Optimal Large View Visual Servoing with Sets of SIFT Features

Thomas Nierobisch, Johannes Krettek, Umar Khan and Frank Hoffmann

Abstract— This paper presents a novel approach to large view visual servoing in the context of object manipulation. In many scenarios the features extracted in the reference pose are only perceivable across a limited region of the work space. The limited visibility of features necessitates the introduction of additional intermediate reference views of the object and requires path planning in view space. In our scheme the visual control is based on decoupled moments of SIFT-features, which are generic in the sense that the control operates with a dynamic set of feature correspondences rather than a static set of individual features. The additional freedom of dynamic feature sets enables flexible path planning in the image space and online selection of optimal reference views during servoing to the goal view. The time to convergence to the goal view is estimated by a neural network based on the residual feature error and the quality of the SIFT feature distribution. The transition among reference views occurs on the basis of this estimated cost which is evaluated online based on the current set of visible features. The dynamic switching scheme achieves robust and nearly time-optimal convergence of the visual control across the entire task space. The effectiveness and robustness of the scheme is confirmed in an evaluation in a virtual reality simulation and on a real robot arm with a eye-in-hand configuration.

I. INTRODUCTION

Vision is expected to play a progressively more important role in service robotic applications in particular in the context of manipulation of daily life objects. Image based visual servoing solely relies on 2D image information for the alignment of the end-effector with an object of unknown pose [1]. The desired pose for grasping is demonstrated to the robot and a set of reference features is extracted from the image. Subsequent approaches of the robot and camera to the goal pose are accomplished by regulating the image error between the current and reference features without explicit geometric reconstruction of the object pose.

Optimal motion control for visual servoing to a static reference view has been discussed in [2], [3] and is based on the decoupling of the translational and rotational degrees of freedom achieved by a partial pose estimate using either the homography or the fundamental constraint. Both approaches require the online estimation of the homography or fundamental matrix in the servo loop, with at least four or eight feature point correspondences. The method in [2], [3] focuses on the optimal control with respect to a fixed set of features, whereas our approach addresses the issue of large view visual servoing with extraction and matching of dynamic sets of SIFT features. The view space is partitioned by an entire set of intermediate, partially overlapping reference views of the object. The authors in [4] integrate a path planner in the image space with a visual controller based on potential fields in order to obtain visual navigation for large displacements. The work in [5] extends these concepts to qualitative visual servoing based on objective functions that capture the progression along the path, the feature visibility and camera orientation. This paper provides a contribution to optimal path planning in the image space considering the residual feature error in conjunction with the quality of the feature distributions in alternative reference views. The additional flexibility of dynamic feature sets provides the basis for opportunistic online switching among reference views while navigating towards the goal view. [6] describes a method for automatic selection of optimal image features for visual servoing in terms of robustness, uniqueness and completeness. Additional performance criteria concern the systems observability, controllability and sensitivity. Our visual features are generic moments computed over a dynamic set of point features. Maximum robustness and observability of the statistical moments is achieved by aggregation over all available and matched features. Our objective is to estimate the quality of feature sets from alternative reference views online based on similar criteria as [6]. Optimal reference view selection relies on an estimate of the time to convergence of residual errors. This estimate is provided by a neural network that is trained with feature errors and distributions as input to predict the time to convergence. The proposed scheme is model-free in so far as it does not depend on a geometric object model or reconstruction of the object or camera pose, which means that navigation and control is entirely performed in the image space.

The paper is organized as follows: Section II provides the definition of decoupled visual features based on weighted moments of SIFT features used for visual servoing in 6 DOF. Due to the limited visibility of SIFT features across different views it is necessary to introduce intermediate reference views. The time optimal reference selection to accomplish large view visual servoing is introduced in III as well as the navigation in the image space. Section IV demonstrates the experimental results on a sphere and a semi cylinder setup and analyzes the convergence behavior of alternative switching strategies. The paper concludes with a summary and outlook on future work in section V.

