A View Direction Planning Strategy for a Multi-Camera Vision System

Tingting Xu, Kolja Kühnlenz, and Martin Buss
Institute of Automatic Control Engineering (LSR)
Technische Universität München
D-80290 Munich, Germany
Email: {tingting.xu, kolja.kuehnlenz, m.buss}@ieee.org

***Abstract*— A multi-camera view direction planning strategy for mobile robots is discussed. Two concurrent tasks are considered: self-localization and object tracking. The approach is to assign the different tasks to different cameras, such that for each task an individual optimal view direction is selected based on the information gain maximization. Thereby, the individual task performance is significantly improved. The performance of the proposed strategy is evaluated in simulations, considering a humanoid robot navigation scenario, and compared with another two coupled or partly coupled gaze control strategies.**

I. INTRODUCTION

Humans, like all animals, respond very readily to novel objects and fast changes in their environment. They can be, for example, significantly attracted by motion in real-world environments, which can help to gather task-relevant information and increase their own safety.

For the purpose of locomotion in an unknown real-world environment, a cognitive mobile robot should also have the human-like ability to react to the world. Particularly useful is an active vision head controlled by a visual attention system that selects task-relevant viewpoints in the environment. Typical tasks are, for instance, self-localization and object tracking based on object position estimation. Therefore, view direction planning has attracted a great attention in the robotics domain.

A variety of approaches for the view direction planning are proposed. A real-time EKF-based SLAM system using active vision is combined with information from odometry and a roll/pitch accelerometer sensor in [1] to perform accurate, repeatable localization while traversing an undulating course. In [2] an approach to an optimal gaze control system for autonomous vehicles is proposed in which the perceptive situation and subjective situation are also predicted. In [3] a viewpoint selection strategy based on entropy maps is introduced, based on a sequential recognition strategy in which object hypotheses are represented as conditional probability density functions. Selection of viewpoints is accomplished using an active vision approach that selects on the basis of minimizing ambiguity of recognition. For gaze control of humanoid robot the basic idea of [4] is based on maximization of the predicted visual information content of a view situation. A task decision strategy is applied to view direction selection.

Due to limited field of view and accuracy of single or stereo camera systems, many works proposed the use of multi-camera vision systems. In [5] a vision system consisting two fixed cameras for the position and orientation tracking of a moving object is proposed. In [6] collaborative multi-camera surveillance with automated person detection is used. Multiple cameras with overlapping fields of view collaborate to confirm results and to implicitly handle non-overlapping regions. In [7] the impact of camera location and multi-camera fusion with real robots in an urban search and rescue task is studied. It is found that having two cameras, one forward-facing and one rear-facing, results in improved situation awareness. In [8] a practical person-tracking system is proposed using two sets of color stereo cameras for locating people with the stereo images and for maintaining their identities with the color images. To achieve both wide field of view and high accuracy, in [9] multi-focal vision is used for measurement and robotics applications. Advantages are significant improvements of control performance and localization accuracy as well as an extension of the workspace compared to conventional approaches.

To accomplish self-localization and tracking of moving objects simultaneously, an active vision system could plan its view direction using the decision strategy mentioned in [4], namely a view direction selected at each time step only for one task. However, due to the simultaneous movements of object and robot, the object position estimation error has an impact on the robot position estimation. Vice versa, the robot position estimation error also affects the object position estimation. To achieve a satisfying accuracy for both tasks, in this paper a view direction planning strategy based on multi-camera vision is proposed. Different stereo cameras are used in order to accomplish different tasks. The view directions of two stereo cameras are controlled independently. Thereby, the individual task performance is significantly improved. The performance of the proposed strategy is evaluated in simulations and compared with another two strategies using the task decision concept mentioned in [4] considering a humanoid robot navigation scenario.

The paper is organized as follows: In Section II, the proposed view direction planning technique is presented. In Section III, the simulations and the results are shown and discussed. The performances of three different view direction planning strategies are compared. Conclusions are given in Section IV.

II. TASK DEFINITION

A. Problem Description

Self-localization is one of the most important tasks for a mobile robot. From visual information in a natural scene, the robot can choose appropriate landmarks such as point features, edges etc. for its localization computation. In a real-world environment moving objects also exist, considered as interesting regions such as humans and vehicles etc.. If the robot's attention is directed to them to acquire better accuracy for object position estimation, the robot ego-motion estimation is corrupted due to simultaneous movements of robot and objects. The object position estimation errors and the robot position estimation error have negative impacts on each other.

