Multi-focal Vision and Gaze Control Improve Navigation Performance

Regular Paper

Kolja Kühnlenz and Martin Buss

Institute of Automatic Control Engineering (LSR) ,Technische Universität München, Germany
* Corresponding author E-mail: kolja.kuehnlenz@ieee.org

Abstract Multi-focal vision systems comprise cameras with various fields of view and measurement accuracies. This article presents a multi-focal approach to localization and mapping of mobile robots with active vision. An implementation of the novel concept is done considering a humanoid robot navigation scenario where the robot is visually guided through a structured environment with several landmarks. Various embodiments of multi-focal vision systems are investigated and the impact on navigation performance is evaluated in comparison to a conventional mono-focal stereo set-up. The comparative studies clearly show the benefits of multi-focal vision for mobile robot navigation: flexibility to assign the different available sensors optimally in each situation, enhancement of the visible field, higher localization accuracy, and, thus, better task performance, i.e. path following behavior of the mobile robot. It is shown that multi-focal vision may strongly improve navigation performance.

Keywords multi-focal vision, foveated vision, gaze control, robotics, navigation

1. Introduction

Multi-focal vision systems comprise cameras with various fields of view and measurement accuracies. Typical examples are so called bi-focal or foveated systems with telephoto and wide-angle cameras aligned in parallel. To date only few works exist making use of the particular characteristics of multi-focal vision and approaches requiring locomotion of mobile or humanoid robots have not been proposed yet.

This article proposes a multi-focal gaze control strategy in combination with multi-focal localization and mapping for a mobile robot. An implementation is done considering a humanoid robot navigation scenario. Evaluation studies are conducted investigating the impact of the multi-focal concepts on navigation performance considering various types of multi-focal vision systems in comparison to a conventional mono-focal embodiment.

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.

Multi-focal active vision for mobile robot navigation is considered novel and provides various benefits shown in this article which may strongly improve navigation performance: flexibility to assign the different available sensors optimally in each situation, enhancement of the visible field, higher localization accuracy, and, thus, better task performance, i.e. path following behavior.

The remainder of this article 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 multi-focal camera coordination in localization and mapping with active vision; evaluation studies comparing various multi-focal concepts to conventional approaches and vision systems are presented in Section 4; conclusions are given in Section 5.

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).
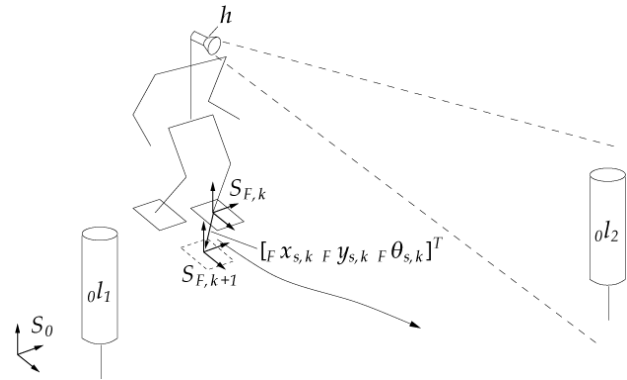

Figure 1. Humanoid robot navigation scenario

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

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

where state-vector $x=[_0x^T \ _Fx_l^T]^T$ contains the robot foot pose estimate $_0x$ in world frame and the landmark position estimates $_Fx_l$ in robot foot frame, d represents system noise capturing dead-reckoning uncertainties, f_s computing the next estimates of the robot pose and landmark positions with respect to the previous ones (basically translation and rotation by u, and adding system noise d), and $g \in \{0; 1\}$ is a binary variable indicating a change of the supporting foot when $g=1$. The commanded step u is expressed by

$$u_k = [_Fx_{s,k} \ _Fy_{s,k} \ _Fq_{s,k}]^T, \qquad (2)$$

including the commanded step position $[_Fx_s \ _Fy_s]^T$ and orientation $_Fq_s$ 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. The landmark positions can be estimated solving a typical 3D reconstruction problem whenever individual landmarks are visible in at least one of the vision sensors yielding a nonlinear measurement equation

$$z_k = f_m(x_k, {}_F P_k, v_{x,k}), \tag{3}$$

with landmark projections in image space x_k, matrix ${}_F P_k$ containing the projection matrices of the individual vision sensors, and sensor noise $v_{x,k}$.