II. VISUAL SERVOING WITH SIFT FEATURES

Scale invariant feature transformations (SIFT) introduced by Lowe [7] occur frequently in textured objects and are invariant to changes in scale, orientation, illumination and affine transformations. They are uniquely identifiable and

T. Nierobisch, J. Krettek, U. Khan and F. Hoffmann are with the Faculty of Electrical Engineering and Information Technology, Chair for Control and Systems Engineering, Universität Dortmund, 44221 Dortmund, Germany
thomas.nierobisch@uni-dortmund.de

robustly matched across different views of the same object. These properties render them particularly suitable for model free image based visual servoing. Our scheme is motivated by the work of [8], which relies on image moments rather than points features to overcome the shortcomings of visual servoing schemes. SIFT features for visual servoing applications were first introduced by [9], in which the authors focus on the robust feature selection and explicitly reconstruct the object pose based on the epipolar geometry. A novel image-based controller that augments conventional point features by the additional attributes scale and keypoint orientation of SIFT features is presented in [10]. This work is improved by establishing a one-to-one correspondence between feature and camera motion based on weighted moments that eliminates or at least minimizes the undesired couplings [11]. A set of reference SIFT features is automatically extracted from an image of the object captured in the demonstrated reference pose. The automatic feature selection detailed in [10] identifies a subset of robust and non-ambiguous features for the ultimate visual servoing of the robot to the reference pose. The visual servoing in 6 DOF relies on six associated statistical moments computed over the location, scale and orientation of the SIFT features. Notice, that the term feature has a dual meaning as it refers to the point like SIFT features as well as the visual features aggregated over a set of matched SIFT features subject to control. A single SIFT-feature F_i contains four attributes, namely its pixel coordinates u_i and v_i, its canonical orientation ϕ_i and its scale σ_i. The scale changes inversely proportional to the distance between the camera and the object. The orientation ϕ_i is consistent with the camera rotation about its optical axis. Scale and orientation are ideal for the control of the distance to the object and the rotation around the optical axis as they prove widely independent of translations and rotations along the other axes. The rotation and translation along the cameras optical axis are captured by the moments

$$f_\gamma = \frac{\sum_{i=1}^n \phi_i}{n}, \quad f_z = \frac{\sum_{i=1}^n \sigma_i}{n} \qquad (1)$$

which correspond to the mean orientation and scale of detected SIFT-features. The translations along x- and y-axis are regulated with respect to the geometric centroid of the point features. The geometric centroid is susceptible to translations along the z-axis causing an undesired coupling with this motion. The decoupling of the centroid from the z-motion is achieved by dynamically weighting the individual point features. The moments

$$f_x = \sum_{i=1}^n w_i u_i, \quad f_y = \sum_{i=1}^n w_i v_i \qquad (2)$$

correspond to the weighted mean of the matched SIFT feature locations and capture translations along the x- and y-axis respectively [11]. The 6 DOF visual control is completed by the moments f_α and f_β associated with rotations along the x- and y- axis.

$$f_\alpha = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{(-\hat{v}_i - \hat{v}_j) \cdot \left\| \vec{p}_j - \vec{p}_i \right\|}{\sum_{k=1}^n \sum_{l=k+1}^n \left\| \vec{p}_k - \vec{p}_l \right\|} \qquad (3)$$

The moments f_α and f_β detect the perspective distortions of lines connecting pairs of SIFT features caused by rotations. The term $\left\| \vec{p}_j - \vec{p}_i \right\|$ denotes the length of the line connecting the two pixels. This length is multiplied by the weight factor $(-\hat{v}_i - \hat{v}_j)$. Its sign indicates whether the line is above or below the u-scan-line through the cameras principal point. The absolute magnitude of the weight increases with the vertical distance from the image center. The moment f_β is defined in an analog manner to f_α, by interchanging the u and v components.

III. TIME-OPTIMAL REFERENCE SELECTION

For large view visual servoing intermediate views are defined to navigate across the entire view hemisphere. It becomes desirable to switch between intermediate views in a stable, robust and time-optimal manner. The cost in terms of number of control cycles to converge from the current view to the reference image is estimated in order to compute the optimal path. Crucial for this purpose is the proper definition of performance criteria for approximation of the cost function and the analysis of their correlation with the cost. In our case, an artificial neural network learns the relationship between the control criteria and the costs in a supervised manner. The training data is obtained from observations of the actual number of control cycles required for transitions between neighboring reference views.

