We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



30

On Foveated Gaze Control and Combined Gaze and Locomotion Planning

Kolja Kühnlenz, Georgios Lidoris, Dirk Wollherr, and Martin Buss
Institute of Automatic Control Engineering, Technische Universität München
D-80290 München, Germany

1. Introduction

This chapter presents recent research results of our laboratory in the area of vision and locomotion coordination with an emphasis on foveated multi-camera vision. A novel active vision planning concept is presented which coordinates the individual devices of a foveated multi-camera system. Gaze direction control is combined with trajectory planning based on information theoretic criteria to provide vision-based autonomous exploring robots with accurate models of their environment.

With the help of velocity and yaw angle sensors, mobile robots can update the internal knowledge about their current position and orientation from a previous time step; this process is commonly referred to as dead-reckoning. Due to measurement errors and slippage these estimations are erroneous and position accuracy degrades over time causing a drift of the estimated robot pose. To overcome the drift problem it is common to take absolute measurements evaluating visual information, which are fused dynamically with the odometry data by applying Kalman-filter or other techniques, e.g. (Dissanayake et al., 2001). The use of active vision systems for navigation is state-of-the-art providing a situation-related selective allocation of vision sensor resources, e.g. (Davison & Murray, 2002; Seara et al., 2003; Vidal-Calleja et al., 2006). Active vision systems comprising only one type of vision sensor face a trade-off between field of view and measurement accuracy due to limitations of sensor size and resolution, and of computational resources. In order to overcome this drawback the combined use of several vision devices with different fields of view and measurement accuracies is known which is called foveated, multi-resolution, or multi-focal vision, e.g. cf. (Dickmanns, 2003; Kühnlenz et al., 2006; Ude et al., 2006). Thereby, the individual vision devices can be independently controlled according to the current situation and task requirements. The use of foveated active vision for humanoid robot navigation is considered novel.

Active vision is also frequently utilized in the context of robotic exploration. Yet, gaze control and locomotion planning are generally decoupled in state-of-the-art approaches to simultaneous localization and mapping (SLAM). An integrated locomotion planning and gaze direction control concept maximizing the collected amount of information is presented in the second part of this chapter. This strategy results in more accurate autonomously acquired environment representations and robot position estimates compared to state-of-the-art approaches.

The chapter is organized as follows: In Section 2 vision-based localization and mapping in the context of humanoid robots is surveyed; Section 3 is concerned with foveated multi-

Open Access Database www.i-techonline.com

Source: Humanoid Robots, New Developments, Book edited by: Armando Carlos de Pina Filho
ISBN 978-3-902613-02-8, pp.582, I-Tech, Vienna, Austria, June 2007

camera coordination; novel concepts of gaze control and path planning coordination are presented in Section 4; evaluation studies comparing the novel concepts to conventional planning approaches and vision systems are presented in Section 5; conclusions are given in Section 6.

2. Vision-Based Localization and Mapping for Humanoid Robots

Most state-of-the-art humanoid robots are equipped with vision systems. The benefits of using these vision systems for providing absolute measurements of the robot pose in the environment are obvious: pose information on landmarks is provided and no additional devices as, e.g., laser scanners are necessary. Being equipped with internal sensors - angular sensors in the joints and widely used gyros and accelerometers in the trunk - humanoid robots are basically capable of dead-reckoning, i.e. the ability to update position and orientation known from previous measurements. Thus, common simultaneous localization and mapping techniques are applicable which are covered by common literature, e.g. (Sabe et al., 2004; Ozawa et al., 2005; Thomson & Kagami, 2005; Stasse et el., 2006).

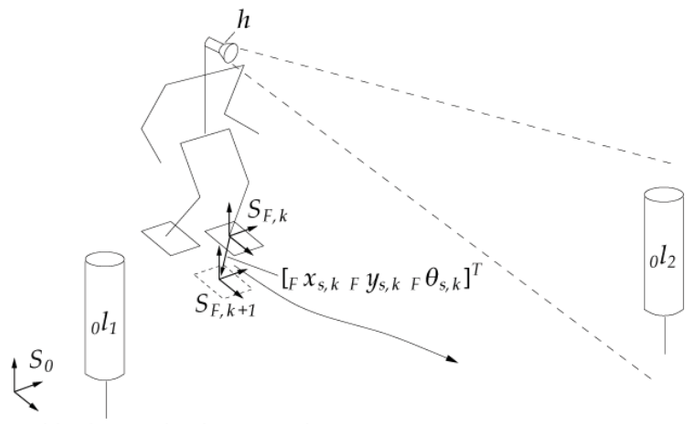


Fig. 1. Humanoid robot navigation scenario.

A fundamental aspect in simultaneous localization and mapping for humanoid walking is the formulation of a state-space model accounting for the footstep sequences of the robot. In vision-based SLAM, the system state, i.e. the robot pose and environment point positions, are predicted based on the dead-reckoning model of the mobile robot. Common Kalman-filter techniques are applied in order to obtain more accurate estimations accounting for uncertainties in the robot locomotion. Whenever visual measurements of environmental points are taken, updates of the robot state are computed. Changing ground contact situations of the feet, however, result in different kinematic chains from a world reference frame to measured environment points. This discontinuous movement of the humanoid robot requires an adaptation of the filter formulation. In earlier works we proposed a hybrid formulation of the state-space model in order to capture this locomotion principle (Seara et al., 2003). Thereby, the robot reference frame is placed in the foot currently in contact with the ground and is switched whenever the supporting foot changes. The dead-reckoning model is expressed by

$$x_{k+1} = x_k(1-\gamma_k) + f_s(x_k, u_k, d_{x,k})\gamma_k \quad , \tag{1}$$

where state-vector x contains the robot foot pose and the landmark positions, d represents system noise capturing dead-reckoning uncertainties, and $\gamma \in \{0; 1\}$ is a binary variable indicating a change of the supporting foot when γ=1. The commanded step u is expressed by

$$u_k = \begin{bmatrix} {}_F x_{s,k} & {}_F y_{s,k} & {}_F \theta_{s,k} \end{bmatrix}^T \quad , \tag{2}$$

including the commanded step position $[x_s\ y_s]^T$ and orientation θ_s with respect to the current supporting foot frame S_F. Figure 1 schematically shows a typical SLAM situation of a humanoid robot with the reference frame currently placed in the left foot.

In vision-based SLAM field of view restrictions of the vision device strongly limit the number of landmarks to be observed simultaneously. Yet, a larger field of view can only be realized accepting a lower measurement accuracy of the vision device mainly due to limitations of sensor size and resolution. Therefore, we propose the use of several vision devices which provide different fields of view and accuracies and a novel gaze control concept for coordinating the individual vision devices in order to provide both, large field of view and high measurement accuracy, simultaneously. These foveated active vision concepts for robot navigation are discussed in the following section.

3. Foveated Multi-Camera Coordination

3.1 Active Vision in SLAM

In order to gather an optimal situation-dependent amount of information the control of the vision system pose is common. To date, there are only few works in the area of active vision-based SLAM, e.g. (Davison & Murray, 2002; Se et el., 2002; Vidal-Calleja et el., 2006) which are based on measures representing the information gathered with respect to the SLAM task. All these approaches are greedy strategies only evaluating the current situation without considering future planning steps. In order to obtain an optimal gaze direction considering also some future planning steps, we proposed a gaze direction planning strategy with limited time horizon (Lidoris et al., 2006). Furthermore, in earlier works (Seara et al., 2003) we introduced a gaze control strategy considering concurrent tasks, localization, and obstacle avoidance for humanoid robots in order to account for navigation in physical environments.