For simplicity, in this paper we only refer to moving objects as objects. All other static objects in the environment could be regarded as landmarks and be used for the localization task. We assume that the robot can distinguish between static and dynamic things using general motion detection algorithms.

B. View Direction Planning Strategy

A view direction planning strategy based on multi-camera vision is proposed. In our approach, different tasks are assigned to different cameras. The view direction planner is divided into several planners, according to the number of the tasks. Each planner acquires different information from the environment, processes it separately, utilizing its individual task-relevant criterion, and chooses the most appropriate view direction for its respective camera, such that each task can be accomplished optimally. For purpose of the robot self-localization and objects tracking in the environment, two view direction planners are needed (see Fig. 1).

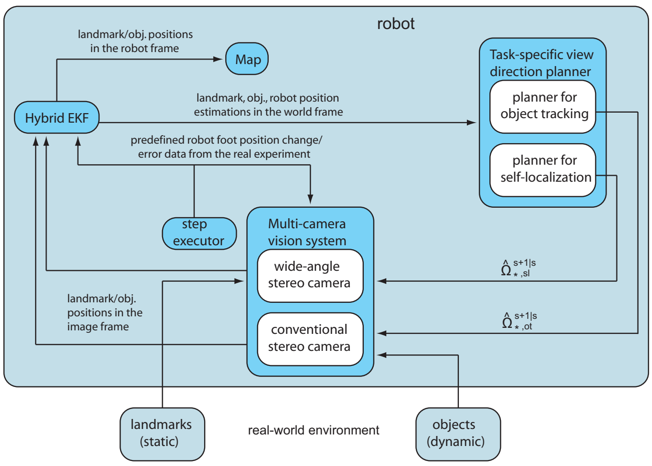

Fig. 1. Multi-focal view direction planning architecture and simulation layout

We use an information-based concept for the planners. The idea is to choose the view direction which could provide a maximum information gain.

One planner is applied for the self-localization (sl) task. The information measure v^s_{sl} for this task at step s is defined as follows:

$$v^s_{sl} = \frac{1}{2}\sum_{j=1}^{2}\sqrt{e^s_j}, \tag{1}$$

where e^s_j are the eigenvalues of the robot position covariance matrices and j is the index for x- and y-direction. We use a hybrid Extended Kalman filter to predict the robot position and calculate covariance matrices. The error covariance matrix is a measure of the estimation accuracy. For possible view directions $\hat{\Omega}$ of the respective camera the respective information measures $\hat{v}^{s+1|s}_{sl}$ will be computed. The view direction $\hat{\Omega}^{s+1|s}_{*,sl}$ with the strongest information content increase, which is defined as the covariance decrease, will be chosen as the optimal view direction by the planner and applied for the next time step $s+1$ to the respective camera.

$$\hat{\Omega}^{s+1|s}_{*,sl} = \arg\max_{\hat{\Omega}}(v^s_{sl} - \hat{v}^{s+1|s}_{sl}). \tag{2}$$

In the view direction planner for the object tracking (ot) task, the distances between the robot and the objects will be checked firstly. The eigenvalues of the position covariance matrix of the object nearest to the robot is considered as information measure v^s_{ot} for this task at step s:

$$v^s_{ot} = \sum_{j=1}^{3}\sqrt{e^s_j}, \tag{3}$$

where e^s_j are the eigenvalues of the object position covariance matrices and j is the index for x-, y- and z-direction. Analogously, a view direction with the strongest information content increase is regarded as the optimal one and should be focused by the other camera.

$$\hat{\Omega}^{s+1|s}_{*,ot} = \arg\max_{\hat{\Omega}}(v^s_{ot} - \hat{v}^{s+1|s}_{ot}). \tag{4}$$

Two view direction planners use two different information measure definitions. There is no overlapping in the computation of the view directions. Applying this strategy, the view direction planner of the self-localization task is independent of the view direction planner of the object tracking task. In other words, the position estimation of the robot ego-motion does not depend on the object movement any more, so that the accuracy of the robot ego-motion estimation will be improved. Consequently, the position estimation of the object will, respectively, also be improved.