A. Performance criteria

1) Feature error: The overall feature error

$$f(I) = \{\Delta f_x, \, \Delta f_y, \, \Delta f_z, \, \Delta f_\alpha, \, \Delta f_\beta, \, \Delta f_\gamma\}$$

constitutes the most significant performance criterion for the estimation of the cost. A single feature error alone does not provide a good estimate of cost, because the actual time until convergence depends on the feature error with the slowest task space motion, usually associated with the translational degrees of freedom. The rotational errors are bounded by the visibility constraint and are usually stabilized within a few control steps. Each element of $f(I)$ is normalized to the interval $[0,1]$ according to its maximum range. The total feature error is the sum of normalized errors.

$$\hat{f}(I) = \sum_{i=1}^6 \left| \hat{f}_i(I) \right| \qquad (4)$$

The feature error already attributes to a substantial amount of variation in the cost, nevertheless the cost estimate is improved by inclusion of additional criteria that capture the quality and robustness of visual control.

2) Number of features: The robustness and the control performance increase significantly if more than the minimal number of correspondences is established. The redundancy of multiple features reduces the noise level and contributes to the beneficial widespread dispersion of features in the image space. A small number of features might cause a compact distribution of point features, which might cause poor or even unstable control in the image space. The number of matched features also provides an estimate of the geometric distance

of the current view to the reference pose. Distant poses only share a subset of mutually visible features, whereas the number of correspondences naturally increases with the proximity of both viewpoints. The criterion $C(I) = n$ is defined as the number of feature correspondences between the current and the reference view. The number alone is not a unique indicator of the expected cost as it also depends on the distribution of these features defined in terms of their entropy and variance around the centroid.

3) Entropy: Entropy measures the order or disorder in a distribution. The image is partitioned into $N = 10$ vertical and horizontal equally spaced columns and rows. The entropy along the two axes is calculated as

$$E_{u,v}(I) = -\sum_{i=1}^{N} H_{u,v}(i) \cdot \log_N (H_{u,v}(i)) \qquad (5)$$

in which $H_{u,v}(i)$ denotes the relative frequency of SIFT features in the ith column respectively row. Whereas high entropy indicates a uniform distribution, a low entropy reveals an inhomogeneous distribution, which harms the robustness and speed of convergence of visual servoing.

4) Centroid location: The visual features f_α and f_β benefit from a distribution uniformly centered around the principal point in order to capture the distortion of line segments. The deviation of the feature centroid from the origin is expressed by

$$|\bar{u}| = \sum_{i=1}^{n} \left| \frac{u_i - u_0}{n} \right| \quad |\bar{v}| = \sum_{i=1}^{n} \left| \frac{v_i - v_0}{n} \right| \qquad (6)$$

in which low values represent desirable feature distributions.

5) Variance of the feature distribution: The variance of the feature positions provides an additional estimate of the quality of the feature distribution. A low variance in particular in conjunction with a dislocated centroid reflects a feature distribution that is suboptimal for visual control and delays the convergence to the reference image. The variances are computed as

$$\sigma_u = \sum_{i=1}^{n} \frac{(u_i - \bar{u})^2}{n}, \quad \sigma_v = \sum_{i=1}^{n} \frac{(v_i - \bar{v})^2}{n} \qquad (7)$$

Notice, that entropy reflects the geometric homogeneity of the feature set, whereas variance captures its width.

B. Correlation between criteria and time to convergence

Control experiments from 150 initial positions randomly distributed over the task space are recorded in order to evaluate the correlation between the performance indicators and the time to convergence. Each control step of an individual run constitutes a training sample for supervised learning of the neural network. A control run is considered as successfully converged to the reference image if all feature errors are reduced to within 10% of their average initial value. The correlation between the performance criteria and the actual time to convergence provides insight into the influence and relevance of the individual indicators. The linear dependency

between two stochastic variables is computed according to Pearson's correlation coefficient:

$$r_{XY} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}, \qquad (8)$$

which assumes values in the interval $[-1, 1]$. Large absolute values indicate strong correlation between the two quantities. Table I specifies the correlations between the performance indicators and the cost in terms of time to convergence.

TABLE I

PEARSON CORRELATION BETWEEN PERFORMANCE FEATURE ERROR RESP. CRITERIA AND TIME TO CONVERGENCE

	Δf_x	Δf_y	Δf_z	Δf_α	Δf_β	Δf_γ	$\hat{f}$	
r_{XY}	0,30	0,14	0,17	0,14	0,13	0,13	0,63	
	$C(I)$	$C_n(I)$	$E_u(I)$	$E_v(I)$	$|\bar{u}|$	$|\bar{v}|$	σ_u	σ_v
r_{XY}	-0,66	-0,72	-0,66	-0,72	0,44	0,32	-0,64	-0,62

The individual feature errors are only slightly correlated with the cost, whereas the normalized summed feature error $\hat{f}$ is indeed a proper indicator of the distance to the reference pose. Notice, that the number of matched features $C_n(I)$ correlates even more with the cost than the summed absolute errors $\hat{f}$. The scalar summed error contains less information than the entire error vector $f(I)$. This is explicable, as the feature errors related to the translational degrees of freedom converge at a slower rate.

