Gaze Control for Goal-Oriented Humanoid Walking

J. F. Seara, O. Lorch, and G. Schmidt

Institute of Automatic Control Engineering,
Technische Universität München,
D-80290 Munich, Germany
eMail: Javier.Fernandez.Seara@ei.tum.de

Abstract

In this article a predictive task-dependent gaze control strategy for goal-oriented humanoid walking is presented. In the context of active vision systems we introduce an information theoretical approach for maximization of visual information. Based on two novel concepts, *Information Content* of a view situation and *Incertitude*, we present a method for selecting optimal subsequent view directions, thus contributing to an improved performance in typical semi-autonomous robot locomotion tasks. Simulations and experimental results dealing with the duality of concurrent tasks during locomotion, i.e. obstacle avoidance and self localization, demonstrate the applicability of our approach on humanoid walking machines.

Keywords

Gaze Control, Active Vision, Vision Guided Humanoid Walking.

1 Introduction and Research Directions

On Humanoids 2000 [8] we presented the ViGWaM emulation environment to support our theoretical studies on the interaction between vision and biped locomotion. This framework allows the validation and extrapolation of our vision-based control approaches to a varity of walking machines. Figure 1 shows the simplified overall control and emulation architecture and its coupling with the experimental set-up. Data from real visual perception and image processing are coupled with ViGWaM — **Vi**sion-**G**uided **V**irtual **Wa**lking **M**achine. Its operation is visualized in an augmented reality display, where an image of an external camera observing the real scenario is overlayed by a virtual 3D graphic animation of ViGWaM.

Theoretical investigations carried out in our group can be subdivided in three major directions: (i) development of strategies for *walking pattern primitive synthesis and concatenation* to a stable step sequence allowing situation-dependent walking, see [4]; (ii) algorithms for robust *image processing* and obstacle reconstruction considering the head motion in 6 DoFs during locomotion; (iii) biologically motivated studies on intention based gaze control with the objective to increase visual information in a task and goal dependent way [5]. In this article we present our approach to predictive task-dependent view direction control.

The article is organized as follows: Section 2 presents the "*where to look next?*" problem in the framework of active vision systems, the information theoretical approach taken and its corresponding mathematical formalization. A general description of the proposed predictive gaze control strategy is presented in Section 3 with the corresponding simulation results. Section 4 summarizes the benefits of the approach for the active vision system of ViGWaM. The article is concluded in Section 5 with an outlook on possible improvements in gaze control obtained by the proposed approach.

2 Active Vision and Gaze Control

Vision is one of the most powerful sensor modalities both for biological and artificial systems. Vision allows them to gather simultaneously "*what*" and "*where*" information about their environments. For biological agents, like humans or animals, vision proves to be the main aid to obstacle avoidance, localization, and navigation. Most of these visually guided locomotion "specialists" make use of an active vision system. An active vision system is a system capable of adjusting its visual parameters in a controlled manner to extract relevant information about the scene in space and time [2]. Figure 2 depicts the vision architecture of ViGWaM with the gaze control blocks highlighted. Image processing algorithms are adapted to the time varying task-depended perception re-

Proceedings of the IEEE-RAS International Conference on Humanoid Robots

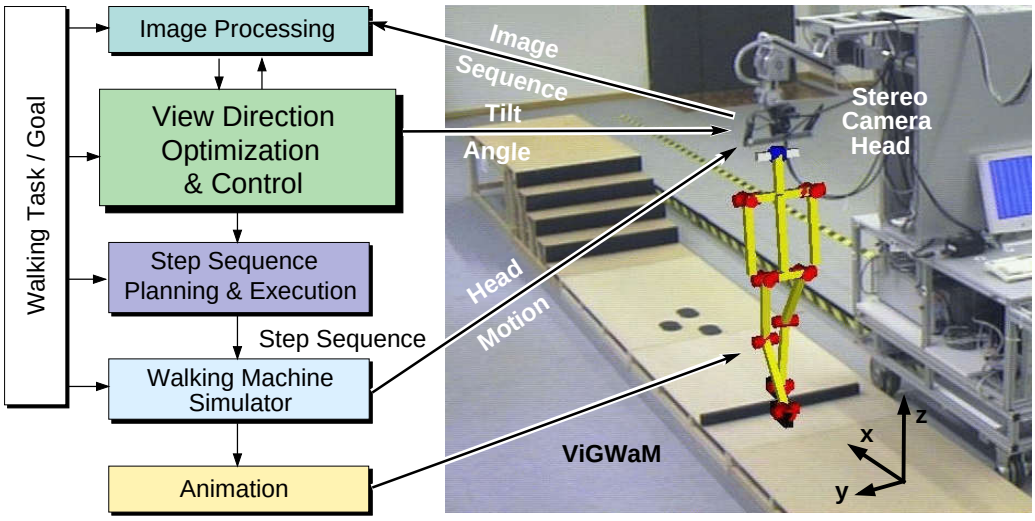

Figure 1: Emulator architecture and experimental set-up with ViGWaM overlayed.

quirements. This is mirrored in the cascaded vision architecture presented here. The results of the image processing are fused with a 3D-map and are employed for step sequence adaptation and gaze control.