3.2 Foveated Active Vision

Vision systems comprising only one type of vision sensors face a tradeoff between measurement accuracy and field of view due to limitations of sensor size and computational resources for image processing. Accuracy and field of view are mainly determined by the focal-length of the lens or mirror optics, respectively. Within the context of robot navigation this tradeoff implies a compromise between localization accuracy and keeping a large part of the scene in view.

With an active vision system this tradeoff could be compensated providing that a sufficiently accurate map of relevant landmarks or structures of interest to be observed is known a priori. Then the highest available focal-length and, thus, the highest measurement accuracy could be chosen. If additionally very fast gaze shifts can be realized, the narrow field of view would be acceptable as visual attention can be directed dynamically towards

the most relevant structure in the current situation. Yet, in a variety of scenarios this approach is unsuitable or even unrealizable. In at least partially unknown environments and in exploration scenarios a sufficient map is not available and thus has to be created online. However, due to the strongly limited field of view the detection of new objects of potential interest is hardly possible. Another aspect are potentially relevant or even dangerous objects or activities in the local surroundings of the robot which cannot be detected.

In order to overcome the common drawback of trading field of view versus measurement accuracy, the combination of wide-angle and telephoto vision devices has been suggested. Such systems provide at the same time both, an observation of a large part of the environment and a selective examination with high accuracy. In common literature these systems are referred to as foveated, multi-resolution or multi-focal systems. The individual vision devices may be fixed with respect to each other or may be independently motion controllable in one or more degrees of freedom. Most common embodiments of foveated systems are used in state-of-the-art humanoid robots comprising two different cameras combined in each eye which are aligned in parallel, e.g. (Brooks et al., 1999; Ude et al., 2006; Vijayakumar et al., 2004). Systems for ground vehicles, e.g. (Apostoloff & Zelinsky, 2002; Maurer et al., 1996; Dickmanns, 2003) are another prominent class. An upcoming area are surveillance systems which strongly benefit from the combination of large scene overview and selective observation with high accuracy, e.g. (Bodor et al., 2004; Davis & Chen, 2003; Elder et al., 2004; Jankovic & Naish, 2005; Horaud et al., 2006). An embodiment with independent motion control of three vision devices with a total of 6 degrees-of-freedom (DoF) is the camera head of the humanoid robot *LOLA* developed at our laboratory which is shown in Figure 2 providing more flexibility and, due to directly driven gimbals, faster camera motions than other known systems, cf. e.g. (Kühnlenz et al., 2006).

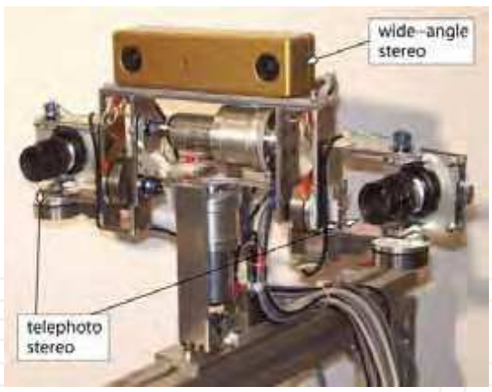


Fig. 2. Multi-focal vision system of humanoid LOLA (Kühnlenz et el. 2006).

Most known methods for active vision control in the field of foveated vision are concerned with decision-based mechanisms to coordinate the view direction of a telephoto vision device based on evaluations of visual data of a wide-angle device. For a survey on state-of-the-art methods cf. (Kühnlenz, 2006). A first approach towards a coordination of foveated multi-camera view direction planning for humanoid walking has been investigated in our laboratory which is presented in the following sections.

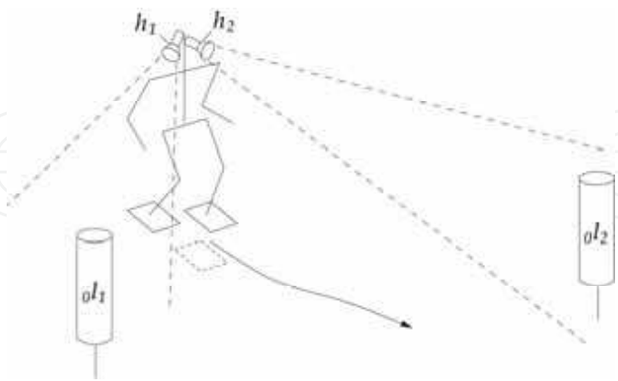


Fig. 3. Humanoid robot navigation scenario with multi-camera vision.

3.3 Considerations for Camera Coordination

In the area of foveated vision a large body of literature exists covering mechanisms to assess peripheral visual data in order to derive control commands to direct foveal attention towards regions of potential interest. The most prominent computational approaches in the biologically inspired field are computational neuroscience models of top-down modulated bottom-up attention weighting particular visual features of the environment, e.g. (Koch & Ullmann, 1984; Itti & Koch, 2001). In the technical field a larger variety of different methods is known. Common approaches solve optimization problems, assess the visual information content, or evaluate the environment towards particular visual features, e.g. (Bodor et al., 2004; Darrell, 1997; Pellkofer & Dickmanns, 2000; Scasselati, 1998; Shibata et al., 2001). To date, only few works have been presented on foveated and multi-camera attention considering locomotion tasks. Prominent examples are the works of (Pellkofer & Dickmanns, 2000) in the field of visual guidance of autonomous ground vehicles and gaze control concepts for the humanoid *LOLA* conducted in our laboratory (Kühnlenz, 2006), where optimal view directions are determined by maximizing the information gain.

In earlier works we proposed a task-related information measure as quality measure termed *incertitude* (Seara et al., 2003) which has been taken as the basis for the coordination of the two stereo-camera devices of *LOLA* with different characteristics. The mission of the humanoid robot is a locomotion task to walk along a certain path or to explore the world. A primary condition for view direction planning, thus, has to consider the quality of locomotion task accomplishment in order to determine an optimal view direction for the next time step. The concept of incertitude captures this task-dependence by evaluating the predicted certainty of the estimated robot foot pose. Therefore, the average of the main axes lengths of the foot pose covariance matrix confidence ellipsoid is computed

$$v_0 = \frac{1}{2}\sum_i \sqrt{e_i}\,, \tag{3}$$

where counter i covers the considered components of the foot pose and e_i are the eigenvalues of the predicted foot pose covariance matrix P_{uu} which is a submatrix of the predicted covariance matrix of a possible target state as estimated by the Kalman-filter, e.g. cf. (Dissanayake et el., 2001)

$$P^{i}{}_{k|k} = \begin{pmatrix} P^{i}{}_{uu} & P^{i}{}_{um} \\ P^{i}{}_{mu} & P^{i}{}_{mm} \end{pmatrix}, \tag{4}$$

where $\boldsymbol{P}^{i}{}_{uu}$ is the error covariance matrix of the robot state estimate, $\boldsymbol{P}^{i}{}_{mm}$ is the map covariance matrix of the landmark state estimates and $\boldsymbol{P}^{i}{}_{um}$ is a cross-covariance matrix between robot and landmark states. Low values of the defined measure (3), thus, indicate a high certainty of the robot pose estimation and, therefore, good task performance for the locomotion task. Additional measures to assess the performance of secondary tasks have been proposed which also may have an indirect impact on the performance of the primary (locomotion) task, e.g. field of view limitations, presence of activities, etc., (Kühnlenz, 2006). These measures are all extensions to the central gaze control concept and, therefore, out of scope of this chapter.

Given such measures to assess the task performance of the humanoid robot the next task is to derive appropriate view directions for the individual vision devices in the following time step in order to achieve a particular desired task performance. This gaze control concept is topic of the following section.

