Proceedings of the 2002 IEEE
International Conference on Robotics & Automation
Washington, DC • May 2002

Perception Errors in Vision Guided Walking: Analysis, Modeling, and Filtering

O. Lorch J. F. Seara K. H. Strobl U. D. Hanebeck G. Schmidt

Institute of Automatic Control Engineering (LSR),
Technische Universität München, D-80290 Munich, Germany,
oliver.lorch@ei.tum.de

Abstract

This article deals with specific aspects concerning the visual perception process of a humanoid walking machine. An active vision system provides the information about the environment necessary for autonomous goal-oriented locomotion. Due to errors in each stage of the perception process, ideal environment reconstruction is not possible. By modeling these errors, stochastic components can be compensated using a hybrid Extended Kalman Filter approach with an alternating reference frame, thus reflecting the discontinuous character of biped walking. The perception results improved by filtering can be used for the autonomous locomotion of the robot. Experiments with the walking machine BARt-UH[1] demonstrate the validity of our approach.

1 Introduction

Robots are becoming more and more common in everyday life and as part of human society. Among the great variety of robots, biped humanoids are most likely to be employed for carrying out jobs typically performed by humans. Therefore, human abilities and even behaviors are favored for this kind of robots. Although human sensory capabilities are too multifaceted to be imitated by humanoids, they can be used as inspiring examples in order to equip the robot with the appropriate perception mechanisms for successful fulfillment of a given task.

This article discusses some of the perception problems when using a vision system on-board a walking machine. The objective of our research, as explained in [1], is the development of a vision-guided humanoid robot with the capability to walk safely in an environment with obstacles, as indicated in Fig. 1. For this task, a sufficient perception of the surroundings and a reliable location estimation of the objects in the walking path is required. The walking machine must recognize and localize relevant objects to perform planning of appropriate step sequences and to react to unexpected situations.

[1] **B**ipedal **A**utonomous **R**obo**t** – **U**niversität **H**annover. Further information concerning BARt-UH can be found at `http://www.biped.irt.uni-hannover.de`.

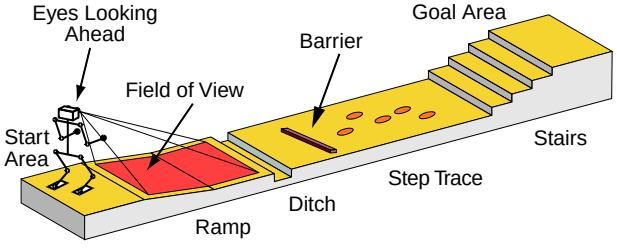

Figure 1: Prototypical scenario for biped locomotion.

After the problem definition in Section 2 and a brief introduction of the specifics of artificial visual perception for biped locomotion, a formulation of the perception process and related errors are described in Section 3. Section 4 presents a novel approach to a Kalman Filter Formulation with hybrid characteristics and its application to a walking machine. Results of experiments are outlined in Section 5, demonstrating the validity of the proposed new filtering approach.

2 Problem Definition

For purposes of recognizing and locating obstacles in its environment, the walking machine considered is equipped with an active stereo-camera pair. Visual information is analyzed by a real-time image processing system. Errors arise in each phase of the perception process, leading to erroneous obstacle location estimates. In addition, walking machine movements perturb image acquisition, causing even larger estimation errors. The active camera head of the walking machine is equipped with a view direction stabilization capability, which can partly compensate the movements of the walking machine. Therefore, some of the errors caused by blurred images can be reduced. However, certain errors remain, thus requiring the application of estimation and filtering techniques.

Perception process errors imply uncertainties in the measurements, which increase together with the uncertainties in the movements of the walking machine. These uncertainties are reduced every time the walking machine "sees" the objects. As the field of view of a vision sen-

sors is limited, not all objects will be seen simultaneously. Therefore, gaze control is needed, directing the gaze to objects of interest. This topic is a part of our research and is treated in another article [2].

In order to improve the object location estimates, the perception errors have to be modeled. Based on such a model, an appropriate Kalman Filter is developed and designed. The filter makes use of redundancy provided by the information contained in the image stream, thus increasing estimation accuracy.