Since in biologically inspired artificial systems the field of view is typically limited, a means of directing the visual sensor is required. This can be accomplished by a gaze control scheme which determines the orientation of the sensor, so that the relevant objects of the scene are within the field of view. Thus, the primary task of the gaze controller is to direct the visual sensor (*overt* attention in biological systems) rather than to determine the high saliency regions in a given scene (*covert* attention). A gaze controller accomplishes the task to determine "*where to look next?*", which is an intention/mission problem.

Quite a few articles have been presented on the topic of view direction control, see [3, 6]. However the "*where to look next?*"–problem has been treated so far in a rather limited way. The goal was to determine the most conspicuous point or feature within an image in order to direct attention towards it, i.e. shifts on *covert* attention. In this article we present a novel approach to the problem with a focus on increasing system autonomy by task-oriented maximization of visual information.

2.1 Information Theoretical Approach

We consider a framework based on Shannon's Information Theory [9] for selecting the camera parameters that maximize the visual information, i.e. task-oriented optimization of the information that the captured image conveys about the state of the observer and environment. The maximization of visual information implies an information quantification (measure of information). Our approach to this quantification is related to the definition of the concept of *Information Content*.

A well-known definition of information based on Shannon's Theory is: *Information is a measure of the decrement of uncertainty*. To our best knowledge there has not been made an information theoretic treatment of how uncertainty can be employed in robotics without following a probabilistic analysis. Uncertainty is a concept normally applied in a probabilistic context (Kalman filtering, Markov models, etc., see [6] for example). To point out that the novel approach presented here does not need an exhaustive probabilistic analysis of the visual sensor capabilities in order to come up with a task-dependent optimal gaze behavior, we will refer to it with the notion *Incertitude*.

We define *Incertitude* as a measure of the uncertainty (lack of reliable knowledge) related to the total amount of information necessary for the successful fulfillment of a given task. When expressing this in percentage, 0% *Incertitude* means an exact and sufficient amount of task-related knowlegde, and 100% *Incertitude* means a total lack of useful information for a given task. We define the *Information Content* of a view situation, i.e. environment status and sensor position and orientation, as the measure of how perception under these terms is capable of reducing *Incertitude*.

Applying these definitions within the framework of active vision systems, we present a methodological approach to quantify the *Information Content* in

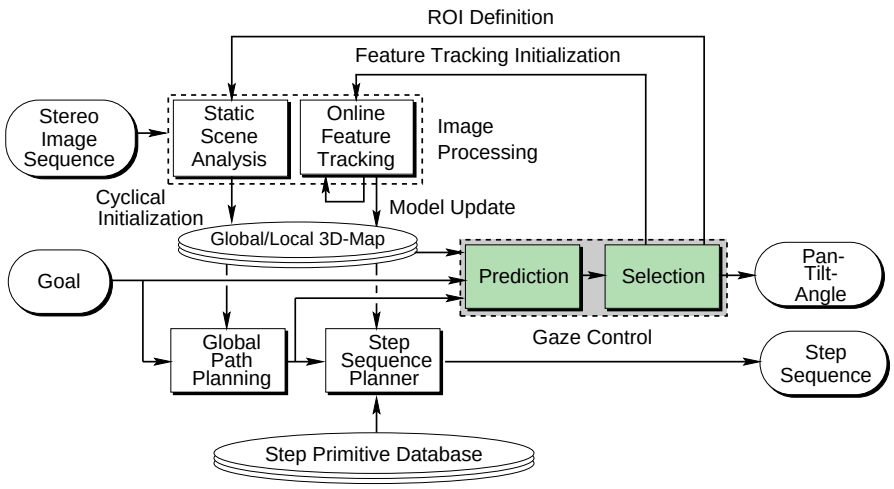

Figure 2: Cascaded vision architecture with gaze control highlighted.

a view situation, i.e. quantification of the *Incertitude* variations in a specific perception status. We assume that the instantaneous visual perception of a scenario by the observer does not reduce its *Incertitude* to zero, i.e. there exist uncertainties in the measurements. It is noteworthy that this approach comes to stand in the general context of active robot perception and requires therefore a geometrical interpretation. Thus the final goal is to improve an agent's performance in different autonomous locomotion tasks such as obstacle avoidance or navigation. The *Information Content* quantification in a view situation is explained as follows:

- The visual perception of objects in the observer's immediate vicinity leads to a substantial reduction in the *Incertitude* of the type, location, and other features of these objects, which can be estimated by means of visual perception. A view of a nearby object implies a richer *Information Content* than the view of a distant one under the same perception conditions. Obviously, quality of measurement decreases with distance as close objects can be sensed with higher resolution. An approximation for this dependency can be given by the proportionality to the inverse of squared object distance, see [7, 8].

- The visual perception of an object projecting on a central region in the visual field yields a greater reduction/decrement in the observer's *Incertitude* than the perception of objects on the image's outer parts (*Fixation* is one of the more important tools in active vision). The *Information Content* of those vision situations is richer due to the improvement in the image quality regarding our purpose. There is less unmodeled optical distortion because only the area near the optical axis is used and the features will be imaged near the principal points. Moreover, when an object projects on the image center, the probability of lying completely within the image boundaries is higher and analysis more advantageous.

- The richness of the *Information Content* in a specific view situation depends not only on the above mentioned aspects but also on the scene perception time history. This is the history of gaze points and the accumulated useful information, which means the dynamical evolution of the whole process, i.e. the current *Incertitude* in the present situation.