III. PERFORMANCE EVALUATION

To evaluate the performance of the view direction planning strategy proposed in Section II, several simulations are conducted. A humanoid robot navigation scenario is considered.

A. Simulation Description

A walking humanoid robot is equipped with two active stereo cameras, providing the robot with visual information. One camera is a wide-angle stereo camera, and the other is a conventional stereo camera. The wide-angle stereo camera which is responsible for the robot self-localization task has aperture angles of 60° in each direction with focal lengths of 2mm. The conventional stereo camera corresponding to the task of object position estimation has aperture angles of 30° in each direction with focal lengths of 20mm. We assume that two stereo cameras are independently controllable such that we can plan their view directions individually. The multi-camera vision system should be controlled such that the position estimation errors of the robot and the present objects are minimized.

Our view direction planning architecture and simulation layout is shown in Fig. 1.

Based on [4] and [9], the step executor models the biped walking of a humanoid robot in the world frame. The robot frame is placed in the current supporting foot. A hybrid Extended Kalman Filter (EKF) is the central part of the whole system. An EKF is an efficient recursive filter which gives the state estimation of a dynamic system from a series of noisy measurements. Accounting for biped locomotion, a hybrid EKF is proposed in [4]. The system state $\mathbf{x}_{k+1}$ at step $k+1$ is represented as follows:

$$\begin{aligned}\mathbf{x}_{k+1} = & \begin{bmatrix} {}_0\mathbf{x}_{0_{k+1}} \\ {}_F\mathbf{x}_{0_{k+1}} \\ \vdots \end{bmatrix} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_{k+1}, \mathbf{w}_k) \\ & = \mathbf{x}_k(1-\gamma_{k+1}) + \mathbf{f}_s(\mathbf{x}_k, \mathbf{u}_{k+1}, \mathbf{w}_k)\gamma_{k+1}\end{aligned} \quad (5)$$

where $\gamma_{k+1} \in \{0,1\}$ and $\gamma = 1$ represents the robot foot changes. ${}_0\mathbf{x}_{0_{k+1}}$ and ${}_F\mathbf{x}_{0_{k+1}}$ are the state vectors in the world frame and in the robot frame, respectively. $\mathbf{u}_{k+1}$ is the current control vector, provided by the step executor and $\mathbf{w}$ is the random variable representing the process noise. $\mathbf{f}_s$ is the transformed system state after the foot change.

The approach is to assign different tasks to different cameras. Different environment informations are captured by the multi-camera vision system and provided to the Hybrid EKF. The wide-angle stereo camera provides the position estimator, the hybrid EKF, with the landmark positions in the image frame, while the conventional stereo camera only brings the position informations of objects to the position estimator. In this hybrid EKF the landmark positions, the object positions, the robot foot positions and the robot camera head positions in the world frame are predicted. This prediction is evaluated by the task-specific view direction planner.

The task-specific view direction planner evaluates the information content increase for all possible view directions in different planners for different tasks. The task-specific optimal view directions are provided to the wide-angle stereo camera and to the conventional stereo camera, respectively.

In our simulations the robot should walk straight forward parallel to the positive x-axis (see Fig. 2). The real path of the camera head is shown with the blue line. The robot walks 24 steps in total. Because of the walking error data which has been taken from real experiments with a humanoid robot [10], the path is not a straight line. The robot foot position in each step is presented by blue stars. The start point of

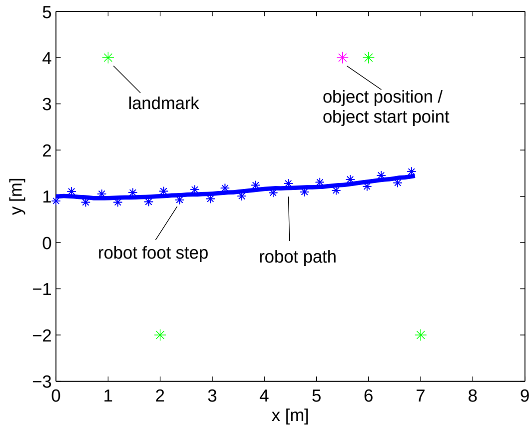

Fig. 2. Simulation scenario: the robot walking path, the landmark positions, the object position.

the right foot is $(0, 0.9)$m. At each step the robot is moving 0.30m in the x-direction. The distance between two robot feet is 0.20m.