3 Visual Perception Process

The specific characteristics of biped locomotion allow obstacle avoidance strategies, which are not feasible with wheel based mobile robots, such as striding over barriers or climbing up stairs. However, the particular stability problems of biped walking imply that these obstacle avoidance strategies require rather precise information about the object not only with respect to the sensor, the stereo-camera head in our case, but also with respect to the feet of the robot. This is the reason why the perception process must be analyzed carefully, taking into consideration not only various sources of errors during visual reconstruction, but also the transformation of these results into one of the two foot coordinate frames.

3.1 Nominal Reconstruction and Frames

Projection Model From 3D to 2D: The perception process in a pair of stereo-cameras used here can be explained with the pinhole camera model by the mapping of a 3D point ${}_F\boldsymbol{x} = [x\ y\ z\ 1]^T$ into a 2D point $\boldsymbol{m} = [u\ v]^T$. The 3D point coordinates are referred to a foot frame S_F, which is located in the stand foot of the walking machine, as shown in Fig. 2. Obviously, there exist two alternative foot frames: left S_{F_L} and right S_{F_R}. During double support, the last foot which contacted the ground is selected. Although S_F is changing in every step, we will suppress subindeces for simplicity in the following. $\boldsymbol{m}$ denotes the undistorted image point and is referred to the camera image frame S_{I_L} and S_{I_R} of each camera.

The mapping from 3D to 2D can be represented by two transformations: one from the foot frame S_F to the camera reference frame S_C of the stereo-rig, which applies for both cameras, cf. Fig. 2,

$$ {}_C\boldsymbol{x} = {}^C\boldsymbol{T}_F\, {}_F\boldsymbol{x}\,, \tag{1} $$

and another transformation from the camera reference frame S_C to each camera image frame $S_{I_{L,R}}$. Following the Tsai-Model [3], this mapping can be accomplished by the perspective perception matrices $\boldsymbol{P}_L$ and $\boldsymbol{P}_R$,

$$ s\,[\boldsymbol{m}_L{}^T\ 1]^T = \boldsymbol{P}_L\, {}_C\boldsymbol{x} \quad , \tag{2} $$

$$ s\,[\boldsymbol{m}_R{}^T\ 1]^T = \boldsymbol{P}_R\, {}_C\boldsymbol{x} \quad , \tag{3} $$

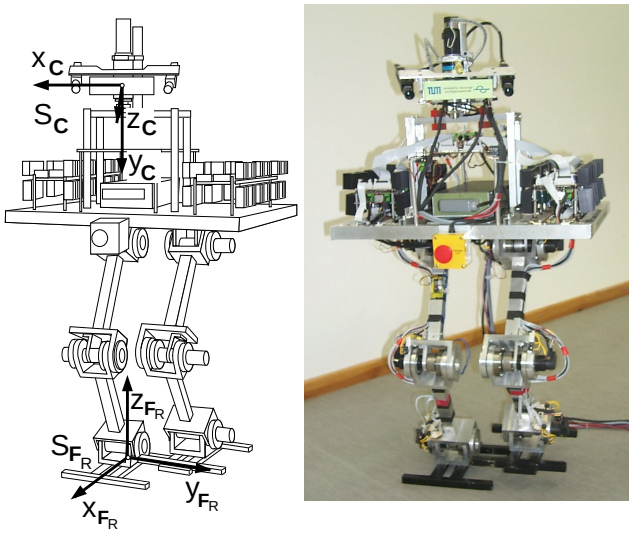

Figure 2: The walking machine BARt-UH with the head reference frame S_C and the (right) foot frame S_{F_R}.

where s is an arbitrary scalar, $\boldsymbol{m}_{L,R}$ are the undistorted image point coordinates for the left and right camera and ${}_C\boldsymbol{x}$ is the point coordinate referred to the camera frame S_C. For simplicity, quantization and distortion are neglected in this section.

The perception matrix $\boldsymbol{P}_L$ can be written as follows,

$$ \boldsymbol{P}_L = \left(\begin{array}{ccc|c} -f\,r_{11} & -f\,r_{12} & -f\,r_{13} & -f\,t_x \\ -f\,r_{21} & -f\,r_{22} & -f\,r_{23} & -f\,t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{array}\right) = \begin{pmatrix} \boldsymbol{q}_{L,1}^T & q_{L,14} \\ \boldsymbol{q}_{L,2}^T & q_{L,24} \\ \boldsymbol{q}_{L,3}^T & q_{L,34} \end{pmatrix}, $$

where f is the focal distance, t_i and r_{ij} are the components of the translation and rotation matrix from each camera frame S_I to the camera reference frame S_C using the roll, pitch, and yaw notation. There exists an analogous matrix $\boldsymbol{P}_R$.