In order to predict the time to convergence two neural networks with different input features are trained with the data acquired during the 150 experimental runs. The multi-layer perceptrons are composed of 16 neurons in the hidden layer and are trained with the standard back-propagation algorithm. The first network only uses the six-dimensional feature error $f(I)$ as input, whereas the second network in addition has considers the performance criteria $c(I) = \{C_n(I), E_u(I), E_v(i), \bar{u}, \bar{v}, \sigma_u, \sigma_v\}$. Figure 1 depicts the relationship between the estimated costs on the x-axis and the true costs for the full input network. It also shows the linear regression for the partially and fully informed network. The neural network only trained with the feature error $f(I)$ achieves a correlation between estimated and true cost of 0.75. This correlation is substantially improved by incorporation of the additional performance criteria to a degree of 0.96. The improvement in prediction accuracy of the fully informed network error compared to the pure feature error based network is confirmed by the reduced training and test set error shown in table III-B. This demonstrates that a distance metric in the image space to the goal view, has a significantly lower correlation with the costs than $f(I)$ in conjunction with the image distribution indicators $C(I)$. This observation confirms that the distribution of the SIFT-feature crucially effects the control performance.
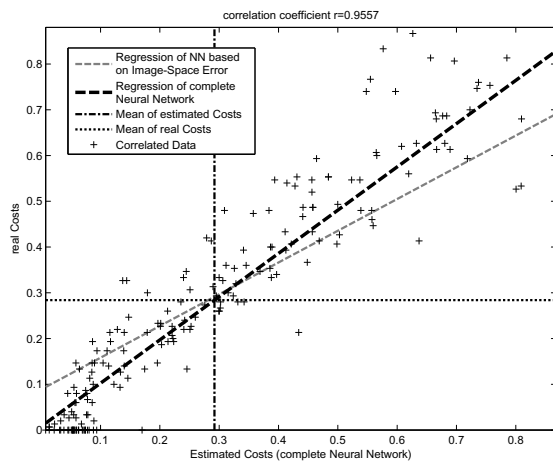
Fig. 1. Neural network estimate versus true cost (+) and regression line for the neural network with $f(I)$ as input and $(f(I), c(I))$ as input.

TABLE II

TRAINING AND TEST SET ERROR FOR NEURAL NETWORK TRAINED WITH $f(I)$ ONLY AND WITH $f(I), c(I)$

	RMSE train	RMSE test	correlation
$f(I)$	0.0149	0.0297	0.75
$f(I), c(I)$	0.0072	0.0092	0.96

C. Navigation in the image space

Our approach neither requires a geometric model of the object nor is it aware of the spatial relationship between the reference views. Instead the optimal path is planned online in the image-space rather than in the task space. For that purpose each reference-view (RV) represents a node in an undirected graph, in which edges define neighborhood relationships between overlapping views. The cost of an edge connecting two views reflects the transition time between the views expressed in terms of number of iterations to converge from the initial view to the neighboring view. The graph supports the global initial path planning from the start view to the desired goal view, but it also forms the basis for the decision when to switch to the next reference view. The cost estimation within the path planning consists of two major steps, an off-line computation of graph costs between the reference view and and an online computation of the cost from the current view to the overlapping reference views. The planner switches between reference view based on a comparison of the accumulated costs of currently feasible reference views.