3.4 Multi-Camera View Direction Planning

Common approaches to optimal view direction planning for mobile systems are based on a maximization of the information gain, e.g. (Davison, 1998; Pellkofer & Dickmanns, 2000; Seara et al., 2003), in order to determine either a selected gaze shift or a sequence of gaze behaviors. Particularly, in the field of foveated and multi-camera vision also visibility conditions are considered, e.g. (Pellkofer & Dickmanns, 2000; Kühnlenz, 2006).

The basic principle of multi-camera coordination in this chapter is an information maximization over a set of possible view directions of independent vision devices. The assumed task of the robot is to follow a path as closely as possible. As a consequence the estimation error of the robot pose within the environment during its motion has to be minimal in order to complete the mission optimally. The presumed objective for view direction planning is to gather the largest possible amount of information with respect to the task to be accomplished. An information gain corresponds to a reduction of uncertainty. In order to maximize the information gain the robot pose error has to be minimized by selecting appropriate view directions of the individual cameras of the foveated multi-camera vision system. Following this, an optimal configuration of view directions for the locomotion task in the next time step satisfies the condition of minimizing the robot pose estimation error. In terms of the task-related information measure defined in the previous section this gaze control strategy can be expressed by

$$\hat{\boldsymbol{\Omega}}^{*} = \arg\min_{\hat{\Omega}} \hat{v}_0, \tag{5}$$

where $\boldsymbol{\Omega}=[pan_1\ tilt_1\ \ldots\ pan_n\ tilt_n]^T$ is a configuration of pan- and tilt-angles of all vision devices, v_0 is the incertitude information measure defined in the previous section, and $(.)^*$ denotes the optimal value. This method constitutes an extension to our earlier works on gaze control for humanoid robots (Seara et al., 2003) generalizing them to multi-camera vision systems. In Section 6, a comparative evaluation of this strategy is presented assuming a humanoid robot navigation scenario with sparsely distributed point landmarks.

The presented gaze control strategy considers a preplanned path of the humanoid robot which is not altered as the robot moves. The following section is concerned with combined

planning of gaze direction and locomotion path in order to provide the mobile robot with capabilities of exploring unknown environments.

4. Combined Gaze Direction and Path Planning

In the previous section a foveated approach to active vision has been presented which optimally controls the devices such that the robot pose error is minimized. This section is concerned with a novel approach which combines locomotion planning and gaze direction control concepts in order to improve autonomous robotic exploration.

4.1 Locomotion Planning for Exploration and SLAM

Robotic exploration is largely understood as investigating an unknown environment such that the area is covered by the robot sensors and a representation is generated allowing the robot to perform its tasks with a certain amount of confidence. Early approaches focused on generating motion commands which minimize the time needed to cover the whole terrain. This was achieved by extracting frontiers between known and unknown areas (Yamauchi, 1998; Koenig & Tovey, 2001) and visiting the nearest unknown area. Such approaches only distinguish between previously visited and unknown terrain without taking into account the amount of information gathered after each action. To incorporate the uncertainty about the state of the environment, (Moorehead et al., 2001) try to minimize the uncertainty of the robot about grid cells, by using entropy as a criterion. Further, (Grabowski et al., 2003) present an approach in which the robot is forced to observe obstacles from different viewpoints so that sharper boundaries between objects and free-space are acquired. However, the techniques mentioned above assume the location of the robot as known.

Recently, some techniques have been proposed which actively control the motion of the robot while simultaneously creating a map of the environment and localizing the robot in it. In (Feder et al., 1999) information gain is introduced as a measure of utility for locally deciding on exploration actions. Formal information measures, as discussed in the introduction of this section, can be used to quantify uncertainty and therefore evaluate the effect of future control actions on the quality of the robot state estimate. Therefore, (Bourgault et al., 2002) introduced a utility function which trades off the cost of exploring new terrain with the utility of selecting future positions that reduce uncertainty. In (Stachniss et al., 2005) a similar decision theoretic framework is used in combination with a particle filter based SLAM solution for robotic exploration. Further, (Bryson & Sukkarieh, 2005) present simulated results which demonstrate the effect of different actions to information gain for unmanned aerial vehicles performing vision-based SLAM.

All the approaches mentioned above, perform a greedy optimization based on information theoretic criteria for the trajectory generation only over the next time step. However, some planning approaches have been introduced which demonstrate improved performance. Such a planning approach that introduces a new measure of map quality is described in (Sim & Little, 2006). However, some initial state estimate of all the landmarks is assumed and sensors are assumed to have an unlimited field of view. Another multi-step planning algorithm for SLAM is described in (Huang et al., 2005).

4.2 Combined Gaze Direction and Path Planning

Most conventional approaches in autonomous exploration and active vision either control the motion of the robot or the active sensors. In this chapter, we adapt the control inputs of

the robot and the sensory system simultaneously so that the best state estimates possible are acquired and as much new terrain as possible is explored.

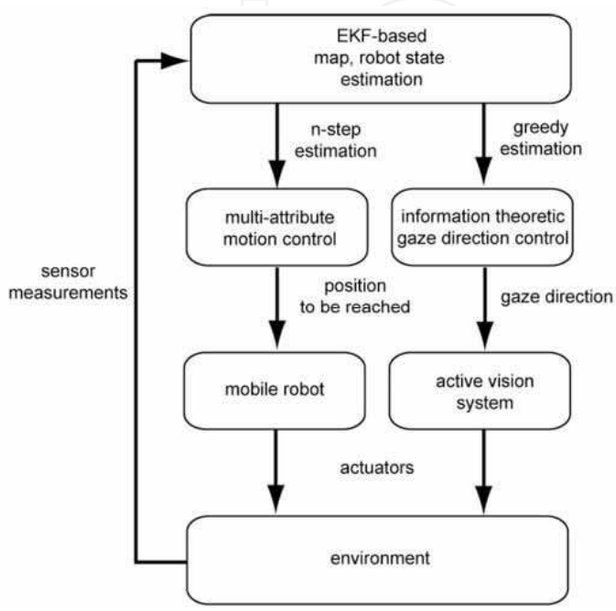


Fig. 4. Proposed motion and gaze direction control scheme

Figure 4 illustrates the proposed motion and gaze direction control scheme. The robot and its active vision system are controlled by two modules which use a common model of the environment. For trajectory planning, a multi-step prediction algorithm is introduced in order to evaluate all possible positions that can be reached by the robot over a finite given time horizon. This estimation forms a multi-attribute function which is used to decide where the robot should move next. A trade-off is made between localization, map accuracy, and proactivity of exploration. For the gaze direction control a greedy information-based optimization is used to choose those view directions that minimize position and map uncertainties. The robot depends on noisy data gained from the visual sensors and at the same time its actions affect the quality of the collected data and its environment.