2.2 Mathematical Formalization

In this section we present the mathematical formalization of the above presented concepts. We represent the pose, i.e. position and orientation, of the active vision system in a global reference frame as

$$\begin{aligned} \mathbf{\Omega} &= \left[\mathbf{\Omega}^{B^T}\ \mathbf{\Omega}^{Ob^T}\right]^T \\ &= \left[[q_1^B \dots q_6^B]\ [q_1^{Ob} \dots q_n^{Ob}]\right]^T, \end{aligned} \tag{1}$$

where $\mathbf{\Omega}^B$ is the pose of the base reference frame (B) of the active vision system. $\mathbf{\Omega}^{Ob}$ describes the present pose of the observer (Ob) relative to B and is related to its internal structural configuration. The number of degrees of freedom n of a e.g. stereo active vision system is $1 \leq n \leq 12$, being for example $n = 3$ in a stereo *pan-tilt-roll*

camera head without vergence control. The case $n = 12$ would indicate independent relative motion between both cameras and between cameras and base reference frame, which is a rare case.

We assume the existence of an initial amount of knowledge about the scene (i.e. statically pre-acquired information which can be used to construct an initial 3D-geometrical map): pose information about the objects (index i), i.e. the position (index p) and orientation (index r) of the objects, in the scenario considered in the task, $\boldsymbol{\Upsilon}_i = \left[\boldsymbol{\Upsilon}_{p,i}^T \;\; \boldsymbol{\Upsilon}_{r,i}^T\right]^T$, where $\boldsymbol{\Upsilon}_{p,i}, \boldsymbol{\Upsilon}_{r,i} \in \mathbb{R}^3$ and the initial values of a parameter set which determines the degree of *Incertitude* ν_i.

In order to describe the perception process we need to establish a model of the vision system. This model defines the mapping between a 3D-point in the environment, $\boldsymbol{\Upsilon}_i$, and its corresponding projected 2D-image point, $\boldsymbol{m}_i = [u_i \quad v_i]^T$. This imaging relationship can be modeled as $\boldsymbol{m}_i = \mathcal{H}(\boldsymbol{\Omega}, \boldsymbol{\Upsilon}_i)$.

The individual *Information Content* ($\mathcal{IC}_i$) due to the presence of one object i in the field of view in a specific view situation at time t is given by

$$\mathcal{IC}_i(t) = \mathcal{F}(t, \boldsymbol{\Omega}, \boldsymbol{\Upsilon}_i, \nu_i) \tag{2}$$

and the total *Information Content* ($\mathcal{IC}$) in a specific view situation in the presence of N visible objects is given by

$$\mathcal{IC}(t) = \sum_{i=1}^{N} \mathcal{IC}_i(t). \tag{3}$$

The function $\mathcal{F}$ in (2) can then be expressed according to the definitions presented in the previous section as

$$\mathcal{F}(t) = \mathcal{F}\left(f_1(t, \nu_i), f_2(t, \boldsymbol{d}_i), f_3(t, \boldsymbol{m}_i)\right), \tag{4}$$

where f_1 represents the influence of the perception time history defined above, f_2 the influence of the relative position between observer and object i which is $\boldsymbol{d}_i = h(\boldsymbol{\Omega}, \boldsymbol{\Upsilon}_{p,i})$, and f_3 represents the influence of the projecting position of a feature on the image.

With this notation, we can express the relation between *Information Content* and variation of *Incertitude* ν_i as

$$\dot{\nu}_i(t) = -\,C_1 \cdot \mathcal{IC}_i(t) + C_2 \cdot \mathcal{I}_{SM}(t), \tag{5}$$

where C_1 and C_2 are positive constant weighting factors and the term $\mathcal{I}_{SM}$ represents the increment of *Incertitude* due to the observer's self motion.

3 Predictive Gaze Controller

3.1 Gaze Control Algorithm

As discussed in Section 1, when working with an active vision system a view direction control algorithm must be employed to determine how to direct the sensor so that the interesting objects of the scene are within its field of view (Fixation). A simple reactive controller neglecting the delay due to the non-zero image acquisition and processing time, seems therefore inappropriate for active vision application which require not only a reactive component but also a purposive task-dependent active behavior. Prediction proves to be an appropriate method to improve performance in such systems. Biological experiments have proven that humans make also heavy use of internal prediction [1]. The human brain uses pre-stored information for navigation and guidance. This information, for example the approximate position of an oncoming obstacle on a walking trail, helps to direct the visual sensor. This is the biological basis of the predictive component incorporated in the gaze controller described here.