1) Initial path-planning and cost-estimation: The initial cost estimation is based upon the graph constructed from the complete set of reference-views which form its nodes. The number of matching features is computed for every possible pair of reference views. An edge is generated between two overlapping views if they share five or more common features. The cost of an edge is estimated by evaluating the set of corresponding features with the neural network described in the previous section. The optimal path from
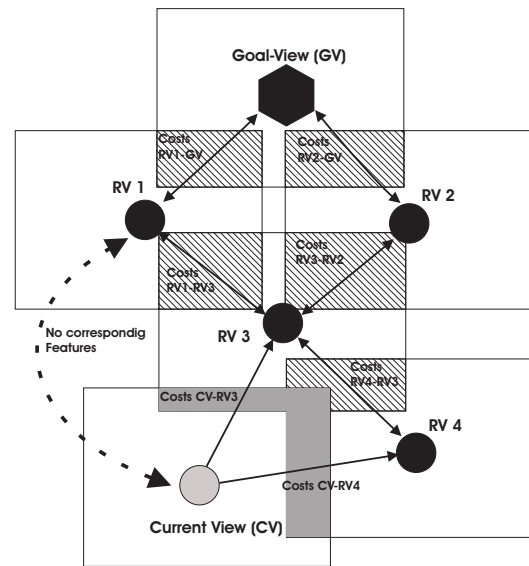

Fig. 2. Reference-, Goal- and Current View represented by a Graph

every reference-view to the goal-view is calculated with the well-known Dijkstra-Algorithm [12] for finding the shortest path in a weighted graph. This calculation is part of the teach-in-process in which reference views are captured across the work space and is performed off-line in advance.

2) Current cost estimation and selection of optimal reference-view: The features extracted from the current view (CV) are continuously compared to those of overlapping reference views in order to identify the optimal current reference view online during control. The time to convergence is estimated for the feasible reference views in the same way as for the initial generation of the graph. The total costs for reaching a specific reference view plus the already estimated cost for the shortest path from that node to the goal view are compared among all feasible views. The node with minimal cost is selected as the next reference view to be included into the shortest path to the goal. The view evaluation is only performed every fifth control-cycle in order to reduce the amount of online computations.

Fig. 2 depicts a section of a graph generated from a set of images with four intermediate reference views $RV_1, \ldots, RV_4$, a goal view GV and the current view CV. The images associated with a view are diagrammed by rectangles, the hatched areas represent the overlap between neighboring images which contain common SIFT features. The cost of the transition from the current view to the two feasible reference views RV_3 and RV_4 depends on the number and quality of common features in the gray areas. The current view has no connection to the reference views RV_1 and RV_2 as the subset of common features is empty, as indicated by the dotted line. A hysteresis in the switching scheme avoids the risk of the visual controller getting trapped in a limit cycle around the optimal switching point due to uncertainties in the cost estimate or fluctuations in the matched features. The initially estimated costs of the optimal path from the current

view to the goal are weighted by the number of intermediate nodes from the candidate reference views to the goal node. That way, switching to a reference view which node is closer to the goal node becomes more attractive, whereas the reverse switching to a more distant node is suppressed even if its estimated cost seems more attractive. A transition to a lower cost reference view is initiated if its superiority is confirmed in two consecutive iterations, thereby gaining additional robustness with respect to cyclic switching.

IV. EXPERIMENTAL RESULTS

This section presents an evaluation of the proposed scheme in visual servoing experiments within a virtual reality environment and on a real 5 DOF robotic arm with an eye-in-hand configuration. In both experimental setups the performance of the cost estimation based switching scheme is compared with two alternative methods. The first method, in contrast to our scheme, assumes that the geometric distance in task space between reference views is known. It switches to the reference view closest to the goal pose, once the minimal number of visual features is perceived. This switching strategy ignores the perceptability and quality of the set of matched feature and is from a control point of view not sufficiently robust. Nevertheless for the purpose of comparison it provides an upper performance limit. The second method computes an optimal static path that connects the start to the goal node based on the static costs. It is not opportunistic as it does not reestimate the costs online, or replans if other reference views not originally included in the plan suddenly appear more attractive. It switches to the next view outlined in the plan upon convergence of the feature error to a current reference view. This method although suboptimal is robust from a control point of view, but is still improvable by relaxing the convergence criterion without sacrificing robustness.