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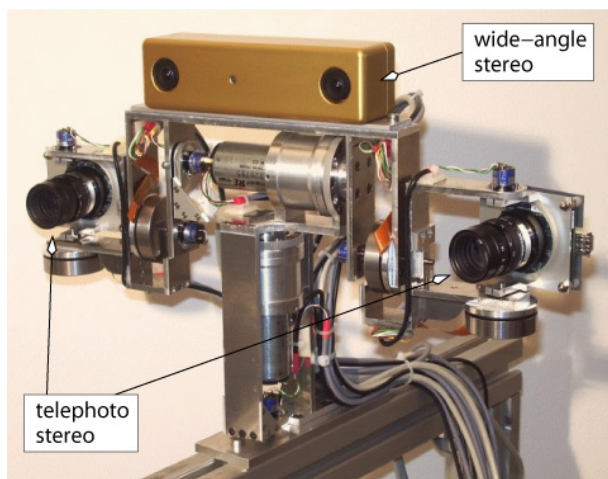
Figure 2. Multi-focal vision system of humanoid *LOLA* (Kühnlenz et el. 2006)

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 multi-focal view direction planning for mobile robots has been investigated in our laboratory which is presented in the following sections.

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

$$v_0 = \frac{1}{2} \sum_i \sqrt{e_i} \; , \qquad (4)$$

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)

$$P^i_{k|k} = \begin{pmatrix} P^i_{uu} & P^i_{um} \\ P^i_{mu} & P^i_{mm} \end{pmatrix}, \qquad (5)$$

where P^i_{uu} is the error covariance matrix of the robot state estimate, P^i_{mm} is the map covariance matrix of the landmark state estimates and P^i_{um} is a cross-covariance matrix between robot and landmark states. Low values of the defined measure (4), 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 article.
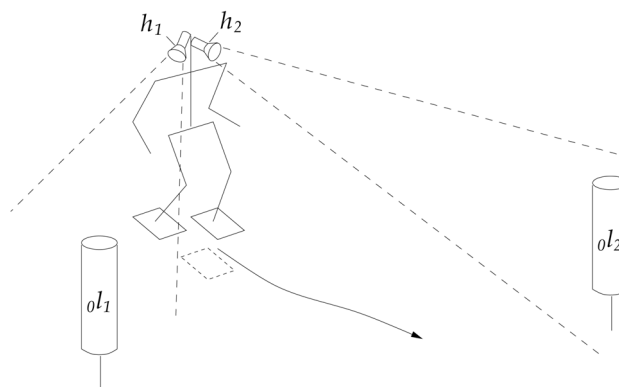

Figure 3. Humanoid robot navigation scenario with (bi-focal) multi-camera vision

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-focal 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 nonlinear measurement equation (3) is extended by allowing for different sensor characteristics, e.g. different projection matrices in $_FP_k$. The basic principle of multi-camera coordination in this article 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

$$\hat{\Omega}^* = \arg\min_{\hat{\Omega}} \hat{\nu}_0 , \qquad (6)$$

where $W=[pan_1 \ tilt_1 \ ... \ pan_n \ tilt_n]^T$ is a configuration of pan- and tilt-angles of all vision devices, n_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. Comparative Evaluation Study

In Section 3 novel concepts of multi-focal gaze control and localization and mapping have been introduced. This section is concerned with comparative evaluation studies in order to assess the performance of the proposed multi-focal approaches in comparison to conventional mono-focal methods and vision systems.