Reconstruction From 2D to 3D: For purposes of obstacle avoidance, the 3D point coordinate, causing the two 2D camera coordinates, is the point of interest. Objects are projected to the two camera image frames $S_{I_{L,R}}$ and the inverse transformation has to be performed to obtain the 3D coordinates of the objects referred to the camera frame.

Let us assume that the image coordinates $\boldsymbol{m}_{L,R} = [u_{L,R}\ v_{L,R}]^T$ are known for both cameras and the stereo correspondence problem is already solved. To find the corresponding 3D point coordinates, ${}_C\boldsymbol{x}$ has to fulfill (2) and (3) for both $\boldsymbol{m}_{L,R}$. From the two views, we obtain a total of four linear equations in the coordinates of ${}_C\boldsymbol{x}$, which can be written and solved in various ways [4, 5]. We solve

$$ \boldsymbol{F}\, {}_C\boldsymbol{x}^* = \boldsymbol{b} \quad , \tag{4} $$

for ${}_C\boldsymbol{x}^*$, ${}_C\boldsymbol{x} = \left[{}_C\boldsymbol{x}^{*T}\ 1\right]^T$, with

$$\boldsymbol{F} = \begin{pmatrix} u_L \boldsymbol{q}_{L,3}^T - \boldsymbol{q}_{L,1}^T \\ v_L \boldsymbol{q}_{L,3}^T - \boldsymbol{q}_{L,2}^T \\ u_R \boldsymbol{q}_{R,3}^T - \boldsymbol{q}_{R,1}^T \\ v_R \boldsymbol{q}_{R,3}^T - \boldsymbol{q}_{R,2}^T \end{pmatrix}, \; \boldsymbol{b} = \begin{bmatrix} q_{L,14} - u_L q_{L,34} \\ q_{L,24} - v_L q_{L,34} \\ q_{R,14} - u_R q_{R,34} \\ q_{R,24} - v_R q_{R,34} \end{bmatrix}. \quad (5)$$

The matrix $\boldsymbol{F}$ and vector $\boldsymbol{b}$ are uncertain, due to errors in the measurements of u and v. The real values $\boldsymbol{F}$ and $\boldsymbol{b}$ are considered as the estimated values $\hat{\boldsymbol{F}}$ and $\hat{\boldsymbol{b}}$ with perturbations $\Delta\boldsymbol{F}$ and $\Delta\boldsymbol{b}$, $\boldsymbol{F} = \hat{\boldsymbol{F}} + \Delta\boldsymbol{F}$ and $\boldsymbol{b} = \hat{\boldsymbol{b}} + \Delta\boldsymbol{b}$. For a first estimation of ${}_C\hat{\boldsymbol{x}}^*$ the perturbations are neglected. The solution can then be found as the least squares solution of $\hat{\boldsymbol{F}}$ given by

$${}_C\hat{\boldsymbol{x}}^* = (\hat{\boldsymbol{F}}^T \hat{\boldsymbol{F}})^{-1} \hat{\boldsymbol{F}}^T \hat{\boldsymbol{b}}. \quad (6)$$

With the first solution of ${}_C\hat{\boldsymbol{x}}^*$ from (6), an error vector $\boldsymbol{e}$ can now be defined as

$$\boldsymbol{e} = \boldsymbol{F}\,{}_C\hat{\boldsymbol{x}}^* - \boldsymbol{b} \approx \Delta\boldsymbol{F}\,{}_C\hat{\boldsymbol{x}}^* - \Delta\boldsymbol{b} = \boldsymbol{g}(\boldsymbol{m}_L, \boldsymbol{m}_R). \quad (7)$$