There are four landmarks around the robot path. Their positions are $(1,4)$m, $(2,-2)$m, $(6,4)$m and $(7,-2)$m. They are marked with green stars. An object is located in the environment at the position $(5.5,4)$m, marked with a pink cross.

We compare three multi-camera planning strategies evaluating the absolute position estimation errors and the estimation covariance of the robot ego-motion as well as the object motion based on visual information:

- Strategy 1 (S1): Two stereo cameras are not independently controllable. The task decision strategy (see [4]) is used to choose the task for both cameras at each step. The view direction planner computes either the optimal view direction for self-localization or the optimal view direction for object position estimation at each step.
- Strategy 2 (S2): Two stereo cameras are independently controllable. The task decision strategy is also applied to the view direction planners. Both stereo cameras execute the same task at each step, but their view directions may be different.
- Strategy 3 (S3): Two stereo cameras are independently controllable. The view direction strategy proposed in Section II is applied. Different tasks are assigned to different stereo cameras. The wide-angle stereo camera is responsible for the self-localization task, while the conventional stereo camera is accountable for the object position estimation task.

All the cameras are assumed ideal. Lens distortion and quantization effects are not considered.

The other system parameters are set as follows: At each step the positions are measured and estimated three times. Step parameters covariance equals $0.005^2 m^2$ in each direction, x and y. The square of the variance of the robot orientation in each step has the value of $\sqrt{3}\ rad^2$. The initial position variance of the object equals $0.02^2 m^2$, while the initial position variance of the robot also equals 0.02^2m^2.

B. Simulation Results

1) In a Static Environment: Firstly, we consider the situation that the object at $(5.5,4)m$ is not moving, such that this object is also treated as a landmark.

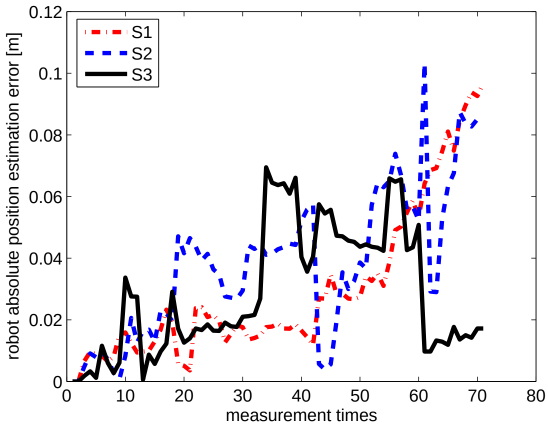

Fig. 3. Absolute position estimation errors of the robot ego-motion in a static environment.

In Fig. 3 the absolute position estimation errors of the robot ego-motion using view direction planning strategy S1, S2 and S3 are compared. The estimation using S3 has no apparent improvement in comparison to the estimation using S1 and S2. The mean value of the robot position estimation error using S3 is 0.0285m, while the mean values using S1 and S2 are 0.0384m and 0.030m, respectively.

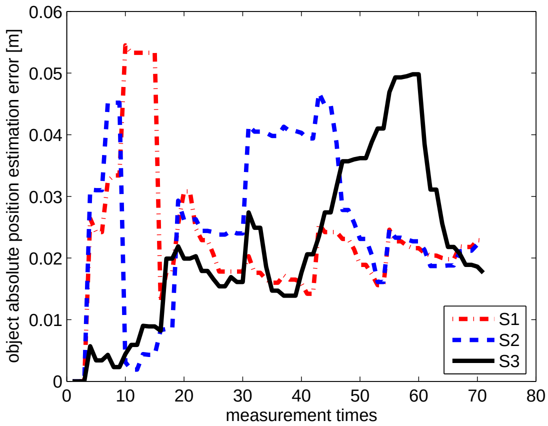

Fig. 4. Absolute position estimation errors of the object in a static environment.

Fig. 4 shows a comparison of the object absolute position estimation errors in this static environment using three strategies. The mean values of the object absolute position estimation errors using S1, S2 and S3 are 0.0229m, 0.0250m and 0.0214m, respectively. The object estimation applying S3 has no apparently better performance than those using S1 and S2, either.

It is established that S3 has no obviously better performance in position estimation in a static environment than S1 and S2.

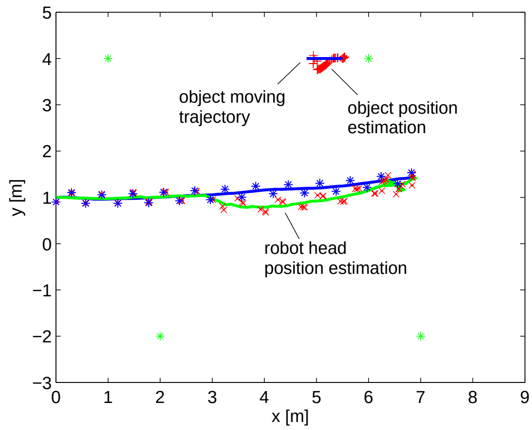

Fig. 5. Simulation result displayed in a overlooking view using S1 in a dynamic environment.

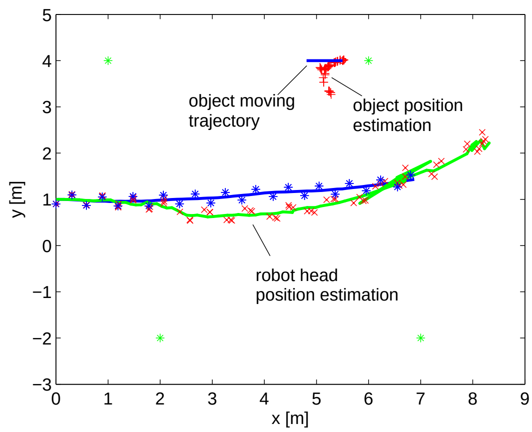

Fig. 6. Simulation result displayed in a overlooking view using S2 in a dynamic environment.

2) In a Dynamic Environment: Now we investigate the case that an object is simultaneously moving along a straight line toward the negative x-direction (see Fig. 5, 6 and 7). The start point is $(5.5,4)$m. The trajectory of the object is displayed by a blue line. During the time interval of one step of the robot, the object moves 0.03m in the negative x-direction. The object position estimation is shown in red crosses. The green line is the estimated trajectory of the robot head. The red stars near the green line are the position estimations of the robot foot currently supporting at each step.

Fig. 5 and Fig. 6 are the overlooking views of the robot motion and the estimation results using S1 and S2. The position estimations of the robot head using S1 and S2 are

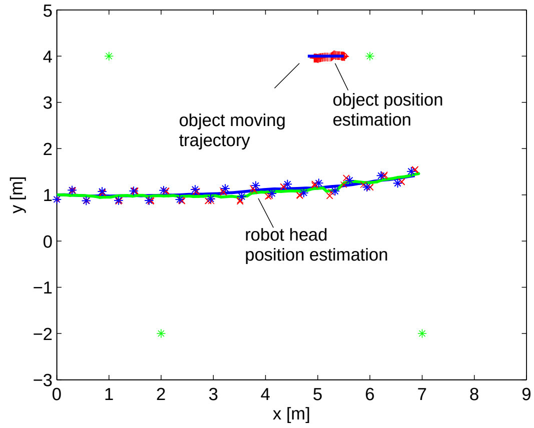

Fig. 7. Simulation result displayed in a overlooking view using S3 in a dynamic environment.

far away from the real path for the reason that the object seen by the robot is also moving. Using S3 we have a significant improvement in the robot position estimation and the object position estimation (see Fig. 7).

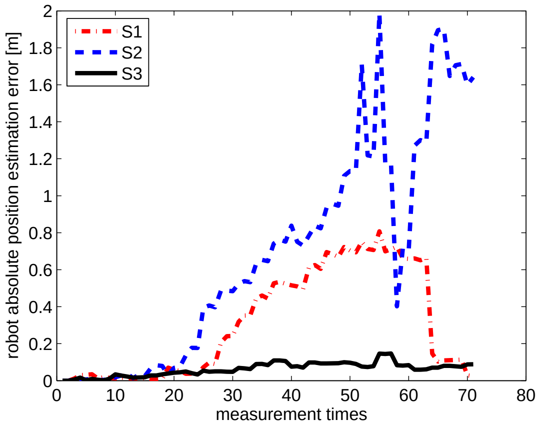

Fig. 8. Absolute position estimation errors of the robot ego-motion in a dynamic environment.

Fig. 8 shows the quantitative comparison of the robot position estimation errors over the simulation time. The mean value of the robot absolute position estimation errors using S1 equals 0.3119m, while the absolute robot position estimation errors applying S2 have the mean value of 0.6841m. If S3 is used, the mean value of the estimation errors is 0.0623m which is obviously smaller than those using S1 and S2. The accuracy is much higher using S3 than those using S1 and S2.

Fig. 9 illustrates a comparison of robot position estimation covariances using S1, S2 and S3 over 71 measure times (three times per step). The covariance of S3 is much smaller and a little smaller than those of S1 and S2. The oscillation is caused by the using of EKF. By the second measurement at each step the covariance of the ego motion is obviously

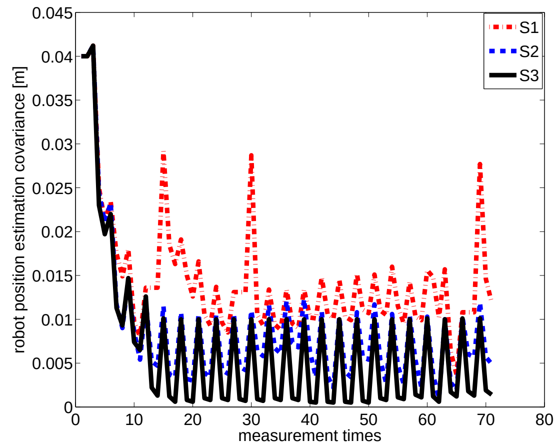

Fig. 9. Robot ego-motion estimation covariance.

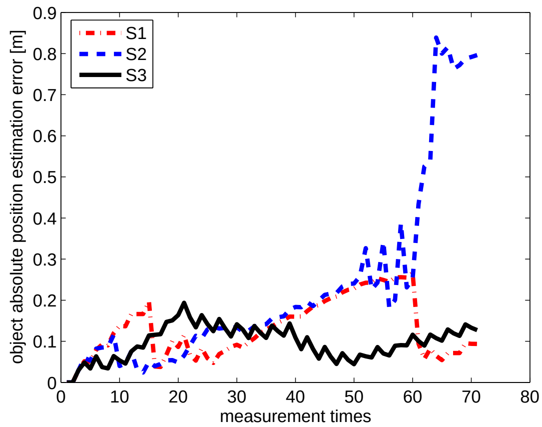

Fig. 10. Absolute position estimation errors of the object motion in dynamic environment.

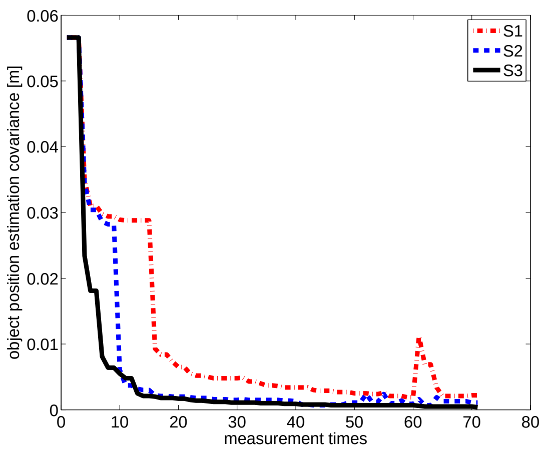

Fig. 11. Object position estimation covariance.

reduced.

Considering the comparison of the absolute errors and covariances, S3 performs better than S1 and S2 for robot ego-motion estimation.

Respectively, S3 also has a better performance of the position estimation of the object motion. A comparison among S1, S2 and S3 is presented in Figure 10. The mean values of the absolute position estimation errors of the object using S1, S2 and S3 are 0.1295m, 0.2300m and 0.0976m, respectively. The average position estimation error using S3 is about 0.03m and 0.13m smaller than those using S2 and S1.

Furthermore, the object position estimation covariances using S1, S2 and S3 are compared in Fig. 11. A same result is achieved as the comparison of robot position estimation. S3 performs much better than S1 and a little better than S2.

From the simulation results above, we can draw a conclusion as follows:

- $S3 > S1 > S2$ for absolute position estimation error,
- $S3 > S2 > S1$ for estimation covariance.