4.1 Scenario Definition: Humanoid Robot Navigation

Considered is a typical locomotion task of a humanoid robot with the robot walking 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 4. 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 state-vector consisting of robot pose and landmark positions is estimated dynamically fusing visual and odometrical information using a Kalman-filter approach depicted in Figure 6 considering the hybrid dead-reckoning model (1) (for detailed description the reader may refer to (Kühnlenz, 2007)). 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 is 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) bi-focal 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) wide-angle stereo-camera with aperture angles of 60°, focal-lengths of 2mm, and a stereo-base of 25cm; d) the bi-focal vision system with aperture angles of the wide-angle cameras 80°; e) two independent stereo-cameras, focal-lengths 2mm and 40mm, respectively, aperture angles 80° 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.
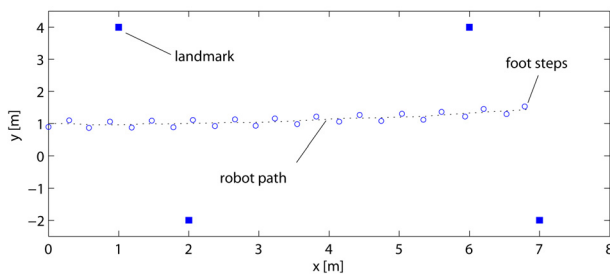

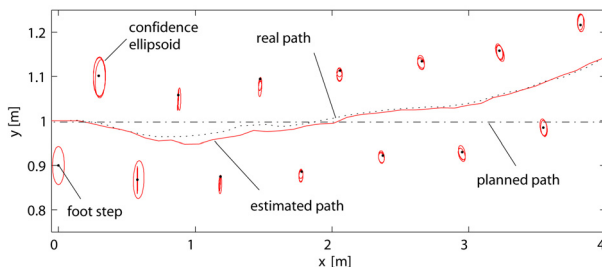**Figure 4.** Top-view of humanoid robot navigation scenario

Figure 5. 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 5 contains a cut-out of Figure 4 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 7 and Figure 8 show the resulting view directions for each step of the robot for the individual vision systems. The resulting fields of view on the ground plane

within the environment are depicted in Figures 9, 10, 11, and 12. The propagations of the areas of the confidence ellipses are shown in Figure 13. 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. The 90%-confidence ellipses for three different vision systems at the same respective step are depicted in Figure 14. These results are discussed in the following.
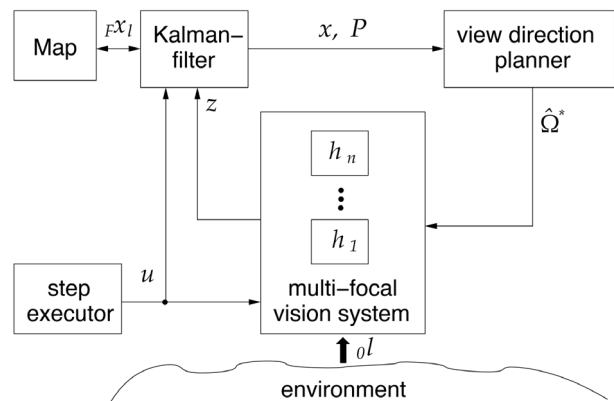

Figure 6. Multi-focal gaze control and simultaneous localization and mapping architecture; state-vector $x=[_Ox^T \ _{FX}l^T]^T$, covariance matrix P, controlled step $_Ou$, landmark positions $_Ol$, estimated landmark positions $_{FX}l$, estimated robot pose $_Ox$, measurements z, vision sensors h_i, and optimized configuration of view directions W^*; $_O(.)$ denoting world frame and $_F(.)$ denoting foot frame
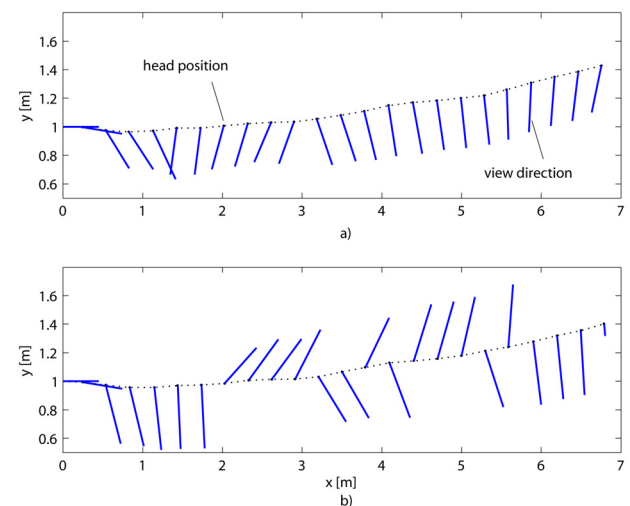