With the earlier defined concepts of *Information Content* and *Incertitude* we present a predictive task-dependent gaze control strategy for active vision systems based on maximization of the predicted visual information. The approach is as follows: Using the *a priori* available information about the scene (for example a 3D-map, i.e. $\boldsymbol{\Upsilon}_i$ and $\nu_i(t)$), the current observer location and actual motion parameters ($\boldsymbol{\Omega}(t)$ and $\dot{\boldsymbol{\Omega}}(t)$), the gaze controller predicts the (quasi-)optimal pose of the visual sensor for a future view situation $\widehat{\boldsymbol{\Omega}}_\star^{Ob}$ (predicted variables are denoted as $\widehat{\ }$ and optimal ones by index $\star$). The optimal pose is found by selecting the state of maximum *predicted Information Content*, $\widehat{\mathcal{IC}}_{max}$, among a set of **P**redicted **S**tates (*PS*). This can be expressed as:

$$\widehat{\mathcal{IC}}_{max} = \max_{\widehat{\boldsymbol{\Omega}}^{Ob}} \sum_{i=1}^{N} \widehat{\mathcal{IC}}_i(\widehat{\boldsymbol{\Omega}}, \boldsymbol{\Upsilon}_i, \nu_i) \quad \text{and} \tag{6}$$

$$\widehat{\boldsymbol{\Omega}}_\star^{Ob} = \arg\max_{\widehat{\boldsymbol{\Omega}}^{Ob}} \sum_{i=1}^{N} \widehat{\mathcal{IC}}_i(\widehat{\boldsymbol{\Omega}}, \boldsymbol{\Upsilon}_i, \nu_i), \tag{7}$$

both subject to

$$\widehat{\boldsymbol{\Omega}}_{min}^{Ob} \leq \widehat{\boldsymbol{\Omega}}^{Ob} \leq \widehat{\boldsymbol{\Omega}}_{max}^{Ob} \quad \text{and} \tag{8}$$

$$\boldsymbol{g}(\dot{\boldsymbol{\Omega}}^{Ob}) = \boldsymbol{0}, \tag{9}$$

where (8) are upper and lower bounds for the active vision system parameters and (9) guarantees

a steady gaze behavior without abrupt view direction changes. The latter would negatively influence the perception process. Equation (9) takes into consideration the observer's system dynamics and represents some sort of penalty function which limits the performance (e.g. speed) of saccades.
A predicted state is therefore specified by the predicted future pose of the observer and the variable value of sensor configuration parameters (for example *pan-tilt*-angles), which determine the field of view. The *Information Content* of a given predicted state is quantified, depending on the current goal according to the relationships described in Section 2.1.
According to (5), the predicted *Incertitude* $\widehat{\nu}$ in the system expressed in discrete time k evolves as follows

$$\widehat{\nu}_i(k{+}1) = \nu_i(k) - C_1 \cdot \widehat{\mathcal{IC}}_i(k) + C_2 \cdot \widehat{\mathcal{I}}_{SM}(k), \quad (10)$$

where $\widehat{\mathcal{IC}}$ and $\widehat{\mathcal{I}}_{SM}$ are the predicted values for the *Information Content* and the predicted gain of *Incertitude* due to the observer's self motion.

3.2 Implementation and Simulation

In order to demonstrate the validity of our approach, we present an example application in which the advantages of using the before presented gaze controller in an arbitrary active visual system are pointed out compared to the traditional passive approach.
The active observer's goal in this example is to gather as much task-oriented information as possible. The two tasks considered are typical problems in autonomous robot navigation and guidance: self localization and obstacle avoidance. While performing the two tasks, the active observer considers two different types of *Information Content*, i.e. two different *Incertitudes*. These are here noted as *Global Incertitude*, ν^G, (referring to self localization) and *Local Incertitude*, ν_i^L, (referring to obstacle positions, where i denotes the number of existing obstacles). The main goal of the active system is therefore to gather a sufficient amount of task-oriented information to keep these two *Incertitudes* within certain safety margins. These margins guarantee that a sufficient amount of information is available for the safe guidance and obstacle avoidance of the observer.
In this experiment we assume an active vision system, which moves in a finite 3D-space with obstacles and landmarks. At the start situation we assume a quite exact knowledge about the observer's pose in space, $\mathbf{\Omega}^{init}$, in this case 30% *Global Incertitude*, and a certain amount of knowledge about the pose of the obstacles in this scenario, $\mathbf{\Upsilon}_i^{init}$, but maximum (100%) *Local Incertitude*, see *Incertitude* initial values in Figure 6.
Furthermore, we assume that the active observer's position remains constant over the ground and in lateral direction: The observer moves only in x-direction, cf. Figure 3, with speed b, i.e. $\mathbf{\Omega}^B = [q_1^B \;\; a_2 \;\; a_3 \;\; 0 \;\; 0 \;\; 0]^T$, with $\dot{q}_1^B = b > 0$ and a_2, a_3, and b constants, see ground trajectory projection in Figure 3. The active observer has only two rotational degrees of freedom: *pan* and *tilt*, i.e. $\mathbf{\Omega}^{Ob} = [q_1^{Ob} \;\; q_2^{Ob}]^T = [pan \;\; tilt]^T$. Note that this assumption does not constrain the applicability of the proposed approach to other hardware set-ups and that this *pan-tilt* configuration is very common in robotic systems with active vision.

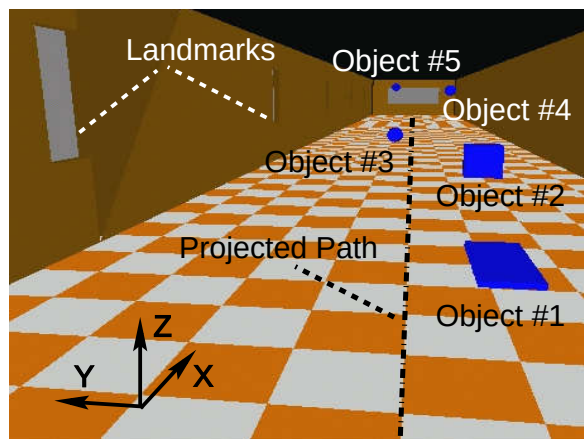

Figure 3: Simulated scenario with landmarks and typical obstacles.