For vision-guided robots one definition for optimally using sensory resources is selecting the next gaze direction such that measurements are obtained which are most informative about the state of the environment and the robot. This raises the question how information gain can be measured. A common measure of uncertainty is entropy. Which has been introduced by (Shannon, 1948). Entropy for a multivariate Gaussian distribution $p(x)$, with covariance $\boldsymbol{P}$ is defined as

$$H(p(x)) = \frac{1}{2}\log((2\pi)^n |P|) \tag{6}$$

Since the determinant of a matrix is a measure for its volume, the entropy measures the compactness and, thus, the informativeness of a distribution. In order to measure the utility of a gaze direction which will result in an observation z, we will use the mutual information

gain $I[x,z]$. The gain of information between any two distributions can be computed as the change in entropy. In our case this change is the difference between the entropies of the state estimates before and after making an observation which are both multivariate Gaussians with covariances $P_{k+1|k}$ and $P_{k+1|k+1}$. Therefore, the information gain evaluates to

$$I[x,z] = H(x) - H(x \mid z) = \frac{1}{2}\log(\left|P_{k+1|k}\right|) - \frac{1}{2}\log(\left|P_{k+1|k+1}\right|) \tag{7}$$

This information gain can be calculated as a function of the state covariance matrix. From (7) it is obvious that information gain $I[x,z]$ becomes maximal, if the determinant of $P_{k+1|k+1}$ is minimized. Starting from the current state estimate the covariances of the states that can be observed next by the vision sensors are predicted. The equations for the prediction step of the classical SLAM algorithm based on an extended Kalman-filter (Dissanayake et al., 2001) are used. After all covariances are predicted the most informative state can be calculated by minimizing $|P_{k+1|k+1}|$. The new optimal gaze direction Ω of the active vision system corresponding to this state is then computed.

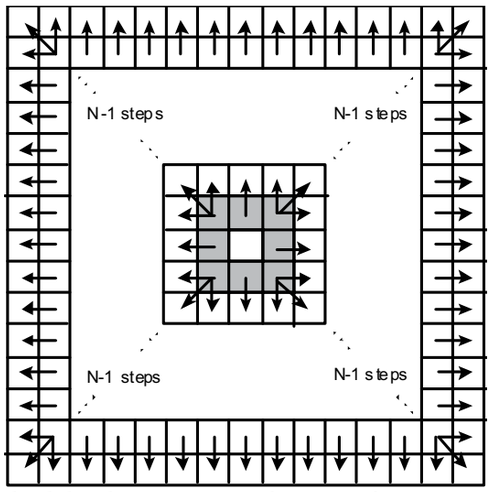


Fig. 5. Region covered while planning over a horizon of N steps. Highlighted grid cells show which cells are taken into account for gaze direction control.

The first step for choosing the next destination for the robot is to estimate the states and covariances of all possible positions that can be reached over its planning horizon. A discretized grid environment is used, where each grid represents a position that can be reached by the robot over future time steps. Therefore, the size of the grid cells depends on the physical properties of the robot. Based on this discretized environment the states and their covariances are computed. While the robot moves, observations are made and used to update the state estimation. This way, all available information is being used. More specifically, based on an initial state estimate and covariance matrix, we calculate all possible robot states and their covariances after N time steps and choose to move to the one that is most informative, namely the one that minimizes relative entropy as described in the previous section. A mathematical description of the algorithm used to produce the multi-step predictions, can be found in (Lidoris et al., 2007). The estimation procedure evolves in a

square-like manner, as shown in Figure 5. Starting from the currently estimated state the first eight neighboring states and covariances are computed. At each step, the estimated state and covariances of the neighboring states are used to infer the next ones until step N. By always using the nearest neighbor in the estimation process, the estimation error is kept minimum. Over each time step k, $8k$ new states are calculated. The control signal, $u_{i,k}$ required in order to drive the robot from a state j to a state i, is chosen as indicated by the arrows in Figure 5.

Using the predicted covariance matrix (4) and the concept of relative entropy mentioned previously, each possible future position of the robot can be evaluated to choose the appropriate target position for the robot. The destination that maximizes the function

$$V_i = \frac{1}{2}\log(\frac{P^i_{uu}}{P^0_{uu}}) - \gamma\frac{1}{2}\log(\frac{P^i_{mm}}{P^0_{mm}}) \tag{8}$$

is chosen as a target for the robot. The first part of the function is a measure of the position uncertainty the robot will encounter in the future position and the second part is a measure of the map quality. The constant γ can be used to adjust the behavior of the robotic explorer. Setting γ to values smaller than one, results in a conservative exploration policy, since the robot will stays near to well-localized features giving more attention to localization. Large values of γ increase the proactivity of the explorer in the sense that it moves to unknown areas neglecting the lower localization accuracy. After selecting the target position which maximizes (8), the robot moves making observations which are used to update the estimated state and covariance. Each time a new state estimate is available, a recalculation of the gaze direction is made. This way we use all new information that becomes available during robot motion. Replanning takes place after N time steps when the target position is reached.

5. Comparative Evaluation Studies

In Sections 3 and 4 novel concepts of foveated active vision and combined gaze and locomotion coordination have been presented. This section is concerned with evaluation studies in order to assess the performance of the proposed approaches in comparison to state-of-the-art planning methods and vision systems.

5.1 Foveated Active Vision

In Section 3 a task-related information measure for the humanoid robot locomotion task and a multi-camera view direction planning strategy have been introduced. This section is concerned with an evaluation study comparing the performance of the novel approach to a conventional single stereo-camera strategy.

Considered is a typical locomotion task of a humanoid robot with the robot moving along a planned path. It has visual and odometrical capabilities such that it is able to localize itself and other objects within the environment. The robot is equipped with a foveated multi-camera vision system consisting of two stereo-camera devices with independently controllable pan- and tilt-angles, different focal-lengths, and different fields of view. The robot's mission is to follow the desired path. Therefore, it has to localize itself continually evaluating odometry data and visual information. Given a particular environmental situation, i.e. configuration of observable objects and robot pose, the objective is to dynamically select appropriate view directions for both vision devices. Figure 3 exemplarily shows a situation in the considered

navigation scenario where a humanoid robot fixates two landmarks with two vision devices of its foveated multi-camera vision system in order to localize itself in the world.

In order to demonstrate the benefits of foveated multi-camera view direction planning the proposed gaze control approach is now evaluated in a structured humanoid robot navigation scenario. Several vision system configurations are evaluated by comparison of the achieved navigation performances. The basic scenario is shown in Figure 6. Four landmarks are distributed within a rectangular environment. The mission of the robot is to follow the planned path in *x*-direction. In order to complete the mission successfully the robot has to localize itself within the environment evaluating available visual information on the positions of the identified landmarks. The robot pose is estimated dynamically using the Kalman-filter approach described in Section 2. In order to maximize the information gain optimal view directions of the individual vision devices are selected dynamically based on the proposed approach in Section 3.4. The positions of the landmarks are not known a priori nor are the number of landmarks. Configurations of the vision system in the considered scenario to be compared are: a) conventional single stereo-camera, focal-lengths 20mm, aperture angles 30°, stereo-base 25cm; b) foveated stereo-camera with two cameras per eye aligned in parallel, focal-lengths 2mm and 40mm, respectively, aperture angles 60° and 10°, respectively, stereo-bases 25cm; c) two independent stereo-cameras, focal-lengths 2mm and 40mm, respectively, aperture angles 60° and 10°, respectively, stereo-bases 25cm. All cameras are ideal, based on the pinhole camera model neglecting lens distortion and quantization effects. Gaussian vision sensor noise with a standard deviation of 1 pixel is considered. Dead-reckoning errors are taken from experiments with the humanoid robot *JOHNNIE*.

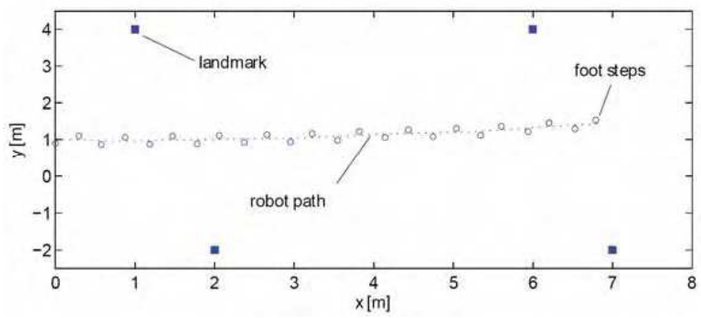


Fig. 6. Top-view of humanoid robot navigation scenario.