Figure 7. Pan-angles of the a) conventional stereo-camera (system a) and b) of the bi-focal stereo-camera (system b) with two aligned cameras per eye

4.2 Mono-Focal Localization Performance

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 6a 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.

An extreme case vision system which is not shown comprising a telephoto camera with very narrow field of view fails in most cases. If no a priori knowledge on landmark positions is available and no landmark is located within the field of view no information for the view direction planner is available. The only possibility is a continuous scan commanding the camera to rotate over the whole range trusting to accidental detection of the one or other landmark. However, this strategy is highly unefficient and unflexible.

A conventional camera with medium accuracy and focal-length has a better chance to capture landmarks. Figure 7a shows the resulting optimal view directions for this scenario. It can be noted that in this case only one landmark is detected at the starting position (see Figure 9) and only accidentally a second one is seen after shifting the view direction (step #3). Thus, the planner only considers the lower landmarks within the environment. However, most of the time only one landmark is visible.

In consequence moving further through the environment at some position the gaze has to be shifted towards another landmark (due to joint limits) resulting in a sudden increase of the robot position covariances (see Figure 13, x=3m).

Interestingly, the wide-angle system shows a significantly better performance which can be noted by the average areas of the foot step confidence ellipses shown in Table 1. Three landmarks are visible at the initial position and can, thus, be considered for optimal view direction planning. In most cases two landmarks are visible at a time.

4.3 Multi-Focal Localization Performance

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 13 and Table 1.

One of the main contributions and benefits of multi-focal vision is the possibility to allocate sensor resources flexibly depending on environmental conditions and current situation. As can be seen in Figure 11 a variety of different configurations of view directions are selected by the planner in order to satisfy the current situational requirements optimally. At each step measurements of two or three landmarks are taken. The significantly better performance is obvious which is noted by the much lower areas of the footstep confidence ellipses compared to all other vision system configurations, see Figure 12 and Figure 13.

The multi-focal vision systems with aligned cameras, i.e. systems with large field of view and small central high-resolution region, which are used in many state-of-the-art robot heads, however, yield very different performances. The considered foveated scenarios vary in the field of view of the wide-angle camera. Looking at the performance measures in Table 1 the average number of visible landmarks differs by more than one landmark at each step. This results in a much higher localization uncertainty for the vision system with slightly smaller wide-angle field of view shown in Table 1.
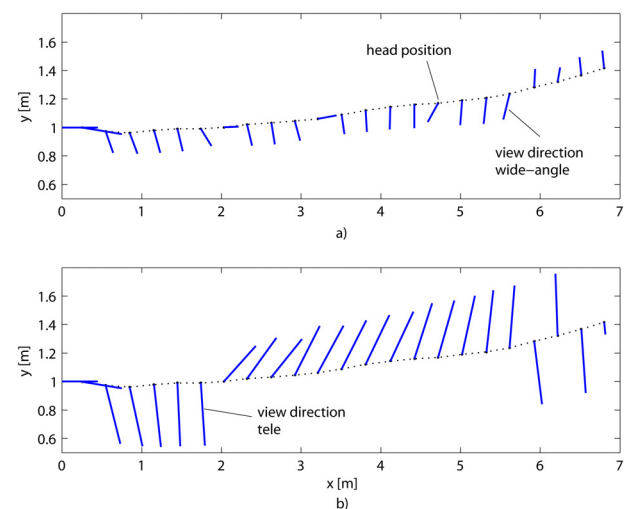