The errors Δu_L, Δv_L in the u- and v-direction are assumed to be uncorrelated. The individual variances are assumed to be known. With the Jacobian of $\boldsymbol{g}(\boldsymbol{m}_L, \boldsymbol{m}_R)$, $\boldsymbol{e}$ can be approximated as

$$\boldsymbol{e} \approx \boldsymbol{J}_{uv} \left[\Delta u_L \; \Delta v_L \; \Delta u_R \; \Delta v_R\right]^T, \quad (8)$$

with

$$\boldsymbol{J}_{uv} = \begin{pmatrix} \boldsymbol{J}_{L,uv} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}_{R,uv} \end{pmatrix}$$

and

$$\boldsymbol{J}_{L,uv} = \begin{pmatrix} \boldsymbol{q}_{L,3}^T {}_C\hat{\boldsymbol{x}}^* - q_{L,34} & 0 \\ 0 & \boldsymbol{q}_{L,3}^T {}_C\hat{\boldsymbol{x}}^* - q_{L,34} \end{pmatrix}.$$

Similar results are obtained for $\boldsymbol{J}_{R,uv}$.

With the Jacobian matrix $\boldsymbol{J}_{uv}$ and the error covariance matrix $\boldsymbol{E}_{uv} = \mathrm{diag}\left(\sigma_{u_L}^2, \sigma_{v_L}^2, \sigma_{u_R}^2, \sigma_{v_R}^2\right)$ in the camera image frame, the influence of the pixel uncertainties on each linear equation from (4) can be computed with the weight matrix $\boldsymbol{E} = \boldsymbol{J}_{uv} \boldsymbol{E}_{uv} \boldsymbol{J}_{uv}^T$. A weighted least squares solution for the estimation of ${}_C\hat{\boldsymbol{x}}^*$ with the inverse of $\boldsymbol{E}$ is then given by

$${}_C\hat{\boldsymbol{x}}^* = (\hat{\boldsymbol{F}}^T \boldsymbol{E}^{-1} \hat{\boldsymbol{F}})^{-1} \hat{\boldsymbol{F}}^T \boldsymbol{E}^{-1} \hat{\boldsymbol{b}}. \quad (9)$$

Equation (9) can be solved for increased accuracy after a first solution of ${}_C\hat{\boldsymbol{x}}^*$ solved with Equation (6), respectively (9) with $\boldsymbol{E} = \boldsymbol{1}$, the identity matrix. Equation (9) can be solved again with the new ${}_C\hat{\boldsymbol{x}}^*$. This iteratively reweighted least squares method terminates when the absolute difference between each component of two consecutive values of ${}_C\hat{\boldsymbol{x}}^*$ is smaller than a prespecified positive threshold ϵ. It is obvious that the method can easily be extended for three and more cameras.

With the known matrix ${}^C\boldsymbol{T}_F$ the point ${}_C\hat{\boldsymbol{x}}$ can be transformed to ${}_F\hat{\boldsymbol{x}}$. Comparative simulations with and without the explained error weighting have shown an improvement of the reconstruction results with good convergence, most notably when the object features are within the borders of the image. For reasonable values of ϵ, typically three iterations are required.

3.2 Sources of Errors in the Perception Process

Fig. 3 shows the transformation relationships of the coordinate frames used, in order to facilitate the localization of the error sources.

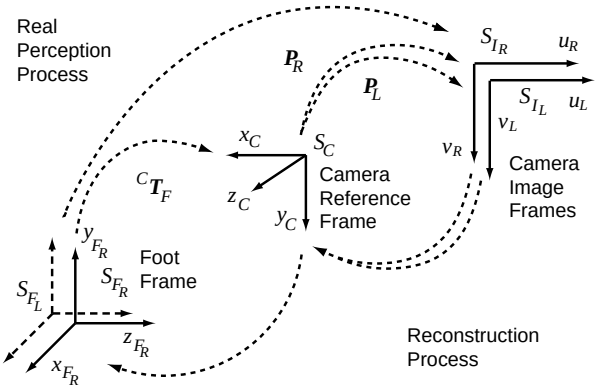

Figure 3: Main reference frames and transformations.

As explained in the previous section, the perception process is mathematically modeled with a transformation of the point coordinates from the foot, S_F, to the camera frame, S_C, using the matrix ${}^C\boldsymbol{T}_F$, and the perception matrices $\boldsymbol{P}_L$ and $\boldsymbol{P}_R$, which relates the 3D point coordinates to the 2D image point coordinates given in camera image frame S_I. The inverse transformation describes the location estimation problem.

The mathematical formulation of the reconstruction process, explained in Section 3.1, is not exact and depends on many factors. Among others, the perception matrix coefficients are only estimated by calibration and the lens distortion and discretization errors normally cannot be neglected.