The heuristically determined function $\mathcal{F}$ used for calculating the *Information Content* in a specific view situation is given by

$$\begin{aligned} \widehat{\mathcal{IC}}_i &= \mathcal{F}\left(f_1(\nu_i), f_2(\widehat{d}_i), f_3(\widehat{\boldsymbol{m}}_i)\right) \\ &= \frac{\nu_i}{\widehat{d}_i^2} \cdot \left[\cos\left(\frac{\widehat{\theta}_i}{\alpha}\right) \cdot \cos\left(\frac{\widehat{\varphi}_i}{\alpha^*}\right)\right], \end{aligned} \quad (11)$$

where α is the camera horizontal view angle, α^* is the camera vertical view angle, $\widehat{d}_i$ is the predicted relative distance to obstacle i, and $\widehat{\theta}_i$ and $\widehat{\varphi}_i$ are the predicted projection angles of the feature reference point on the image plane. These variable values are calculated by

$$\widehat{\boldsymbol{m}}_i = [\widehat{\theta}_i \;\; \widehat{\varphi}_i]^T = \mathcal{H}(\widehat{\mathbf{\Omega}}, \mathbf{\Upsilon}_i^p), \quad (12)$$

$$\widehat{d}_i = h(\widehat{\mathbf{\Omega}}, \mathbf{\Upsilon}_i^p). \quad (13)$$

Figure 4 shows the function f_3 (4) for a horizontal camera view angle $\alpha = 60°$. This function represents the influence of the projecting position of

a feature on the image. Notice that the distribution representing this influence can be interpreted as an approximate model of a biologically *foveated retina*.

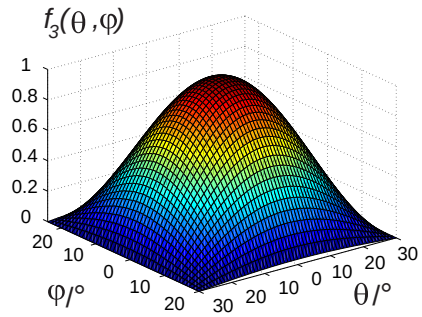

Figure 4: Influence of a feature's projecting position on the image on the *Information Content*.

The evolution of *Global (G) and Local (L) Incertitude* is specified by the following equation

$$\nu_i^{G/L}(k+1) = \nu_i^{G/L}(k) - C_1 \cdot \mathcal{IC}_i(k) + C_2 \cdot \mathcal{I}_{SM}(k) \quad (14)$$

with C_1 and C_2 denoting adequately chosen weighting constants. This pair of constants has different values for *Global Incertitude* ν^G and *Local Incertitude* ν^L. In this simulation the increment of *Incertitude* due to the observer's self motion, $\mathcal{I}_{SM}$, is a function of the covered distance that exponentially approaches the upper bound of 100% *Incertitude*. The prediction interval is determined by the processing time required by the visualization task.

3.3 Simulation Results

Figure 5 shows a top (upper figure) and side view (lower figure) of the gaze behavior resulting from the simulation according to the scenario presented in Figure 3. Five landmarks (□) and five objects (○), represented for this example by their center of gravity as point-objects, are present. Objects #1 to #4 are static while #5 is moving with constant velocity during the interval $t_1 \leq t \leq t_2$ along the marked path (○ ▷ ▷ ▷ ○), see Figure 5. The gaze behavior is represented as a vector field indicating the view direction while the observer is moving through the scenario. Some properties of the gaze controller can be analyzed based on this figure.

It is obvious that the observer's attention is attracted by close objects. When moving the observer's *Incertitude* about its position increases and self localization is needed. Therefore, the view direction is selected to point towards a landmark after some distance, see the results for *Global Incertitude* in Figure 6. It can also be seen how the

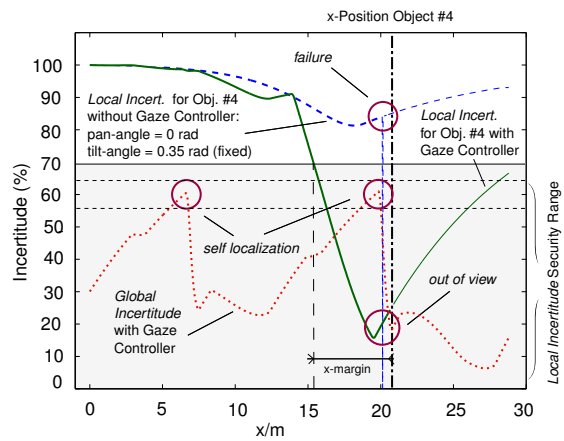

Figure 6: *Local Incertitude* results with (—) and without (- - -) gaze control for single lateral object and *Clobal Incertitude* results with gaze control (· · ·).

object in motion is followed by the gaze in order to gather the maximum information about it.