Figure 8. Pan-angles of bi-focal vision system with two independent stereo cameras (system e); a) telephoto and b) wide-angle stereo-camera

The confidence ellipse areas for this multi-focal system only reach the values of a conventional camera (see Figure 13) while the one with the larger field of view performs almost equally to the multi-focal system with independent cameras, see also Table 1. This can be explained by the fact that the wide-angle region is always shifted with the telephoto region. There is no possibility to adjust the wide-angle pose such that a sufficient number of landmarks is captured. Thus, the number of visible landmarks depends

strongly on the field of view of the wide-angle camera. In the scenario with smaller wide-angle field of view only occasionally the wide-angle camera captures more than one landmark which can be seen in Figure 10. At most steps a configuration as seen in step #12 is given. So at many steps merely one landmark is visible.

Assessing these results, the advantages of multi-focal active vision in mobile robot navigation are obvious. Localization accuracy is strongly improved and in case of multi-focal vision with independent cameras sensor resources can be flexibly allocated depending on the current task and situational requirements. In case of multi-focal active vision with aligned cameras a certain field of view has to be provided depending on the environmental setting in order to achieve similar performance.

4.4 Discussion

The concept of multi-focal view direction planning for mobile robot navigation has been introduced in order to overcome various drawbacks of conventional strategies comprising only one type of or kinematically coupled vision devices. The performance of various vision system embodiments has been investigated based on a standard navigation scenario localizing a humanoid robot walking on a straight path through a structured environment. It has been demonstrated that multi-focal vision outperforms all other vision system embodiments due to its flexibility in resource allocation and additional high-accuracy sensors resulting in significantly improved mission performance, i.e. localization accuracy, of the

mobile robot. Foveated devices which are increasingly used perform similarly only if a sufficient field of view of the wide-angle camera is provided.

The main reason for the weaker performance of mono-focal system configurations is the trade-off between field of view and accuracy which substantially reduces the number of reference objects to be observed or the accuracy of object position measurements. A certain field of view is needed in order to continually make out next reference objects along the robot's path strongly limiting accuracy. In the extreme case of a telephoto camera with very narrow aperture angles no objects are detected at all. The foveated version, i.e. a multi-focal system with relative sensor poses fixed, suffers from the shortcoming that the view direction has to be determined based on the foveated region which potentially prevents objects to be detected by the wider-angle region if its aperture angles are too small.

Even though a relatively straight forward perception model has been used generalization to more complex models with distortions, quantization, etc., is possible. However, also in terms of nonlinear distortions a better performance of multi-focal vision can be expected as distortions in higher-resolution devices are comparably small due to narrow aperture angles. In this work only point landmarks have been considered. In order to evaluate the impact of quantization effects on multi-focal vision performance also the impact of image processing algorithms has to be considered.
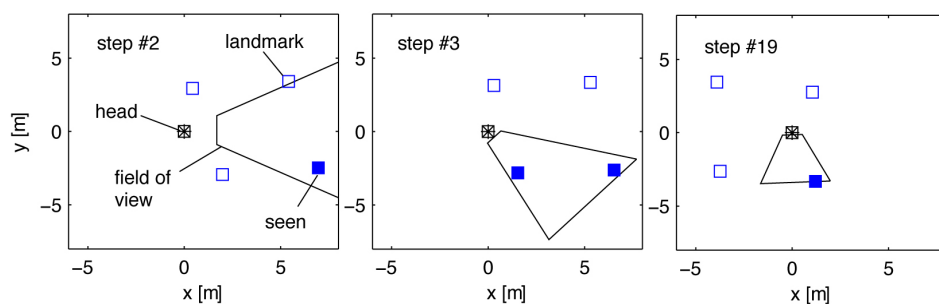