In the following, the main causes of errors are resumed:

- discretization errors during data acquisition,
- errors due to the effects of motion in half frames,
- varying lighting conditions,
- calibration errors in the perception matrices $\boldsymbol{P}_L$ and $\boldsymbol{P}_R$ used,
- errors in ${}^C\boldsymbol{T}_F$ due to many subsequent transformations from S_C to S_F which are sensitive to the structural integrity and angle-uncertainties in the robot's kinematic chain,
- optimization criteria used in the reconstruction process.

4 Filtering

4.1 Preliminaries

The sources explained in Section 3.2 cause remaining errors in the spatial reconstruction of an object relative to the camera frame and foot frame respectively.

In this section, the use of an Extended Kalman Filter (EKF) to reduce the location estimation is explained. The EKF formulation is given in the context of walking machines. The filter is adapted to reflect the hybrid characteristics of biped walking due to alternating foot reference frames.

4.2 Extended Kalman Filter

The Kalman Filter is a tool for state estimation of linear systems based on a sequence of measurements. With the nonlinear transformations of the estimated matrix ${}^{C}\boldsymbol{T}_{F}$ the system and measurement equations are written as follows

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}\left(\boldsymbol{x}_k,\, \boldsymbol{u}_{k+1},\, \boldsymbol{w}_k\right) \tag{10}$$

$$\boldsymbol{z}_{k+1} = \boldsymbol{h}\left(\boldsymbol{x}_{k+1},\, \boldsymbol{v}_{k+1}\right). \tag{11}$$

The random variables $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ represent the process and measurement noise respectively – they are assumed to be white and zero mean with diagonal non-zero covariance matrices $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$.

Prediction: The basic idea of the EKF is to linearize the state transition function $\boldsymbol{f}$ and the measurement function $\boldsymbol{h}$ around the estimated state vector. The problem can then be defined by the following nonlinear stochastic difference equations

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{f}\left(\hat{\boldsymbol{x}}_{k|k},\, \boldsymbol{u}_{k+1},\, \boldsymbol{0}\right) \tag{12}$$

$$\hat{\boldsymbol{z}}_{k+1|k} = \boldsymbol{h}\left(\hat{\boldsymbol{x}}_{k+1|k},\, \boldsymbol{0}\right) \tag{13}$$

$$\boldsymbol{C}_{k+1|k} = \boldsymbol{A}_k \boldsymbol{C}_{k|k} \boldsymbol{A}_k^T + \boldsymbol{W}_k \boldsymbol{Q}_k \boldsymbol{W}_k^T, \tag{14}$$

where

$$\boldsymbol{A}_k = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}_k} \;\text{ and }\; \boldsymbol{W}_k = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{w}_k},$$

and $\boldsymbol{C}$ denotes the covariance matrix representing the uncertainty of $\boldsymbol{x}$.

Update: The next step is to correct the estimated values found by the prediction equations (12)-(14) with the new measurement $\boldsymbol{z}_{k+1}$. These update equations can be expressed as follows

$$\hat{\boldsymbol{x}}_{k+1|k+1} = \hat{\boldsymbol{x}}_{k+1|k} + \boldsymbol{K}_{k+1}\left(\boldsymbol{z}_{k+1} - \hat{\boldsymbol{z}}_{k+1|k}\right), \tag{15}$$

$$\boldsymbol{C}_{k+1|k+1} = \boldsymbol{C}_{k+1|k} - \boldsymbol{K}_{k+1} \boldsymbol{S}_{k+1} \boldsymbol{K}_{k+1}^T, \tag{16}$$

where the Kalman gain $\boldsymbol{K}$ is defined as

$$\boldsymbol{K}_{k+1} = \boldsymbol{C}_{k+1|k} \boldsymbol{H}_{k+1}^T \boldsymbol{S}_{k+1}^{-1}, \quad \text{where} \tag{17}$$

$$\boldsymbol{S}_{k+1} = \boldsymbol{H}_{k+1} \boldsymbol{C}_{k+1|k} \boldsymbol{H}_{k+1}^T + \boldsymbol{V}_{k+1} \boldsymbol{R}_{k+1} \boldsymbol{V}_{k+1}^T,$$

$$\boldsymbol{H}_{k+1} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}_{k+1}} \quad \text{and} \quad \boldsymbol{V}_{k+1} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{v}_{k+1}}.$$

The influences of nonlinear characteristics, arising from errors in the orientation of the robot, have been considered. However, the analysis of this nonlinearity is not within the focus of this article.