Figure 6 presents some results of *Local Incertitude* for one of the objects (#4) obtained in the same scenario and the comparison of the two before mentioned approaches: fixed gaze strategy and active vision approach with predictive gaze controller, which corresponds to the gaze behavior described in Figure 5. In this example the fixed gaze approach is defined by two constant angles: *pan*-angle $= 0.0\,rad$ and and *tilt*-angle $= 0.35\,rad$ (visual sensor slightly pointing down). This visual configuration seems to be a reasonable compromise among the feasible fixed gazes for this scenario. The *Local Incertitude* values reached are comparatively high compared to other configurations, i.e. in this situation objects remain within the field of view for a long time.

This figure shows also the security ranges considered. This means that when passing an object its corresponding *Local Incertitude* must be in this area. This property is evaluated just before the observer reaches the object position. If this value is outside the security range considered, the simulation indicates a failure situation as it is shown in the figure. On the other hand can it be observed that in this experiment this failure does not occur when applying gaze control.

Figure 7 shows the results obtained for the five objects positioned in different configurations (Object #1 to #5). In all five cases the results are better for the gaze controller approach: The *Incertitude* values are lower and the objects remain within the field of view for a longer period of time, as indi-

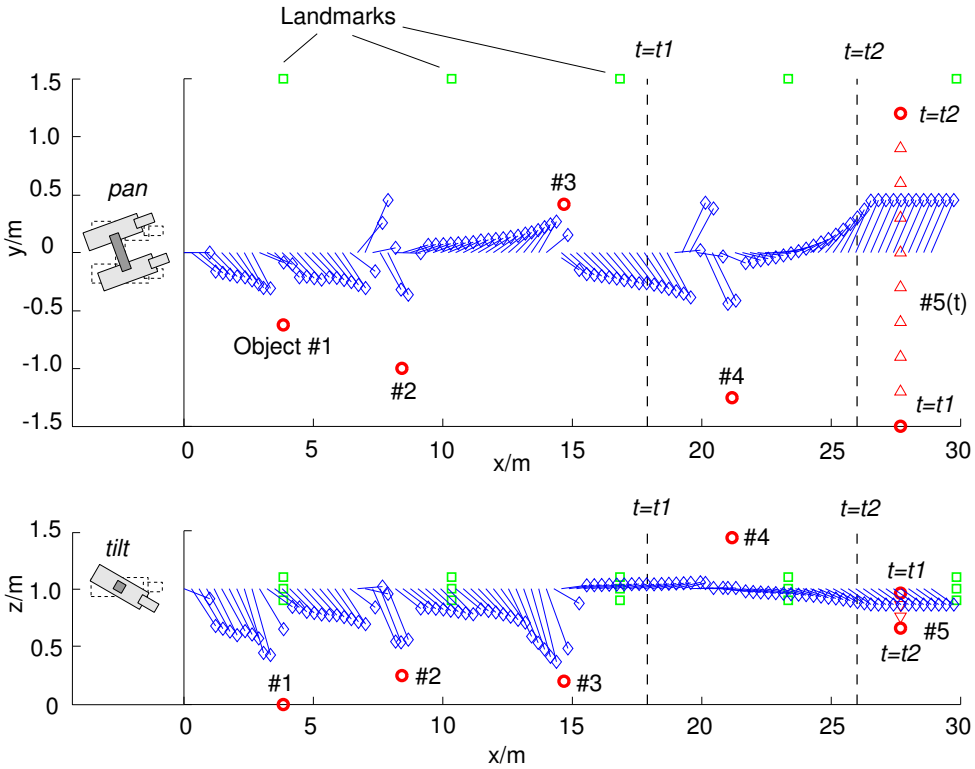

Figure 5: Top view (upper figure) and side view (lower figure) of the resulting gaze behavior.

cated by the global minima in Figure 7. This is especially distinctive when the object is located quite aside from the observers path, see Object #4 in Figure 5.

It is obvious that the corresponding results for *Global Incertitude* are much better with the gaze controller than without due to the lack of direct fixation in the fixed gaze approach. In addition, it can be observed how the *Local Incertitude* curves for the active vision system with gaze controller present non smooth regions, which correspond to the fact that in these regions the gaze is directed towards a landmark to assure the *Global Incertitude* values remaining below the security margin, see Figure 6.

4 Experiments with ViGWaM

4.1 Hardware Setup

A reduced version of this goal-oriented predictive gaze controller has been implemented in the active vision system of ViGWaM [8]. In that article we proposed an emulation environment that allows validation of our vision-based control approach for humanoid walking. In this emulation environment, see Figure 1, a set of stereo cameras mounted on a *pan-tilt*-head can be moved by use of a mobile platform over a prototypical scenario. The head motion is synchronized with a dynamically simulated walking machine to achieve realistic perception conditions. The cameras can travel over the scenario and emulate the ViGWaM head motion in 6 DoFs.