Figure 9. Projections of the field of view of a conventional vision device (system a) of footsteps #2, #3, and #19 of Figure 6 and Figure 9a
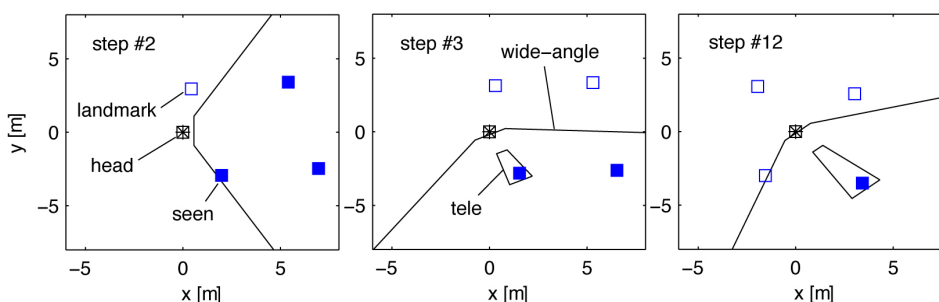

Figure 10. Projections of the field of view of a bi-focal vision device (system b) with cameras aligned in parallel of footsteps #2, #3, and #19 of Figure 4 and Figure 7b
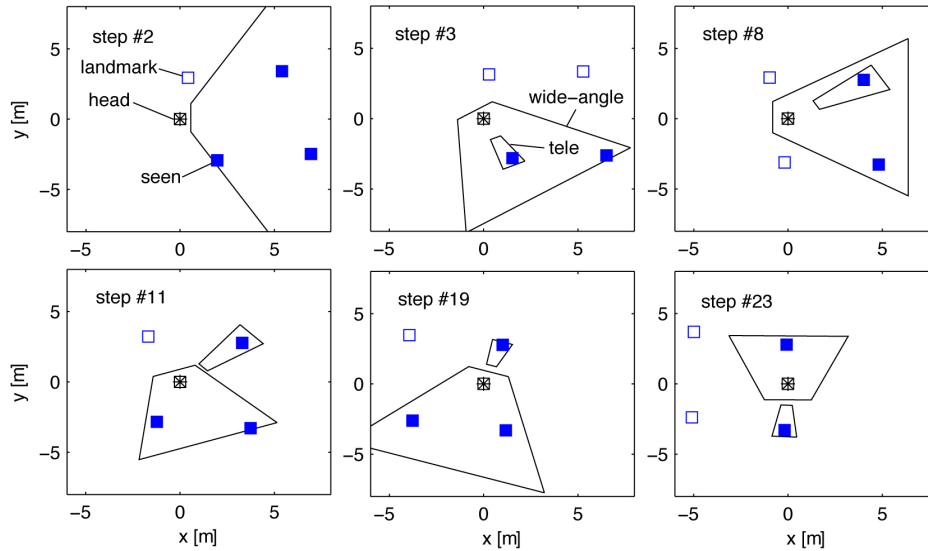
Figure 11. Projections of the field of view of a bi-focal vision device with independent cameras (system e) of footsteps #2, #3, #8, #11, #19, and #23 of Figure 4 and Figure 8
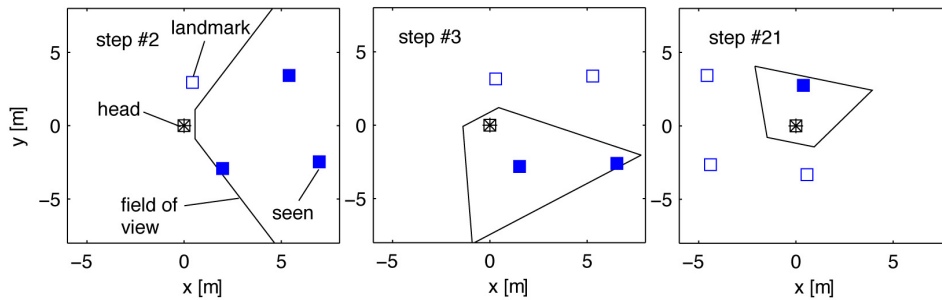