***Hybrid* Extended Kalman Filter:** The application of an EKF to the measurements made by a mobile wheeled robot has already been studied extensively, see e.g. [6]. However, in case of a walking machine the movements are not continuous and require an adaptation of the filtering method. This adaptation leads us to a new filtering formulation reflecting this discontinuity, by choosing an alternating reference frame S_F in the feet in which the measurements are filtered.

On the one hand, the movements of the walking machine cause the relative coordinates of a sample point fixed on the floor to change continuously in the camera reference frame S_C. On the other hand, the object reconstruction process calculates the 3D point coordinate referred to S_C. Subsequently, it translates the object coordinates to frame S_F. The latter transformation is uncertain due to several unmodeled effects, e.g. backlash in the joints, torsion in the limbs, compliance of the floor etc. These uncertainties cause an error which has to be taken into account.

In the context of walking machines, we use the EKF in a new way. Filtering for the position estimation process is much simpler, if the foot frame S_F is used, since a fixed 3D point in the environment does not change its coordinates during one step relative to the foot frame of the fixed standing foot. Therefore, the information of a series of measurements can be combined in a fixed coordinate frame in order to achieve higher accuracy in the location estimation of the objects. This means, previous measurements can be easily used to predict the actual measurement and to reduce uncertainty.

When S_F is fixed, the function $\boldsymbol{f}$ in (12) can be represented as follows

$$\hat{\boldsymbol{x}}_{k+1|k} = \hat{\boldsymbol{x}}_{k|k}\left(1 - \gamma_k\right) + \boldsymbol{f}_s\left(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{0}\right)\gamma_k, \tag{18}$$

where γ_k is a binary variable representing the current control vector: $\gamma_k = 0$ means no change in the coordinate frame S_F, $\gamma_k = 1$ indicates change in the coordinate frame. Consequently, the function $\boldsymbol{f}_s$ is defined as the transformation of the state vector $\boldsymbol{x}$ when a step takes place. The predicted uncertainty matrix of the state vector also has two different formulations depending on the value of γ_k, i.e.

$$\boldsymbol{C}_{k+1|k} = \boldsymbol{C}_{k|k}, \qquad \text{if } \gamma_k = 0, \tag{19}$$

$$\boldsymbol{C}_{k+1|k} = \frac{\partial \boldsymbol{f}_s}{\partial \hat{\boldsymbol{x}}_k} \boldsymbol{C}_{k|k} \frac{\partial \boldsymbol{f}_s}{\partial \hat{\boldsymbol{x}}_k}^T + \boldsymbol{W}_k \boldsymbol{Q}_k \boldsymbol{W}_k^T, \quad \text{if } \gamma_k = 1, \tag{20}$$

where the process noise $\boldsymbol{Q}_k$ represents the dead reckoning error accumulated during a single step. The major advantage of the novel filter formulation is, that $\boldsymbol{Q}_k$ is included at once, when the current step terminates and the foot frame S_F changes. The noise considered in the measurements, $\boldsymbol{R}_{k+1}$, comprises sub-matrices $\boldsymbol{R}^i_{k+1}$ of the uncertainty in the 3D measurement of each object i. This matrix cannot be considered as a constant as it depends on the object position. Hence, it has to be recalculated at every time step.

5 Experiments

5.1 Experimental Set-up

The sample experimental set-up consists of the vision-based guidance system [1] with the pan-tilt stereo-camera head mounted on the biped robot BARt-UH [7], see Fig. 4.

In the experiment described here, we consider the constant three footprints on the walking trail to be stepped on by the robot, cf. Fig. 4. During the locomotion, blob-tracking algorithms to compute the footprint locations are started once every step. When the head is moving, a new view direction is to be set. Therefore, the tracking algorithms are stopped and reinitialized in the next step. This task is accomplished by an underlying predictive gaze controller based on the maximization of the visual information [8]. First, the correspondence problem is solved, the footprint locations are estimated, as explained in Section 3.1, and then transformed into the foot frame of the current standing foot $S_{F_{L,R}}$. This is the input to the filter.