A. Navigation across a sphere

A virtual-reality simulation of a free moving camera allows the verification of the large view visual servoing scheme without being constrained by the robot kinematics or workspace. The camera navigates in 6 DOF around a sphere textured with a schematic map of the globe. The reference views are equidistantly located along longitudes and latitudes. The task is to guide the camera visually from the north to the south pole. Fig. 3 depicts the distribution of reference-views together with the path pursued by the three methods under comparison. Even though the camera is initially located above the north pole, all schemes immediately transit to an initial reference view that is already closer to the goal. The distance based method picks a different great circle route than the other two schemes as it ignores the issue of feature quality. A better rationale is to select the great circle route which guarantees visibility of a sufficient number of features for a stable traverse to the south pole. We term this effect the *Pacific-problem*, as for our globe example, the equal-distant paths either moving over America or Africa contains more features due to the texture and
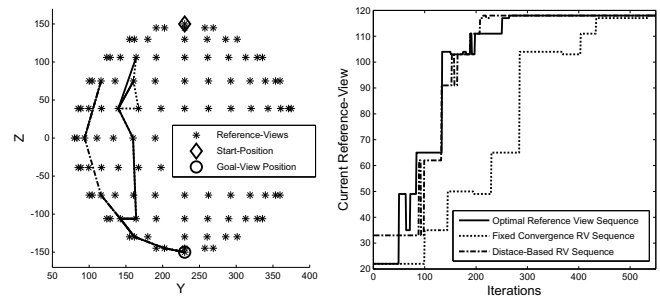

Fig. 3. Alignment of reference-views an comparison of chosen sequences from pole to pole on a sphere

text on the continents than crossing the Pacific with sparse features. The right part of fig. 3 compares the sequence
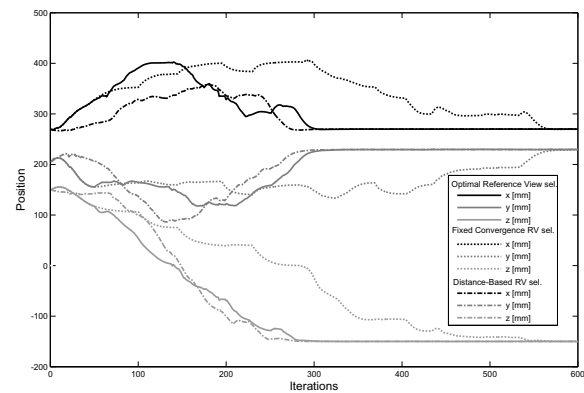

Fig. 4. Pole to pole trajectories of the compared methods

and progression of reference views followed by the three alternative methods. Fig. 4 shows the evolution of the task space error in terms of iterations. The number of iterations until convergence is approximately the same for the optimal image-based and the distance based navigation method. For the former the goal pose is reached within 300 iterations, for the later in about 290 iterations, whereas the static scheme with complete convergence takes about 560 iterations.

B. Navigation across a semi cylinder

The scheme is also evaluated in an experiment on a 5DOF Katana robot with an *eye-in-hand* camera configuration. As the workspace of the manipulator is rather limited, the camera navigates across the inner surface of a semi cylinder with a circumference of $1.8m$ and a height of $0.4m$. The inside of the semi cylinder is textured with a panoramic photo of our campus. This cylindric configuration is optimal with respect to the workspace of the robot as it allows a maximal number of sufficiently distinct reference views. The reference views form a 15×6 grid, horizontally separated by 10 °, vertically by $5cm$. The kinematics of the specific robot limit the camera motion to 5 DOF. At the start pose the camera points at the upper left part of the image and the goal is located in the lower right corner of the cylinder. As shown in Fig. 5, all methods follow largely a similar view-sequence. The only significant deviation occurs halfway

through the path in a region which mostly contains sky and ground and therefore few distinctive features. The optimal switching scheme takes a small vertical detour in order to exploit the higher concentration of features in the textured band between sky and ground. The number of iterations
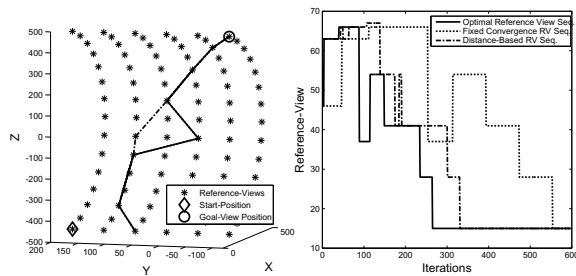

Fig. 5.   Alignment of reference-views and the chosen sequences

until final convergence is about 300 for the optimal method, 400 for the distance-based approach and 600 for the fixed-convergence-method. The difference in time to convergence results from the fact, that the two other methods require a much longer time to traverse the region of sparse features as the visual control fluctuates to the poorer quality of feature distributions. This observation is confirmed by an
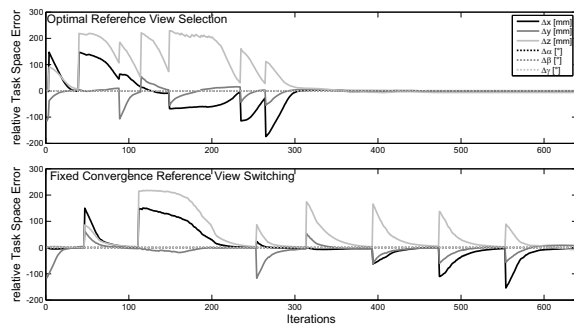