Figure 12. Projections of the field of view of a wide-angle vision device (system c) of footsteps #2, #3, and #21 of Figure 4.
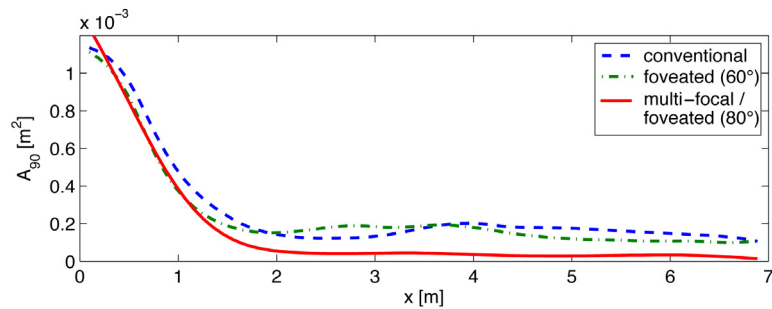

Figure 13. Comparison of the areas A_{90} of the 90%-confidence ellipses of the footstep position covariance matrix using a conventional and a foveated vision system
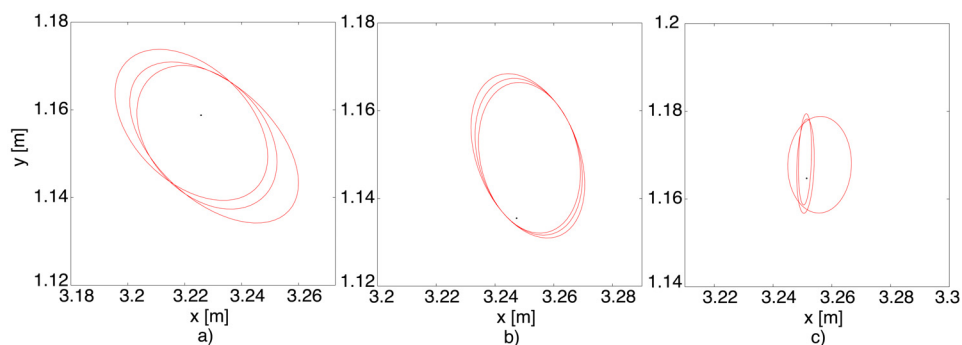

Figure 14. Comparison of the 90%-confidence ellipses of the footstep position estimates at step #12 using a a) conventional (system a), b) wide-angle (system c), and c) bi-focal (system e) vision system with three measurements per step

Vision system	A_{90} [10^{-4}m^2]	average number of visible landmarks
a) conventional	2.7	1.1
b) bi-focal (60°)	2.4	1.3
c) wide-angle	2.1	1.8
d) bi-focal (80°)	1.7	2.3
e) bi-focal (ind.)	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. Conclusions

This article presents a multi-focal 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 multi-focal, respectively, foveated active vision in the context of mobile robots, humanoid robots as well as localization and mapping. In a typical robot navigation scenario the benefits of multi-focal active vision have been demonstrated: an improved localization accuracy combined with an extended visible field compared to conventional active vision, and higher flexibility of sensor resources allocation. 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.

6. Acknowledgments

7. References

[1] 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

[2] 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

[3] 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

[4] 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

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

[6] 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

[7] 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

[8] 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

[9] 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

[10] 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

[11] 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

[12] 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

[13] 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

[14] 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* (IROS), 2005, pp. 641–648, Sendai, Japan

[15] 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

[16] 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

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

[18] 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

[19] 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

[20] 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

[21] 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

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

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

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

[25] Kühnlenz, K., Lidoris, G., Wollherr, D. & Buss, M. (2007). On Foveated Gaze Control and Combined Gaze and Locomotion Planning, *In Humanoid Robots: New Developments*, Obinata, G. & Dutta, A, eds., 2007, Advanced Robotic Systems International and I-Tech, Vienna, Austria.

[26] 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

[27] 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

[28] 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

[29] 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

[30] 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

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

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

[33] 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

[34] 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

[35] 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

[36] 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

[37] 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

[38] 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

[39] 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

[40] 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.

[41] 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

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

[43] 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

[44] 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

[45] 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

[46] 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

[47] 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

[48] 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

[49] 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

[50] 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