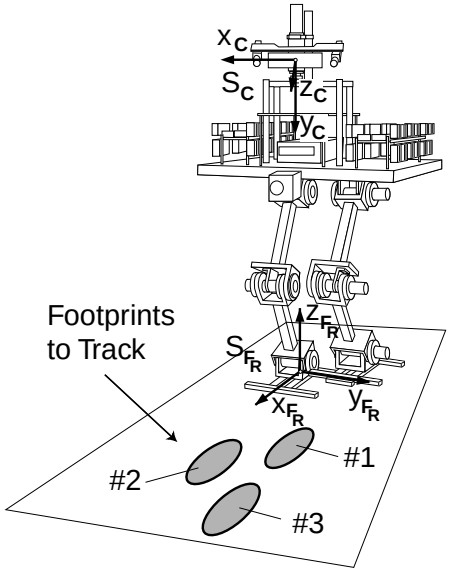

Figure 4: BARt-UH with stereo-camera head walking along a trail with three footprints.

5.2 Experimental Results

In Fig. 5 the evolution of the x_F-position of the footprints is shown. The positions are given relative to the currently standing foot during locomotion of BARt-UH. The vertical lines denote a step-change, i.e. changes in S_F, where the state vector is updated with (18). Each measurement of the vision system is plotted as ($\times$). The output of the filter is illustrated by the three stair-like dashed lines, reflecting the hybrid nature of the walking process. Note, that the distance between the first two footprints is $7\,cm$ as opposed to $34\,cm$ between the last two. It can also be observed, that the footprints do not always remain within the field of view – depending on the operation of the view direction controller– and therefore cannot be tracked continuously.

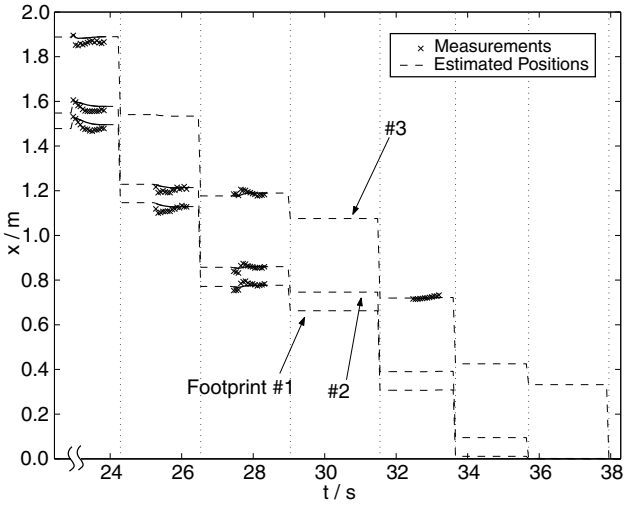

Figure 5: Evolution of the x_F–position of the footprints relative to the alternating foot frames $S_{F_{L,R}}$.

Fig. 6 presents a detailed view of Fig. 5 around $t = 28\,s$. The filter results are shown as a dashed line, changing to a solid line when the footprints are being tracked. The 90% confidence margin of the filter result is indicated by dash-dotted lines.

The evolution of the square root of the eigenvalue in x-direction σ_x, representing the position error variance for each footprint, is plotted in Fig. 7.

When a footprint is in view, the object location estimation can be performed and the standard deviation is decreasing rapidly. This case is indicated in Fig. 7 with a solid line, else with a dash-dotted line, thus corresponding with the plotted crosses ($\times$) in Fig. 5. It is noteworthy, that the standard deviation of the footprint locations also decreases even for those which are currently not in view, due to consideration of the coupling in the covariance matrix of the filter.

With the presented filter results, the information was enough for the robot BARt-UH to step on the footprints

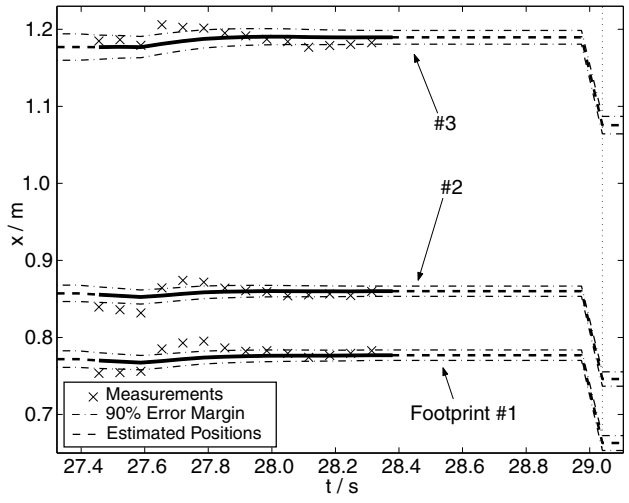

Figure 6: Evolution of the x_F–position of the footprints in detail.