Fig. 6.   Relative task-space error for fixed-convergence and optimal method

analysis of the evolution of the relative task space error with respect to the intermediate reference views shown in fig. 6. The upper graph depicts the progression of task space error and switching sequence for the proposed scheme the lower graph for the static scheme. The static scheme wastes iterations in phases at which the feature error is already low but not yet fully converged. The optimal cost based scheme avoids delayed transition to the next reference view, as it already switches for substantially larger residual errors without compromising the stability of the control. The sample rate of the visual control loop is approximately 4 Hz, limited by the computational effort for feature extraction (160 ms), online path planning and matching (70 ms) and computation of visual features and differential kinematics (30 ms). The time for feature extraction is proportional to the size of the image and the average number of detected features. The time for extraction of a single feature on a 1.8Ghz P4-System for a resolution of $320 * 240$ pixels is approximately $1ms$.

V. CONCLUSION

This paper presents a novel approach for optimal large view visual servoing based on decoupled moments of SIFT-features. The workspace is partitioned into a set of overlapping reference views in order to navigate visually from start to the goal pose. The switching between reference views occurs on the basis of the estimated time to convergence taken the quality of matched features into account. The cost of reference views is evaluated online throughout progression to the goal view, such that the scheme opportunistically selects the reference view that is optimal in the current context. The experimental results in virtual reality and on the real robot demonstrate that the approach minimizes the time to convergence without sacrificing the robustness and thereby stability of the visual control. Future work is concerned with the transfer of the large view visual servoing with SIFT features from the domain of manipulation to navigation of mobile robots in unstructured environments. In this context our research focuses on the development of an heuristic switching scheme for large visual servoing, that is independent of the object and does not require an offline exploration of the view space for prior cost estimation.

REFERENCES

[1] S. Hutchinson, G. D. Hager, I. P. Corke, *A Tutorial on Visual Servoing Control*, IEEE Transactions on Robotics and Automation, vol.12, pp.651-668, 1997.
[2] K. Deguchi, *Optimal motion control for image-based visual servoing by decoupling translation and rotation*, in IEEE Int. Conf. Intell. Rob. and Syst., vol.2, pp. 705-711, Oct. 1998.
[3] F. Chaumette, E. Malis, *2 1/2 D Visual Servoing: A Possible Solution to Improve Image-Based and Position-Based Visual Servoings*, IEEE International Conference on Robotics and Automation, 2000.
[4] Y. Mezouar and F. Chaumette, *Path planning for robust image-based control*, IEEE Trans. on Robotics and Automation, 18(4):534-549, August 2002.
[5] A. Remazeilles, F. Chaumette and P. Gros, *3D navigation based on a visual memory*, IEEE International Conference on Robotics and Automation, 2006.
[6] J. T. Feddema, C. S. G. Lee, and O. R. Mitchell, *Weighted selection of image features for resolved rate visual feedback control*, IEEE Trans. on Robotics and Automation, 7(1):31–47, February 1991.
[7] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.
[8] O. Tahri and F. Chaumette, *Point-based and region-based image moments for visual servoing of planar objects*, IEEE Transactions on Robotics and Automation, Vol. 21, No. 6, 2005.
[9] A. Shademan and F. Janabi-Sharifi, *Using scale-invariant feature points in visual servoing*, Optomechatronic Sensors, Actuators, and Control, 2004, Proceedings of the SPIE, Volume 5603, pp. 63-70 (2004).
[10] F. Hoffmann, T. Nierobisch, T. Seyffarth and G. Rudolph, *Visual Servoing with Moments of SIFT Features*, IEEE International Conference on Systems, Man, and Cybernetics, SMC 2006
[11] T. Nierobisch, J. Krettek, U. Khan and F. Hoffmann, *Weighted moments of SIFT-Features for decoupled visual servoing in 6DOF*, Advances in Cybernetic Systems 2006, United Kingdom
[12] E. W. Dijkstra, *A note on two problems in connexion with graphs*, Numerische Mathematik, pp. 269-271, 1959.