4.2 Results on Gaze Control Behavior

In this new experiment, the mission of ViGWaM is constrained to walking straight ahead towards a staircase in a prototypical sidewalk scenario, not only avoiding collision with a barrier, but keeping its *Global Incertitude* ν^G (referring to self localization, here only one landmark is considered for this purpose) within a safety margin, cf. Figure 8. In this application the time evolution of the *Incertitude* must be considered. Now, the task of the gaze controller is not only the instantaneous minimization of the *Local Incertitude* with respect to the location of the obstacles as in [8]. The objective of gaze control is to ensure that objects of interest remain within the field of view as long and often as necessary for step sequence adaptation and gather enough information for navigation, cf. Section 3.2. In our experimental setup, the camera pair is mounted on a *pan-tilt*-head. This means that the gaze controller possesses of two variables *pan*- and *tilt*-angle, $\mathbf{\Omega}^{Ob} = [q_1^{Ob}\ q_2^{Ob}]^T = [pan\ tilt]^T$, to control perception. In this case the position and ori-

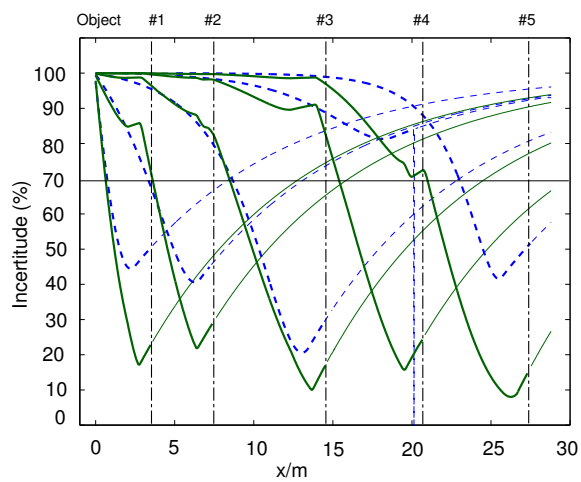

Figure 7: *Local Incertitude* results with (—) and without (- - -) gaze control for the five objects in scenario.

entation of the observer's base reference frame $\mathbf{\Omega}^B$ is defined by the movement of the platform.
The type and dimensions of the obstacles are visually determined in an initialization process with an individual degree of *Incertitude*. The *Information Content* of a given state is quantified, depending on the current goal, taking into account the number of objects in view, their distance to the walking machine, their relative position to the center of the camera view, and the present *Global Incertitude*. The term $\mathcal{F}$ used to calculate the *Information Content* in a view situation for a predicted state (PS) is given by

$$\widehat{\mathcal{IC}}_{PS}(\mathbf{\Omega}^{Ob}) = \mathcal{F}(f_1(\nu^G), f_2(\widehat{d}_i), f_3(\widehat{\boldsymbol{m}}_i)). \quad (15)$$

The (quasi-)optimal *pan-tilt*-angles are obtained by maximizing (15) for future locations of the observer. Figure 9 shows the resulting view directions adjusted by the gaze controller during motion. Images from the upper left to lower right corner were taken in equally spaced intervals of $0.2\,m$ from the left eye. The distance to the barrier from our starting point is $1.5\,m$ and to the staircase $3.8\,m$. The landmark is located at a distance of $7.9\,m$ ahead and $4.2\,m$ to the right of the observer's initial position. It can be seen how the current relevant objects remain focused while being analyzed and how the gaze is modified according to the locomotion task.

5 Conclusions and Future Work

In this article we have presented a novel biologically inspired predictive gaze control strategy for active vision systems. Our information theoretical approach leads to the definition of two new concepts: *Information Content* and *Incertitude*, together with the corresponding mathematical formalism stated. Based on these concepts a systematic approach to the development of a task-dependent gaze control strategy was proposed. Simulations have shown the validity of the method for visual information maximization for robot navigation and obstacle avoidance tasks. The experimental results achieved in an emulation environment demonstrate the applicability and efficiency of the proposed approach for real robot navigation tasks.

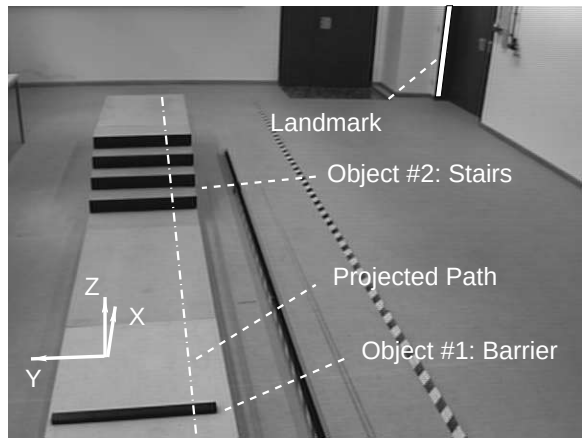

Figure 8: Observer's view of the scenario for gaze control experiments with a barrier (#1), staircase (#2), and landmark.

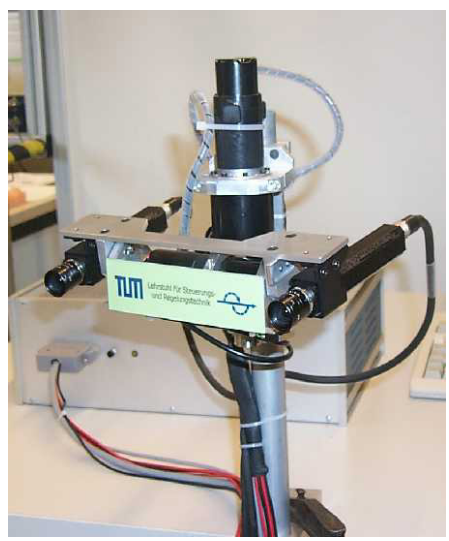

Figure 10: High-perfomance stereo camera head for active vision applications.