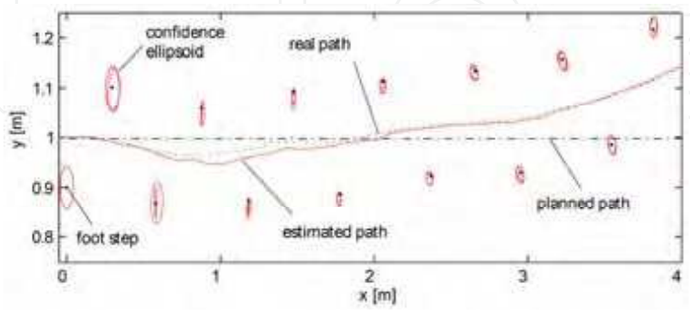


Fig. 7. Real, estimated, and planned paths and footsteps.

The navigation performance is rated assessing the localization accuracy. Therefore, the covariance matrix of the robot position is evaluated computing the areas of the 90%-confidence ellipses of the footstep position uncertainties of the humanoid robot. Figure 7 contains a cut-out of Figure 6 showing the planned and real paths, the path estimated by the Kalman-filter, the foot step positions, and their covariance ellipses. It is noted that due to dead-reckoning errors the real path deviates increasingly from the planned path as locomotion control is open loop. The estimated path follows the real path well.
Figure 8 and Figure 9 show the resulting view directions for each step of the robot for the individual vision systems. The propagations of the areas of the confidence ellipses are shown in Figure 10. Table 1 shows a comparison of the means of the confidence ellipse areas and the average number of landmarks visible for the vision systems for all scenarios. These results are discussed in the following.

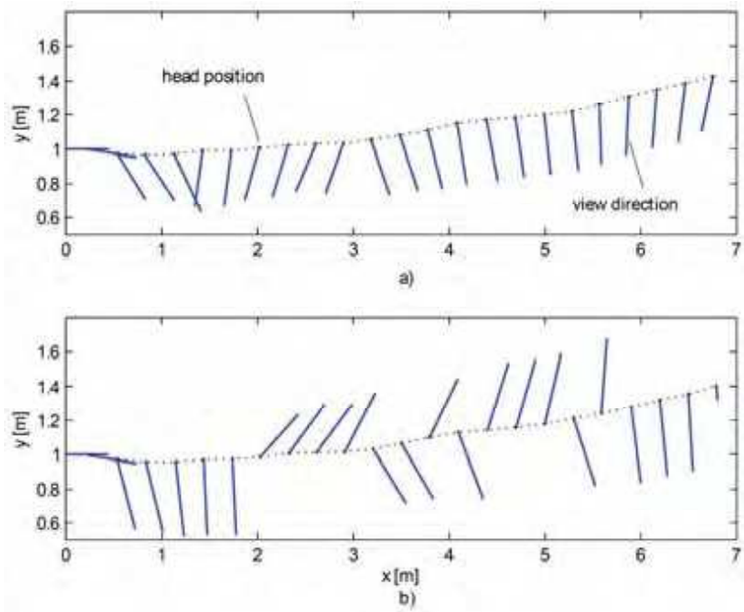


Fig. 8. Pan-angles of the a) conventional stereo-camera and b) of the foveated stereo-camera with two cameras per eye.

Mono-focal vision systems, i.e. systems comprising only one sensor type, suffer from a trade-off of accuracy versus field of view. In robot localization not only measurement accuracy, but also the number of visible landmarks is an important factor in order to determine the current robot position. Depending on the distribution of landmarks and the current situation it may be better to observe more landmarks with lower accuracy in one situation and fewer landmarks with higher accuracy in another. Thus, mono-focal systems are always a compromise working well only for a very limited class of environmental conditions and situations, however, failing in others. This problem is reflected by the results shown in Figure 8a and Table 1. The upper two of the sparsely distributed landmarks are not detected at the initial position and, thus, the planner only

considers the lower landmarks. Most of the time only one landmark is visible and used for localization.

The foveated vision systems provide much more flexibility. Due to the large field of view of the wide-angle device more landmarks are detected in the initial position to be considered by the planner. So, at each step two or more landmarks are used for localization whereas at least one landmark is focused with telephoto cameras resulting in a significantly higher certainty of the estimated footstep positions of the humanoid robot as shown in Figure 10 and Table 1.

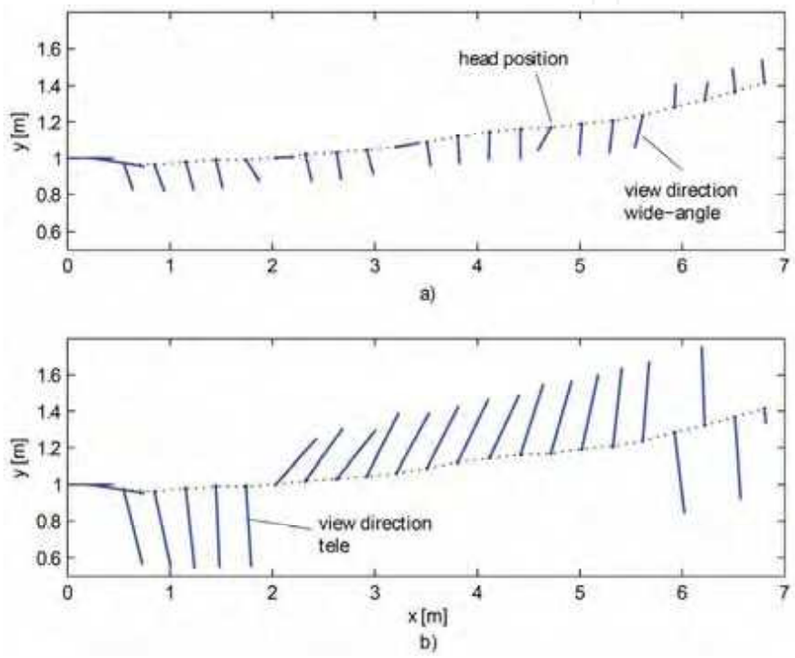


Fig. 9. Pan-angles of foveated vision system with two independent stereo cameras; a) telephoto and b) wide-angle stereo-camera.

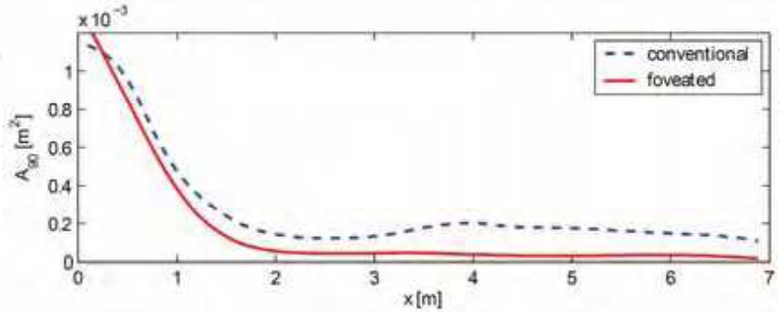


Fig. 10. Comparison of the areas A_{90} of the 90%-confidence ellipses of the footstep position covariance matrix using a conventional and a foveated vision system.

vision system	A_{90} [$10^{-4}m^2$]	average number of visible landmarks
a) conventional	2.7	1.1
b) foveated	1.7	2.3
c) foveated	1.6	2.4

Table 1. Mean of the areas A_{90} of the 90%-confidence ellipses of the footstep covariance matrix and average number of visible landmarks.

