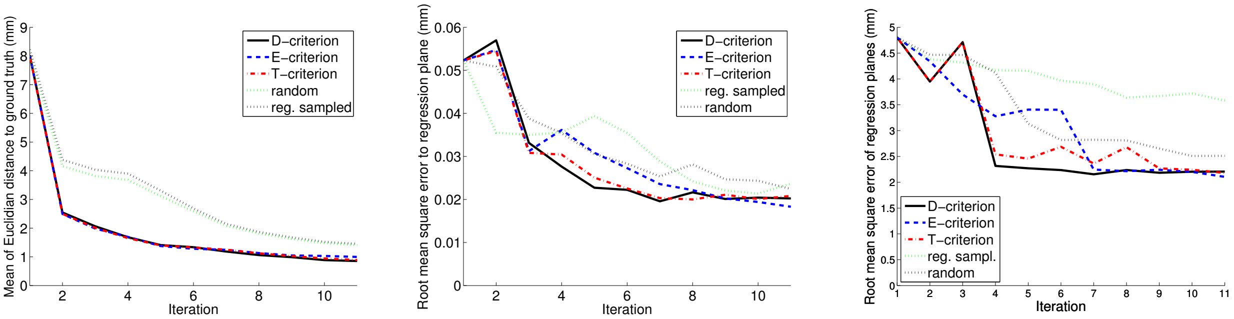
5. Conclusion and Future Work

We have presented an approach for planning the next best view for 3-D reconstruction. For this problem, we formulated a probabilistic approach based on the Kalman filter. The Kalman filter allows us to predict the a posteriori covariance matrix. Since the covariance is a measure for the uncertainty, the basic idea is to reduce the uncertainty, using



(a) Viewpoints selected by the D-criterion (b) Viewpoints selected by the E-criterion (c) Viewpoints selected by the T-criterion

Figure 3: The points of the simulated plane and the view points, which are shown as pyramids.



(a) Simulation (b) Mouse pad (c) L-shaped object

Figure 4: RMS of the reconstruction of the simulated plane (left), of the mouse pad (middle) of the L-shaped object (right)

a metric on this matrix. Several metrics exist in literature; the most common are the D-, E-, and T-criterion.

We developed an adaptation of the D-criterion for the task at hand. We also tested all three criteria for our application of planning next best views in 3-D reconstruction. A comparison shows that the reconstruction accuracy does not significantly depend on the chosen criteria, although the determined view points are not equal if the criterion is switched. The computation time was also similar in the real world experiments, although the computation of each metric has different computational complexity, as can be observed in the simulation. We also showed that next best view planning improves the reconstruction accuracy over unplanned views, even in best case scenarios.

These results are preliminary; we are already working on the problem of self occulsions, which directly affect estimating the probability of a point's visibility. Without handling this constraint, our approach cannot reconstruct complex objects optimally. The calculation of self occlusions requires an object surface to be estimated, in order to determine where points may be occluded by other surface parts.

Though the approach remains the same if there are self occlusions, we expect that the optimality criteria give different results. In the experiments done in this paper, the probability that a point is outside the field of view is quite low, due to the setup. But if we consider self occlusions, then each point may be occluded in a large number of viewpoints. Then the weighting has different effects, because trace reduction is at a different ratio than eigenvalue reduction, for example. So the weighted sum for the optimality measurement may be minimal for other view points.

References

[1] F. Chaumette, S. Boukir, P. Bothemy, and D. Juvin. Optimal estimation of 3d structures using visual servoing. In *Proc. Computer Vision and Pattern Recognition Conf.*, pages 347–354, Seattle, USA, June 1994. 1

[2] F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin. Structure from controlled motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(5):492–504, 1996. 1

[3] A. Davison. Active search for real-time vision. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 1, pages 66–73, Bombay, 2005. 1

[4] J. Denzler and C. Brown. An information theoretic approach to optimal sensor data selection for state estimation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(2):145–157, 2002. 1

[5] R. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basis Engineering*, 82:35–44, 1960. 2

[6] K. Kutulakos and S. Seitz. A theory of shape by space carving. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 1, pages 307–314, Corfu, Greece, 1999. IEEE Computer Society Press. 1

[7] K. N. Kutulakos and C. R. Dyer. Global surface reconstruction by purposive control of observer motion. In *Proc. Computer Vision and Pattern Recognition Conf.*, pages 331–338, 1994. 1

[8] K. N. Kutulakos and C. R. Dyer. Recovering shape by purposive viewpoint adjustment. *Int. J. of Computer Vision*, 12(2):113–136, 1994. 1

[9] E. Marchand and F. Chaumette. Active vision for complete scene reconstruction and exploration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(1):65–72, Januar 1999. 1

[10] E. Marchand and F. Chaumette. An autonumous active vision system for complete and accurate 3d scene reconstruction. *International Journal of Computer Vision*, 32(3):171–194, 1999. 1

[11] W. Niem and M. Steinmetz. Camera viewpoint control for the automatic reconstruction of 3d objects. In *Proceedings of the International Conference on Image Processing*, volume 3, pages 655 – 658, Los Alamitos, Calif., Sep. 1996. IEEE Signal Processing Society. 1

[12] F. Pukelsheim. *Optimal Design of Experiments*. Wiley Series in Probability and Mathematical Statistics. John Whiley & Sons Inc., New York, Chichester, Brisbane, Toronto, Singapur, 1993. 2

[13] C. Tomasi and T. Kanade. Detection and Tracking of Point Features. Technical Report CMU-CS-91-132, Carnegie Mellon University, 1991. 4

[14] G. Welch and G. Bishop. An Introduction to the Kalman Filter. Technical report, University of North Carolina at Chapel Hill, Department of Computer Science, 2006. 2

[15] 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 *Proc. International Conference on Pattern Recognition*, volume 1 of *IEEE Computer Society Press*, pages 103–106, Los Alamitos, California, Washington, Tokyo, 2006. IEEE Computer Society. 1

[16] T. Zinßer, C. Gräßl, and H. Niemann. Efficient feature tracking for long video sequences. In *Pattern Recognition, 26th DAGM Symposium*, volume 3175 of *Lecture Notes in Computer Science*, pages 326–333. Springer-Verlag, Berlin, 2004. 4