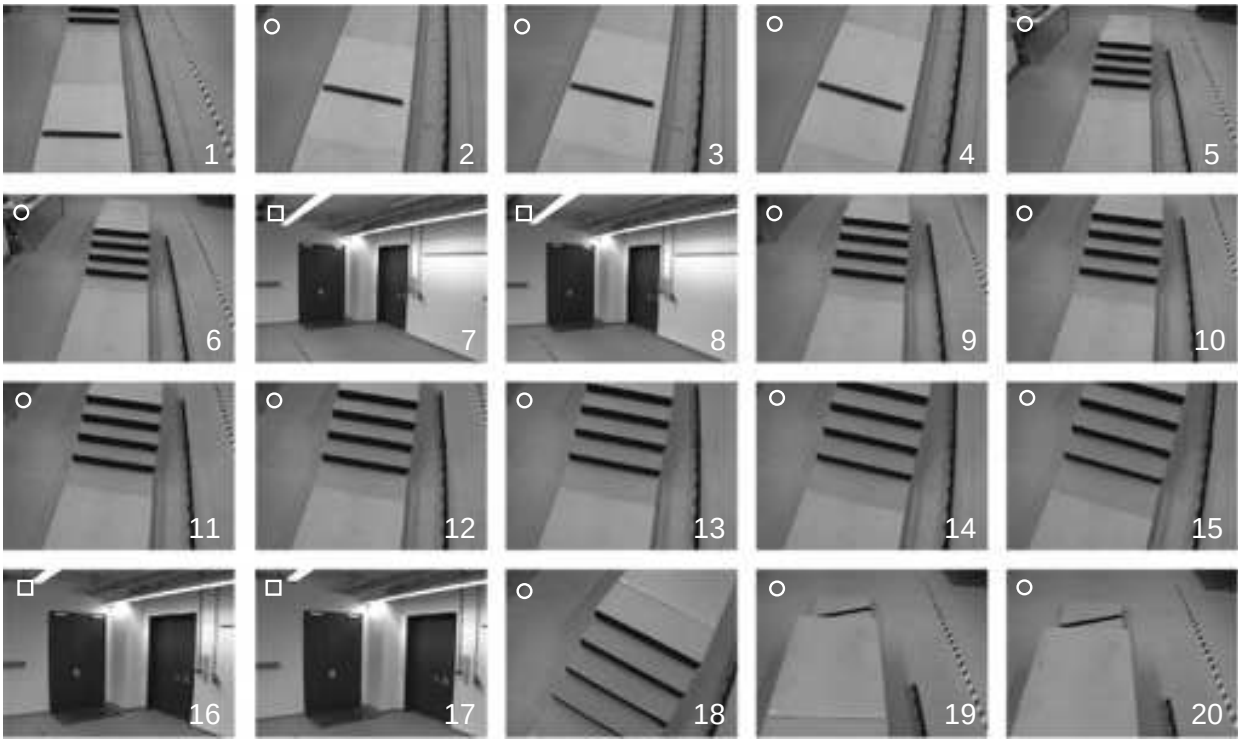

Figure 9: Image sequence while moving over the scenario shown in Figure 8, with optimized view direction for self localization (□) and obstacle avoidance (○).

These results on gaze control as well as the complete biologically inspired vision architecture for goal oriented humanoid walking, presented in Humanoids 2000 [8], will be validated on real walking machine hardware. As a first step towards the integration into a real biped, a high-performance stereo camera head has been designed and constructed, cf. Figure 10, to reach the high dynamical requirements arising from the gaze control for biped locomotion.

Acknowledgements : This work was supported in part by the German Research Foundation (DFG) within the "Autonomous Walking" Priority Research Program.

References

[1] A. E. Patla. Understanding of roles of vision in the control of human locomotion. *Gait and Posture*, 5:54–69, 1997. 4

[2] Y. Aloimonos. Introduction: Active Vision Revisited. Technical report, University of Maryland, 1999. 1

[3] A. J. Davison and N. Kita. Active Visual Localisation for Cooperating Inspection Robots. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2000. 2

[4] J. Denk and G. Schmidt. Walking Primitive Synthesis for an Anthropomorphic Biped using Optimal Control Techniques. In *Proc. of the Int. Conf. on Climbing and Walking Robots (CLAWAR)*, pages 819–826, Karlsruhe, Germany, 2001. 1

[5] J. F. Seara et. al. ViGWaM Active Vision System— Gaze Control for Goal-Oriented Walking. In *Proc. of the Int. Conf. on Climbing and Walking Robots (CLAWAR)*, pages 265–272, Karlsruhe, Germany, 2001. 1

[6] I. Moon et. al. On-line Selection of Stable Visual Landmarks under Uncertainty. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 781–786, Kyongju, Korea, 1999. 2

[7] L. Matthies and S. A. Shafer. Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation*, 3(3), June 1987. 3

[8] O. Lorch et. al. ViGWaM — An Emulation Environment for a Vision Guided Virtual Walking Machine. In *Proc. of the IEEE/RAS Int. Conf. on Humanoid Robots*, Cambridge, Massachusetts, USA, 2000. 1, 3, 7, 9

[9] C. E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:397–423, 623–659, 1948. 2