5.2 Combined Gaze and Trajectory Planning

In this section the performance of the proposed combined gaze direction and motion planning scheme described Section 4 is evaluated. The simulated environment consists of an area of size 20x20 meters with randomly allocated features. The active simulated head which is mounted on top of the robot is assumed to have a field of view of 60° and a maximum viewing range of 6 meters. No previous knowledge of the environment is assumed. Since it is out of the scope of this paper to deal with the correspondence problem of SLAM or feature selection, feature association is considered known and all observed features are used. A harsh odometry error of 10% is chosen, since are interested to see how our algorithm performs with very inaccurate robot models. A sensor model with a variance proportional to the distance for bearing and range measurements is used having the same high noise level. The active head can be moved with high angular velocities, so that saccadic movements are simulated. Finally, γ in (8) is chosen heuristically such that the robot balances between keeping good pose estimates and exploring the environment. The simulations time is set to 60s.

We first ran a simulation assuming a passive sensor, directed always straight ahead of the robot and a greedy policy for motion control which considers only the eight neighboring states. Only seven features were observed and map and localization uncertainty were very high as shown in Figure 11.

Next we conducted simulation with the proposed gaze direction control and again a greedy policy, followed by a simulation with a three-step planning horizon. In Figure 11 the absolute position error is depicted, for all three cases. It is evident that error reduces significantly as the planning horizon grows and gaze direction control is used.

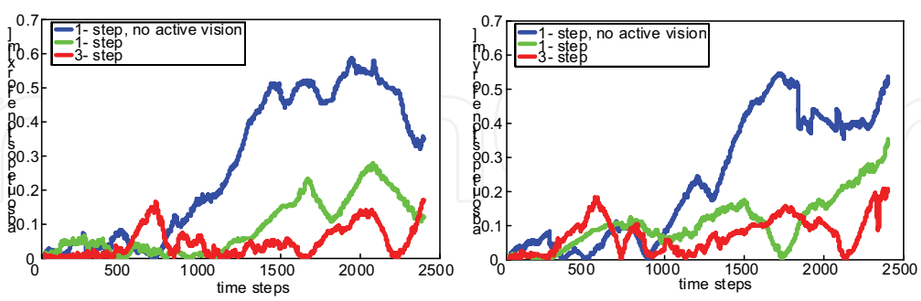


Fig. 11. Absolute position error as a function of time for one-step planning horizon without gaze direction control and one-step and three-step planning horizons with gaze direction control.

The map accuracy is illustrated in Figure 12 by the error ellipsoids for each observed feature for the final map. It is obvious that map accuracy grows as the planning horizon becomes larger. Also more features are observed if gaze direction control is used. From the final map acquired in the case of a three-step planning horizon with gaze direction control, it becomes clear that the proposed approach balances well by observing a large number of features and also building an accurate map.

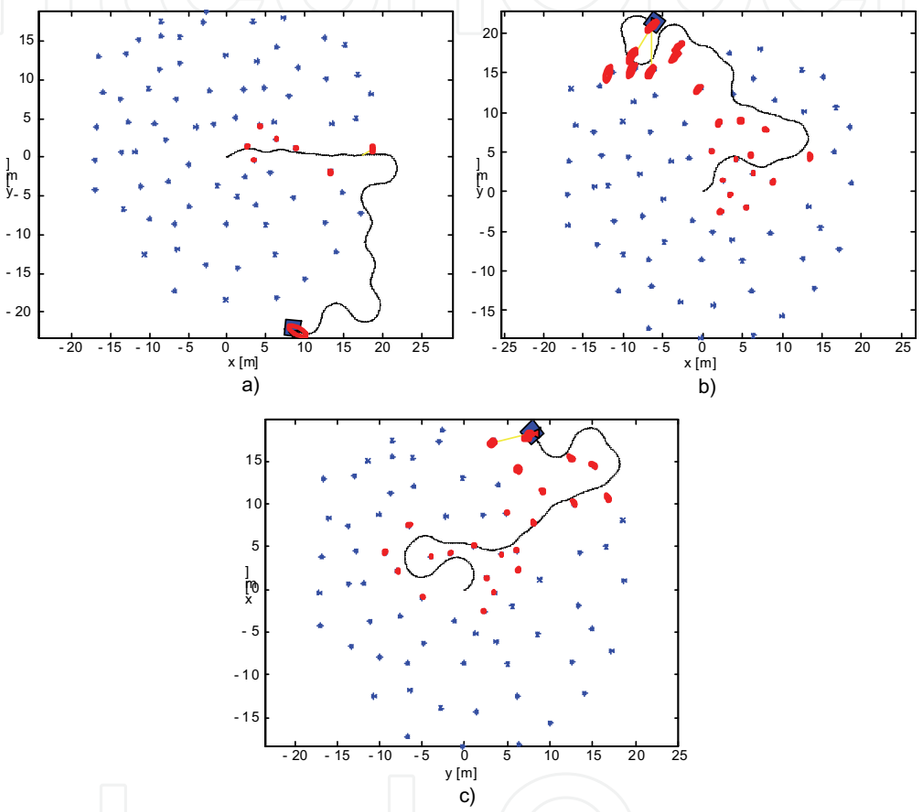


Fig. 12. Map accuracy illustrated by the error ellipsoids of each observed feature for the final map; a) one-step planning horizon without gaze direction control, b) one-step, and c) three-step planning horizons with gaze control. The estimated robot trajectory is illustrated by the black lines, while the red triangle on-top of the robot represents the active head and its gaze direction.

Figure 13 shows the reduction of the entropy over time. Each time a new feature is observed entropy reduces. For that reason it is step-formed. The greedy approaches need more time to reduce entropy and the larger the planning horizon is, the more entropy is reduced. Furthermore, when the planning horizon is small, more time is needed to observe the same number of features. Without gaze direction control entropy is not satisfactorily reduced. This results from the fact, that the gaze direction control module chooses to direct the sensor

system mostly towards already observed and more certain features when the environment is known. Therefore, localization error and feature position uncertainties are kept at a minimum.

The results show that the proposed approach combining gaze direction control and motion planning based on information theoretic concepts for the exploration task, gives superior results in comparison to greedy approaches and approaches that neglect active sensor control.

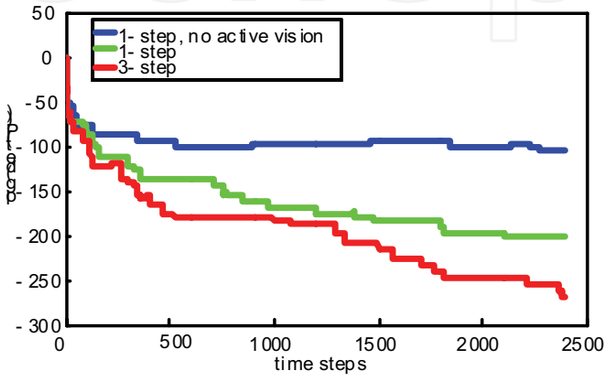


Fig. 13. Entropy measure $\log(|P_{k+1|k+1}|)$ over time for a one-step planning horizon without gaze direction control, one-step and three-step planning horizons with gaze direction control.

6. Conclusions

This chapter presents a foveated active vision planning concept for robot navigation in order to overcome drawbacks of conventional active vision when trading field of view versus measurement accuracy. This is the first approach of task-related control of foveated active vision in the context of humanoid robots as well as localization and mapping. In a typical robot navigation scenario the benefits of foveated active vision have been demonstrated: an improved localization accuracy combined with an extended visible field compared to conventional active vision. As a generic information maximization principle has been used the gaze control strategy is generalizable to other scenarios depending on the definition of the task-related information measures, thereby, allocating vision sensor resources optimally.