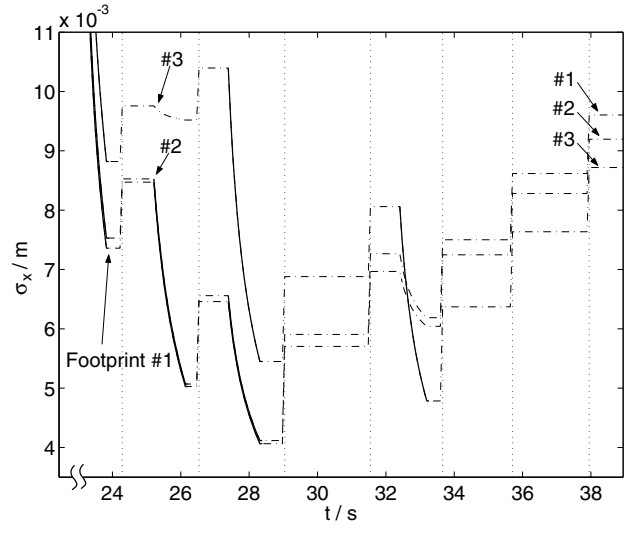

Figure 7: Evolution of the σ-value in x-direction of the object given in the foot frame.

within a margin of about $2\,cm$ providing an active walking behavior. This and other results are to be considered a major step forward towards autonomous, goal-oriented walking.

6 Conclusions

For the purpose of decreasing location uncertainty in vision-guided biped locomotion, the visual perception process has been studied in detail. By analyzing the various sources of errors, a corresponding error model was derived. Based on the error model, an algorithm for object location estimation using an iteratively reweighting least squares method was developed. Remaining stochastic error components are reduced by means of a modified Extended Kalman Filter. The filter takes into account the hybrid character of robot walking, which is reflected in an alternating reference frame for object representation in the feet. This novel type of filter is rather general and can be applied to different kinds of walking machines. Experiments with a real walking machine equipped with a stereo-vision perception system showed the validity of our approach.

Acknowledgments: The authors would like to thank members of the Institute for Automatic Control, Universität Hannover, for their cooperation in conducting these experiments with BARt-UH. This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG) within the "Autonomous Walking" Priority Research Program.

References

[1] O. Lorch, J. Denk, J. F. Seara, M. Buss, F. Freyberger, and G. Schmidt, "ViGWaM — An Emulation Environment for a Vision Guided Virtual Walking Machine," in *Proceedings of the IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Cambridge, Massachusetts, USA, September 2000.

[2] J. F. Seara, G. Schmidt, and O. Lorch, "ViGWaM Active Vision System — Gaze Control for Goal-Oriented Walking," in *Proceedings of the International Conference on Climbing and Walking Robots (CLAWAR)*, Karlsruhe, Germany, September 2001, pp. 265–272.

[3] R. Y. Tsai, "A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses," *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 4, August 1987, pp. 323–344.

[4] Richard I. Hartley and Peter Sturm, "Triangulation," *Journal of Computer Vision and Image Understanding (CVIU)*, vol. 68, no. 2, November 1997, pp. 146–157.

[5] O. D. Faugeras, *Three-Dimensional Computer Vision*, Artificial Intelligence. MIT Press, Cambridge, MA, USA, 1992.

[6] Andrew J. Davison and N. Kita, "Active Visual Localisation for Cooperating Inspection Robots," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Takamatsu, Japan, 2000, pp. 1709–1715.

[7] A. Albert and W. Gerth, "New Path Planning Algorithms for Higher Gait Stability of a Bipedal Robot," in *Proceedings of the International Conference on Climbing and Walking Robots (CLAWAR)*, Karlsruhe, Germany, September 2001, pp. 521–528.

[8] J. F. Seara, O. Lorch, and G. Schmidt, "Gaze Control for Goal-Oriented Humanoid Walking," in *Proceedings of the IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Tokio, Japan, November 2001, pp. 187–195.