The accuracies of the position estimation of the robot motion and the object motion are apparently improved using S3.

C. Discussion

The simulation results above show a better performance by measuring the position estimation applying the view direction planning strategy S3 than those applying S1 and S2 in a dynamic environment. Without moving objects the accuracy of the position estimation using S3 is not obviously improved. But in a dynamic environment, the robot absolute position estimation error using S3 is even five times smaller than that using S1 and ten times smaller than that using S2. The small estimation covariance of S3 also provides a better accuracy of the estimation. The reason is the decoupled processing of different tasks. Without the influence of the object motion, the wide-angle stereo camera estimates the robot motion accurately.

However, the view direction planning strategy S3 also has its limitations. Firstly, if we have a system in which the robot should accomplish more than two tasks, it is not always comfortable to use more than two cameras. To keep the improvement of accuracy what this view direction planning strategy brings, we can classify the system into several task groups. In each group there are only the tasks which do not have interactive impacts on the computation results of each other. Then we can combine our approach with the task decision strategy in [4].

Secondly, the increase of the absolute position estimation error of the object motion shown in Figure 10 using S1 and S2 is principally due to the increase of the absolute position estimation error of robot motion. It is not overseen that the absolute position estimation error of the object applying S3 increases at first. Then it decreases between the 20. measurement and the 50. measurement despite the fact that the absolute position estimation error of the robot motion grows in this time interval. This phenomenon probably means the object lies out of the field of view of the conventional stereo camera in the intervals from 1. to the 20. measurement and from the 50. measurement to the 71. measurement, which causes the increasing position estimation error of the object.

IV. CONCLUSIONS

An information-based multi-camera view direction planning strategy for mobile robots is discussed. Two concurrent tasks are considered: self-localization and object tracking.

To avoid the interactive estimation error of moving robot and objects, different tasks are assigned to different cameras. Different cameras acquire different informations from the environment, process them separately and plan their view directions individually, using their task-specific criteria.

The performance of our strategy is evaluated in simulations considering a humanoid robot navigation scenario, compared with another two view direction planning strategies. Using the proposed view direction planning strategy, the accuracies of the robot motion estimation and the object motion estimation are significantly improved in dynamic environment. The impact from the object motion on the robot motion estimation is eliminated. Using the proposed strategy, the performance of accomplishing concurrent tasks is improved.

ACKNOWLEDGMENTS

This work is supported in part within the DFG excellence initiative research cluster *Cognition for Technical Systems – CoTeSys*, see also `www.cotesys.org`.

REFERENCES

[1] A. J. Davison and N. Kita, "3d simultaneous localisation and map-building using active vision for a robot moving on undulating terrain," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[2] M. Pellkofer and E. Dickmanns, "Ems-vision: Gaze control in autonomous vehicles," in *Proceedings of the IEEE Intelligent Vehicles Symposium 2000*, 2000.

[3] F. P. F. T. Arbel, "Viewpoint selection by navigation through entropy maps," in *ICCV 1999, pp. 248-254*, 1999.

[4] J. Seara and G. Schmidt, "Gaze control strategy for vision-guided humanoid walking," *at-Automatisierungstechnik*, vol. 2, pp. 49–58, 2005.

[5] B. S. V. Lippiello and L. Villani, "Visual tracking using a fixed multi-camera system," in *Proceedings of the 2003 IEEE International Conference on Robotics & Automation*, 2003.

[6] T. Ahmedali and J. Clark, "Collaborative multi-camera surveillance with automated person detection," in *Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV'06)*, 2006.

[7] H. Y. B. M. B. Keyes, R. Casey and Y. Georgiev, "Camera placement and multi-camera fusion for remote robot operation," in *Proceedings of the IEEE International Workshop on Safety, Security and Rescue Robotics, Gaithersburg, MD*, 2006.

[8] S. H. J. Krumm *et al.*, "Multi-camera multi-person tracking for easyliving," in *3rd. IEEE International Workshop on Visual Surveillance, Dublin, Irland*, 2000.

[9] K. Kühnlenz, *Aspects of Multi-Focal Vision*. PhD thesis, Technische Universität München, 2006.

[10] J. Seara, *Intelligent Gaze Control for Vision-Guided Humanoid Walking*. PhD thesis, Technische Universitt Mnchen, 2004.