The second part of this chapter has presented novel concepts of coordinating gaze direction and locomotion planning. Through this planning procedure, all possible trajectories and viewing angles over a specific time horizon are anticipated. This results in significantly improved performance compared to known approaches, which neglect the limitations of the sensors. It is demonstrated that a very accurate environmental model is autonomously produced even if very inaccurate robot and sensor models are used. Such a scheme is ideal for vision-based humanoid robots.

The proposed approaches assume time-invariant scenarios. Path planning and gaze control methods for dynamically changing environments are not yet covered and subject to ongoing research.

7. Acknowledgments

The authors like to gratefully thank Dr. Javier Seara and Klaus Strobl-Diestro for reference simulation code for performance comparison. This work has been supported in part by the German Research Foundation (DFG) grant BU-1043/5-1 and the DFG excellence initiative research cluster *Cognition for Technical Systems - CoTeSys*, see also www.cotesys.org.

8. References

Apostoloff, N.; Zelinsky, A. (2002). Vision in and out of vehicles: Integrated driver and road scene monitoring, Proceedings of the 8th International Symposium on Experimental Robotics (ISER), 2002, Sant Angelo d'Iscia, Italy

Beis, J. F. & Lowe, D. G. (1997). Shape Indexing Using Approximate Nearest-Neighbour Search in High-Dimensional Spaces, Proceedings of the 1997 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1997

Bodor, R.; Morlok, R. & Papanikolopoulos, N. (2004). Dual-camera system for multi-level activity recognition, Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2004, pp.643-648, Sendai, Japan

Bourgault, F.; Makarenko, A. A.; Williams, S. B.; Grocholsky, B. & Durrant-Whyte, H. (2002). Information Based Adaptive Robotic Exploration, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Lausanne, Switzerland

Bryson, M. & Sukkarieh, S. (2005). An Information-Theoretic Approach to Autonomous Navigation and Guidance of an Unihabited Aerial Vehicle in Unknown Environments, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Edmonton, Canada

Brooks, R. A.; Breazeal, C.; Marjanovic, M.; Scasselati, B. & Williamson, M. M. (1999). The Cog Project: Building a Humanoid Robot, In: Computation for Methaphors, Analogy, and Agents, C. Nehaniv, (Ed.), pp. 52-87, Springer, Heidelberg, Berlin, Germany

Darrell, T. (1997). Reinforcement learning of active recognition behaviors, Interval Research Technical Report 1997-045,http://www.interval.com/papers/1997-045, 1997

Davis, J. & Chen, X. (2003). Foveated observation of shape and motion. Proceedings of the 2003 IEEE International Conference on Robotics and Automation (ICRA), pp. 1001-1005, 2003, Taipei, Taiwan

Davison, A. J. & Murray, D. W. (2002). Simultaneous Localization and Map-Building Using Active Vision, In IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No.7, pp. 865-880, 2002

Davison, A. J. (2003). Real-time simultaneous localization and Mapping with a single camera, Proceedings of the International Conference on Computer Vision, 2003, pp. 1403-1410, Nice

Dickmanns, E. D. (2003). An advanced vision system for ground vehicles. Proceedings of the International Workshop on In-Vehicle Cognitive Computer Vision Systems (IVC2VS), 2003, Graz, Austria

Dissanayake, G.; Newman, P.; Clark, S.; Durrant-Whyte, H. F. & Csorba, M. (2001). A Solution to the Simultaneous Localization and Map Building Problem, IEEE Transactions on Robotics and Automation, 2001, pp. 229-241

Doucet, A.; de Freitas, N.; Murphy, K. & Russell, S. (2000). Rao-blackwellised particle filtering for dynamic bayesian networks, Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, 2000

Elder, J. H.; Dornaika, F.; Hou, B. & Goldstein, R. (2004). Attentive wide-field sensing for visual telepresence and surveillance, In: Neurobiology of Attention, L. Itti, G. Rees & J. Tsotsos, (Eds.), 2004, Academic Press, Elsevier

Eustice, R.; Pizarro, O. & Singh, H. (2004). Visually augmented navigation in an unstructured environment using a delayed state history, Proceedings of the IEEE International Conference on Robotics and Automation, 2004, pp. 25-32, New Orleans, USA

Eustice, R.; Walter, M. & Leonard, J. (2005). Sparse extended information filters: insights into sparsification, Proceedings of the IEEE/RSJ International Conference On Intelligent Robots and Systems, 2005, pp. 641-648, Sendai, Japan

Feder, H. J. S.; Leonard J. J. & Smith, C. M. (1999). Adaptive Mobile Robot Navigation and Mapping, International Journal of Robotics and Research, Vol.18, No.7, pp. 650-668, 1999

Seara, J. F.; Strobl, K. H.; Martin E. & Schmidt, G. (2003). Task-oriented and situation-dependent gaze control for vision guided autonomous walking, Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids), 2003, München and Karlsruhe, Germany

Grabowski, R.; Khosla, P. & Choset, H. (2003). Autonomous exploration via regions of interest, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2003, Las Vegas, USA

Gutmann, J.-S. & Konolige, K. (1999). Incremental Mapping of Large Cyclic Environments, Proceedings of International Symposium on Computational Intelligence in Robotics and Automation (CIRA), 1999, Monterey, CA, USA

Horaud, R.; Knossow, D. & Michaelis, M. (2006). Camera cooperation for achieving visual attention, Machine Vision and Applications, Vol. 15, No. 6, 2006, pp. 331-342

Huang, S.; Kwok, N. M.; Dissanayake, G.; Ha, Q. P. & Fang, G. (2005). Multi-Step Look-Ahead Trajectory Planning in SLAM: Possibility and Necessity, Proceedings of the International Conference on Robotics and Automation (ICRA), 2005, pp. 1103-1108, Barcelona, Spain

Itti, L. & Koch, C. (2001). Computational modeling of visual attention, Nature Reviews Neuroscience, Vol. 2, No. 3, 2001

Jankovic, N. D.; Naish, M. D. (2005). Developing a modular spherical vision system, Proceedings of the 2005 IEEE International Conference on Robotics and Automation (ICRA), pp. 1246-1251, 2005, Barcelona, Spain

Jung, I. & Lacroix, S. (2003). High resolution terrain mapping using low attitude aerial stereo imagery, Proccedings of the IEEE International Conference on Computer Vision, 2003, pp. 946-951

Karlsson, N.; Goncalves, L.; Munich M. E. & Pirjanian, P. (2005). The vSLAM Algorithm for Robust Localization and Mapping, Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2005, Barcelona, Spain

Kim J. & Sukkarieh S. (2003). Airborne Simultaneous Localisation and Map Building, Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2003, Taipei, Taiwan

Koch, C. & Ullmann, S. (1984). Selecting one among the many: A simple network implementing shifts in visual attention. MIT AI Memo No. 770, 1984

Koenig, S. & Tovey, C. (2003). Improved analysis of greedy mapping, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2003, Las Vegas, USA

Kühnlenz, K.; Bachmayer, M. & Buss, M. (2006). A multi-focal high-performance vision system, Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA), pp. 150-155, 2006, Orlando, FL, USA

Kühnlenz, K. (2006). Aspects of multi-focal vision, Ph.D. Thesis, Institute of Automatic Control Engineering, Technische Universität München, 2006, Munich, Germany

Kwok, N. M. & Dissanayake, G. (2004). An efficient multiple hypothesis filter for bearing-only SLAM, Proccedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004, pp. 736-741

Leonard, J. J.; Feder H. J. S. (2000). A Computationally Efficient Method for Large-Scale Concurrent Mapping and Localization, Proceedings of the 9th International Symposium on Robotics Research, 2000, pp. 169-176, Salt Lake City, Utah

Leonard, J. J.; Rikoski, R. J.; Newman, P. M. & Bosse M. (2002) The International Journal of Robotics Research, Vol. 21, No. 10-11, 2002, pp. 943-975

Lidoris, G.; Kühnlenz, K.; Wollherr, D. & Buss, M. (2006). Information-Based Gaze Direction Planning Algorithm for SLAM, Proceedings of IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS), 2006, Genova, Italy

Lidoris, G.; Kühnlenz, K.; Wollherr, D. & Buss, M. (2007). Combined trajectory planning and gaze direction control for robotic exploration, Proceedings of the International Conference of Robotics and Automation (ICRA), 2007, Rome, Italy

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, Vol. 60, No. 2, 2004, pp. 91-110

Lu, F. & Milios, E. (1997). Globally Consistent Range Scan Alignment for Environment Mapping, Autonomous Robots, Vol. 4, 1997, pp. 333-349

Maurer, M.; Behringer, R.; Furst, S.; Thomanek, F. & Dickmanns, E. D. (1996). A compact vision system for road vehicle guidance, Proceedings of the 13th International Conference on Pattern Recognition (ICPR), 1996

Montemerlo, M.; Thrun, S.; Koller, D. & Wegbreit, B. (2002). FastSLAM: A Factored Solution to Simultaneous Localization and Mapping, Proceedings of the National Conference on Artificial Intelligence (AAAI), Edmonton, Canada

Moorehead, S. J.; Simmons, R. & Whittaker, W. L. (2001). Autonomous exploration using multiple sources of information, Proceedings of IEEE International Conference on Robotics and Automation (ICRA), 2001, Seoul, Korea

Ozawa, R.; Takaoka, Y.; Kida, K.; Nishiwaki, J.; Chestnutt, J.; Kuffner, J.; Kagami, S.; Mizoguchi, H. & Inoue, H. (2005). Using visual odometry to create 3D maps for online footstep planning, Proceedings of IEEE International Conference on Systems, Man and Cybernetics (SMC), 2005, pp. 2643-2648, Hawaii, USA

Pellkofer, M. & Dickmanns, E. D. (2000). EMS-Vision: Gaze control in Autonomous vehicles, Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 296-301, 2000, Dearborn, MI, USA

Sabe, K.; Fukuchi, M.; Gutmann, J. S.; Ohashi, T; Kawamoto, K. & Yoshigahara, T. (2004). Obstacle Avoidance and path planning for humanoid robots using stereo vision,

Proceedings of IEEE International Conference on Robotics and Automation (ICRA), 2004, pp. 592-597

Scassellati, B. (1998). Eye finding via face detection for a foveated, active vision system, Proceedings of the 15th National Conference on Artificial Intelligence (AAAI), pp. 969-976, 1998, Madison, WI, USA

Se, S.; Lowe D. & Little J. (2002). Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks, Intl. Journal of Robotics Research, Vol. 21, 2002, pp 735-758

Se, S.; Lowe, D. & Little, J. (2005). Vision-based Global Localization and Mapping for Mobile Robots**,** IEEE Transactions on Robotics, Vol. 21, No. 3, 2005, pp. 364-375.

Shannon, C. E. (1948). A mathematical theory of communication, Bell System Technical Journal, vol. 27, 1948, pp. 379-423 and 623-656

Shibata, T.; Vijayakumar, S.; Conradt, J. & Schaal, S. (2001). Humanoid oculomotor control based on concepts of computational neuroscience, Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids), 2001, Tokyo, Japan

Sim, R. & Little, J. J. (2006). Autonomous vision-based exploration and mapping using hybrid maps and Rao-Blackwellised particle filters, Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS), 2006, Beijing, China

Smith, R. C.; Self, M.; Cheeseman, P. (1990). Estimating Uncertain Spatial Relationships in Robotics, In Autonomous Robot Vehicles, I. J. Cox and G. T. Wilfong, eds., 1990, pp. 167-193, Springer-Verlag

Stachniss, C.; Grisetti, G. & Burgard, W. (2005). Information Gain-based Exploration Using Rao-Blackwellized Particle Filters**,** Proceedings of Robotics: Science and Systems (RSS), 2005, pp. 65-72, Cambridge, MA, USA

Stasse, O.; Davison, A. J.; Sellaouti, R. & Yokoi, K. (2006) Real-time 3D SLAM for humanoid robot considering pattern generator information, Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS), 2006, Beijing, China

Thomson S. & Kagami, S. (2005). Humanoid robot localization using stereo vision, Proceedings of IEEE-RAS Intl. Conf. on Humanoid Robots (HUMANOIDS), 2005, pp. 19-25, Tsukuba, Japan

Thrun, S.; Liu, Y.; Koller, D.; Ng, A. Y.; Ghahramani, Z. & Durrant-Whyte, H. F. (2004). Simultaneous localization and mapping with sparse extended information filters, International Journal of Robotics Research, Vol. 23, No. 7-8, 2004

Ude, A.; Gaskett, C. & Cheng, G. (2006). Foveated Vision Systems with Two Cameras Per Eye, Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA), pp., 2006, Orlando, FL, USA

Yamauchi, B. (1998). Frontier-based exploration using multiple robots, Proceedings of International Conference on Autonomous Agents, pp. 47-53, Minneapolis, USA

Vidal-Calleja, T.; Davison, A. J.; Andrade-Cetto, J. & Murray, D.W. (2006). Active Control for Single Camera SLAM, Proceedings of the International Conference of Robotics and Automation (ICRA), Orlando, USA

Vijayakumar, S.; Inoue, M. & Souza, A. D. (2004). Maveric - Oculomotor experimental vision head, http://homepages.inf.ed.ac.uk/svijayak/projects/maveric/index.html, 2004



Humanoid Robots: New Developments

Edited by Armando Carlos de Pina Filho

ISBN 978-3-902613-00-4

Hard cover, 582 pages

Publisher I-Tech Education and Publishing

Published online 01, June, 2007

Published in print edition June, 2007

For many years, the human being has been trying, in all ways, to recreate the complex mechanisms that form the human body. Such task is extremely complicated and the results are not totally satisfactory. However, with increasing technological advances based on theoretical and experimental researches, man gets, in a way, to copy or to imitate some systems of the human body. These researches not only intended to create humanoid robots, great part of them constituting autonomous systems, but also, in some way, to offer a higher knowledge of the systems that form the human body, objectifying possible applications in the technology of rehabilitation of human beings, gathering in a whole studies related not only to Robotics, but also to Biomechanics, Biomimmetics, Cybernetics, among other areas. This book presents a series of researches inspired by this ideal, carried through by various researchers worldwide, looking for to analyze and to discuss diverse subjects related to humanoid robots. The presented contributions explore aspects about robotic hands, learning, language, vision and locomotion.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Kolja Kuhnlenz, Georgios Lidoris, Dirk Wollherr and Martin Buss (2007). On Foveated Gaze Control and Combined Gaze and Locomotion Planning, Humanoid Robots: New Developments, Armando Carlos de Pina Filho (Ed.), ISBN: 978-3-902613-00-4, InTech, Available from: http://www.intechopen.com/books/humanoid_robots_new_developments/on_foveated_gaze_control_and_combined_gaze_and_locomotion_planning

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2007 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen