Practical Experience with Vision-Based Biped Walking

Robert Cupec, Joachim Denk, and Günther Schmidt

Institute of Automatic Control Engineering, Technische Universität München,
D-80290 Munich, Germany, http://www.lsr.ei.tum.de

Abstract. This paper presents key ideas of our approach to perception-based biped robot walking. Computer vision techniques are employed for reactive adaptation of the step-sequence taking into account the restrictions imposed by dynamics and mechanical construction of the walking machine. Our approach and guidance system were evaluated by experiments with the biped robot BARt-UH.

1   Introduction

Biped robots are able to move more freely than wheeled robots in environments designed for humans. They can climb stairs or step over obstacles. To enable a robot to perform tasks autonomously appropriate perception capabilities together with close coordination between perception and locomotion are required. Results in simulated [7,13] and experimental [5,6,12] perception-guided walking in scenarios with obstacles have been reported.

Our research focuses on the interaction between visual perception and biped walking [9]. To support our studies we developed a stand-alone vision-based guidance system for walking machines. Information about obstacles in the walking trail is gathered using a line-based stereo vision algorithm for visual feedback. This information enables a predictive step sequence planner to determine the parameters of adequate step sequences for guiding the robot securely over a set of obstacles to a goal position.

In recent experiments we validated our approach by applying the guidance system to supervisory control of BARt-UH – a biped robot developed at IRT of the Universität Hannover, Germany [1]. The results reported in this article are considered to represent a new contribution to the challenging field of autonomous vision-guided biped walking.

2   Concept of Vision-Based Guidance

For performing a given locomotion task, the guidance system of a walking machine should provide the robot with the parameters of an appropriate step sequence. The sequence must allow the robot to reach a goal position taking into account obstacles in the walking trail. To select appropriate steps for overcoming obstacles, the reactive *step-sequence planner* needs sufficiently

accurate information about obstacle locations relative to the walking machine as well as obstacle dimensions.

In the guidance strategy described in this article, obstacle information is supplied by *visual feedback*. The architecture of the visual-based guidance system is shown in Fig. 1.
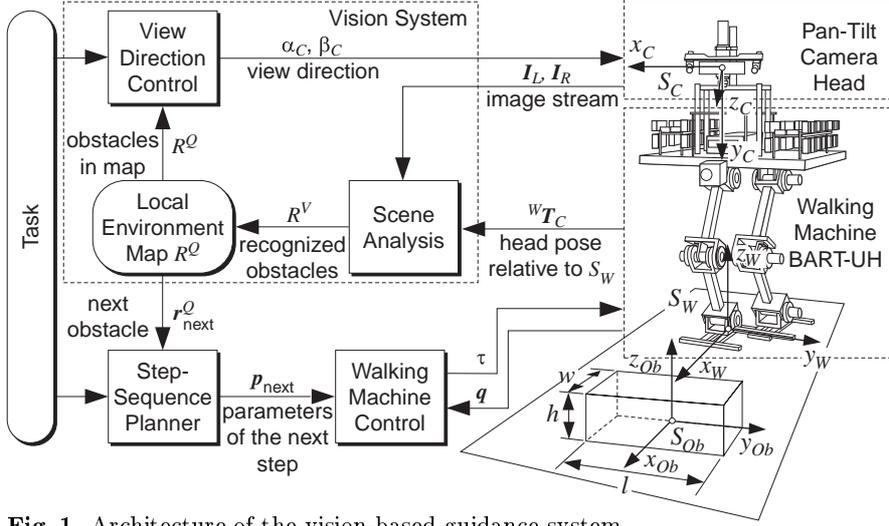
Fig. 1. Architecture of the vision-based guidance system

The environment of the walking machine is perceived by a stereo-camera pair mounted on a pan-tilt head. The resulting stream of camera images $\boldsymbol{I}_L$ and $\boldsymbol{I}_R$ represents the input to a *scene analysis algorithm*. The algorithm recognizes rectangular objects laying on the walking trail and estimates their size and pose relative to the coordinate system S_W which is fixed in the center of the current standing foot of the walking machine. Object recognition and visual estimation is performed once per step using the pose of the camera reference frame S_C relative to S_W. The transformation $^W\boldsymbol{T}_C$ between these two coordinate systems is calculated by use of the data from the encoders in the robot joints.

Each object recognized in the field of view is represented by a vector $\boldsymbol{r}_i^V = [\,c_i^V\ x_i^V\ y_i^V\ \alpha_i^V\ w_i^V\ h_i^V\ l_i^V\,]^T$, where c_i^V denotes the class of object; x_i^V, y_i^V and α_i^V represent the position and orientation of the object reference frame S_{Ob} relative to S_W and w_i^V, h_i^V and l_i^V are object dimensions, cf. Fig. 1.

Since the cameras' field of view is limited, the camera system should be directed in such a way that the currently most relevant objects in the walking trail are visible. This is obtained by a *gaze controller*, which selects the pan and tilt angle α_C and β_C of the camera system corresponding to the view direction with the maximal visual information content [9].

The set of objects $R^V = \{r_1^V, r_2^V, \cdots, r_n^V\}$ recognized in one pair of images $\boldsymbol{I}_L$ and $\boldsymbol{I}_R$ is used for updating the *local environment map* R^Q. This map represents the set of all obstacles r_j^Q appearing in the walking trail during one experiment.

The vector $\boldsymbol{r}_{next}^Q$ containing the information about the next obstacle in the walking trail is provided as input to the step sequence planner. Since the space of steps which a robot can execute is restricted by dynamics and mechanical design, only particular step combinations can be executed. This fact has to be regarded by the planning algorithm when selecting an appropriate sequence for overcoming the obstacle. Since uncertainty in visual estimation decreases as the robot approaches the obstacle, the step sequence is replanned in each step using current visual information. The parameters $\boldsymbol{p}_{next}$ of the next step to be executed are transfered to the robot locomotion controller.

3 Visual Feedback

Different approaches to scene interpretation for walking machines have been reported including depth-map generation [5] and model-based stair recognition [2,11]. In our approach a model-based object recognition algorithm is adopted, which recognizes separate rectangular objects using line segments extracted from a *single image* together with the information about camera system orientation. Size and position of objects relative to the walking machine are obtained by *stereo vision*.

3.1 Recognition of Rectangular Objects

Scene analysis is based on 2D straight line segments, obtained by edge detection and segmentation of edge contours. For each detected 2D line two hypotheses are generated. One hypothesis assumes a 2D line to represent a projection of a *vertical* 3D edge of an object and the other assumes the line to represent a projection of a *horizontal* 3D edge. The information about the orientation of the camera reference frame S_C relative to the gravity axis and the camera parameters obtained by calibration are used to reduce the number of false hypotheses about *vertical* edges [8], as explained next. Let $_C\boldsymbol{x}_i$ be a 3D-point represented in the camera reference frame and $\boldsymbol{m}_i = \begin{bmatrix} u_i & v_i \end{bmatrix}^T$ a 2D-point representing the projection of $_C\boldsymbol{x}_i$ to the image $\boldsymbol{I}$, cf. Fig. 2. The relation between $\boldsymbol{m}_i$ and $_C\boldsymbol{x}_i$ is given by

$$\begin{bmatrix} \boldsymbol{m}_i^T & 1 & s_i \end{bmatrix}^T = s_i \boldsymbol{P} \begin{bmatrix} _C\boldsymbol{x}_i^T & 1 \end{bmatrix}^T , \tag{1}$$

where $\boldsymbol{P}$ is the homogeneous projection matrix of the camera obtained by camera calibration and s_i is a scaling factor defining the position of $_C\boldsymbol{x}_i$ on the projection ray. Since a *vertical* 3D edge $\overline{\boldsymbol{x}_i\boldsymbol{x}_j}$ is parallel to the gravity axis $\boldsymbol{g}$, the following equation is satisfied

$$_C\boldsymbol{x}_j(s_j) - {_C\boldsymbol{x}_i(s_i)} = \parallel {_C\boldsymbol{x}_j(s_j)} - {_C\boldsymbol{x}_i(s_i)} \parallel {_C\boldsymbol{g}} , \tag{2}$$

where unit vector $_C\boldsymbol{g}$ parallel to the gravity axis represents the information on camera system orientation. Only those 2D lines $\overline{\boldsymbol{m}_i\boldsymbol{m}_j}$ which satisfy (1) and (2) can represent the projections of *vertical* 3D edges.
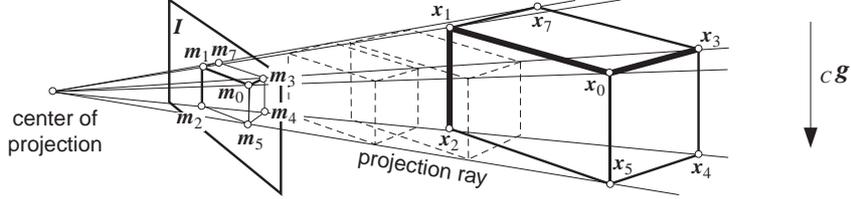
Fig. 2. Projection of a 3D object to a camera image $\boldsymbol{I}$. Groups of orthogonal edges (*bold lines*) are used as indicators for rectangular objects.

On the other hand, it is not possible to use the information about the camera system orientation $_C\boldsymbol{g}$ and the projection matrix $\boldsymbol{P}$ in the same manner in order to reject hypotheses about *horizontal* 3D edges. If no additional information is available, every 2D line $\overline{\boldsymbol{m}_i\boldsymbol{m}_j}$ which doesn't intersect the horizon line can represent a projection of a *horizontal* 3D edge.

Assuming that a 2D line $\overline{\boldsymbol{m}_i\boldsymbol{m}_j}$ is a projection of a *horizontal* 3D edge $\overline{\boldsymbol{x}_i\boldsymbol{x}_j}$, the orientation of $\overline{\boldsymbol{x}_i\boldsymbol{x}_j}$ can be determined using the fact that a *horizontal* 3D edge is perpendicular to the gravity axis $\boldsymbol{g}$, i. e.

$$\left(_C\boldsymbol{x}_j(s_j) - {_C}\boldsymbol{x}_i(s_i)\right)^T {_C}\boldsymbol{g} = 0\,. \tag{3}$$

The orientations of the hypothetical 3D edges are used for edge grouping. Chains of three mutually orthogonal edges, cf. Fig. 2 and groups of two or three parallel edges are used as indicators for rectangular objects. For each edge group an endpoint $_C\boldsymbol{x}_0(s_0)$ is selected as a reference point and coordinates of all other corner points of the hypothetical object are expressed as functions of s_0. Hence, the hypothetical object defined by the scaled corner points $_C\boldsymbol{x}_i(s_0)$ actually represents a family of objects corresponding to the same projection, as illustrated by Fig. 2. Verification of the object hypotheses is performed by matching the projections of the hypothetical object edges to the 2D lines detected in image $\boldsymbol{I}$.

3.2 Object Reconstruction Using Stereo Vision

The scaling factor s_0 needed for full reconstruction of the object is obtained using stereo vision. The stereo vision algorithm determines the correspondences between lines $\overline{\boldsymbol{m}_i^L\boldsymbol{m}_j^L}$ detected in the left camera image $\boldsymbol{I}_L$ and lines $\overline{\boldsymbol{m}_k^R\boldsymbol{m}_l^R}$ detected in the right camera image $\boldsymbol{I}_R$. Two corresponding lines represent the projections of a 3D edge to $\boldsymbol{I}_L$ and $\boldsymbol{I}_R$. The coordinates of the endpoints of the detected 3D edges relative to the camera reference frame S_C are obtained by triangulation.

The scaling factor s_0 is estimated by minimizing the sum of squared distances between the corner points $_C\boldsymbol{x}_i(s_0)$ provided by the recognition algorithm and the corresponding points $_C\hat{\boldsymbol{x}}_i$ estimated by stereo vision, i. e.

$$\hat{s}_0 = \arg\min_{s_0} \sum_{i=1}^{N} \parallel {}_C\hat{\boldsymbol{x}}_i - {}_C\boldsymbol{x}_i(s_0) \parallel^2, \tag{4}$$

where N is the number of corner points of the object confirmed by stereo vision.

The object poses relative to the camera reference frame S_C determined by stereo vision are transformed into the coordinate system S_W. The transformation $^W\boldsymbol{T}_C$ from S_C to S_W is obtained by direct kinematics using the information from the joint encoders.

3.3   Practical Considerations

Line-based vision systems are successfully applied in wheeled robotics [4,8] for self-localization and obstacle avoidance in indoor environments. Simple and reliable detection and precise 3D reconstruction of linear features make them an appropriate basis for robot perception.

In case of walking robots reliable detection of obstacles and precise estimation of their size and pose relative to the robot's standing foot is essential for correct step sequence planning. Unlike a wheeled robot, however, a walking robot represents a rather long kinematic chain connecting the camera system to the ground. Complex kinematics of a walking robot can cause a substantial uncertainty in determining the transformation $^W\boldsymbol{T}_C$. This uncertainty increased by movements of the camera system during walking can degrade the precision of visual estimation.

Furthermore, collisions between the robot's feet and the ground can generate vibrations of the cameras resulting in blurred images. In order to reduce this effect, our vision system triggers the image acquisition at an instant during the step when camera movement is slow. The state during execution of a step corresponding to the slowest head motion is defined by the robot's trajectory planning algorithm [1].

4   Reactive Step Sequence Planning

We assume a robot control system allowing execution of steps from a discrete set of M feasible walking primitives, cf. Fig. 3, as basis for situation dependent adaptive biped locomotion [3,7,9,13]. For purposes of step sequence planning and to permit fast and efficient collision checking, walking primitive properties are characterized in terms of the vector $\boldsymbol{p}^\zeta = [l_I^\zeta \ l_{II}^\zeta \ c^\zeta]^T$, $\zeta \in \{1, \ldots, M\}$, with l_I and l_{II} denoting step-length of the preceding and current step, respectively. Step-clearance c and the parameters $\Delta_{b/a}$, which are assumed to be constant for simplicity, define the collision free space needed for stepping over an obstacle.
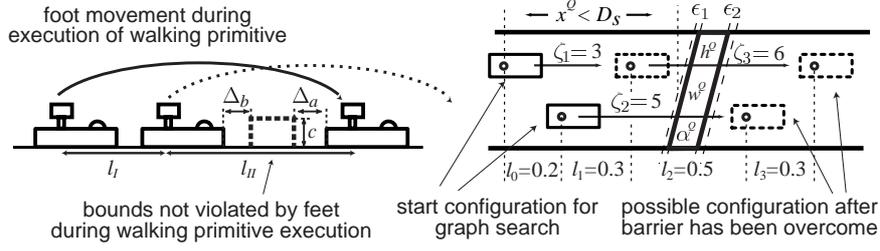
Fig. 3. First step over a barrier by execution of a walking primitive defined by step-lengths l_I, l_{II}, clearance c and the parameters Δ_b, Δ_a (**left**). Situation before a barrier, with distance x^Q, height h^Q, width w^Q, angle α^Q, and security regions given by ϵ_1, ϵ_2. Feasible step-sequence ζ_1, ζ_2, ζ_3 dashed (**right**)

During locomotion the set of walking primitives needs to be searched for a step combination allowing to traverse the next obstacle [7,13]. e.g. a barrier as illustrated in Fig. 3. The distance x^Q to the obstacle and other relevant obstacle information are provided by the vision system. Uncertainty in these parameters resulting from limited precision of image processing and dead-reckoning are regarded by introducing security areas ϵ_1, ϵ_2. Thereby, the obstacle boundaries are virtually enlarged.

For structured representation and a systematic search, the parameters of all walking primitives and the information on their possible combinations are represented by a directed graph, see Fig. 4. A node represents the state of the robot after execution of a walking primitive and is labeled with the corresponding step-length l. The walking primitives themselves are represented by the edges labeled with the corresponding parameters $\boldsymbol{p}^\zeta$, $\zeta \in \{1, \dots, M\}$ and costs k^ζ as a measure for the efficiency of the step.

ζ	l_I	l_{II}	c	k
1	0.3	0.3	0.0	0.30
2	0.3	0.2	0.0	0.35
3	0.2	0.3	0.0	0.35
4	0.2	0.2	0.0	0.50
5	0.3	0.5	0.1	1.10
6	0.5	0.3	0.1	1.10
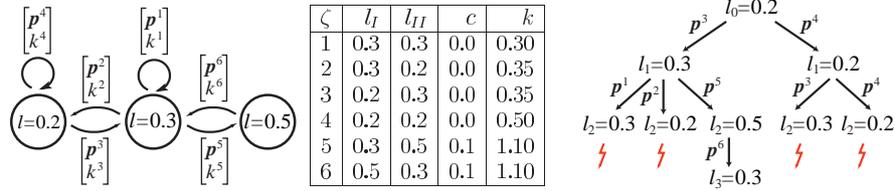
Fig. 4. Graph structure defining feasible walking primitives $\boldsymbol{p}^\zeta = [l_I^\zeta \; l_{II}^\zeta \; c^\zeta]^T$, their costs k^ζ and their possible combinations (**left**). Tree structure resulting from graph search for situation illustrated in Fig. 3. Flashes indicate collisions (**right**)

To determine an appropriate step sequence for overcoming the next obstacle, the graph is searched recursively by depth-first search expanding into a tree structure Fig. 4. The starting node corresponds to the state of the robot at the beginning of the step sequence. Each branch in the tree represents a possible sequence of walking primitives $P = [\zeta_1 \; \zeta_2 \; \dots]^T$. A branch

is searched until the obtained step sequence violates the current obstacle situation or until a feasible solution allowing to pass the obstacle has been found.

Usually there exist multiple feasible solutions $P^j = [\zeta_1^j \ \zeta_2^j \ \ldots \zeta_{N_j}^j]^T$, $j = 1 \ldots F$. From this set the solution is chosen, which fits the requirements of the actual task best. In the current implementation the step sequence minimizing the summarized costs k relative to the traveled distance is selected as

$$\nu = \arg \min_{j \in \{1, \ldots, F\}} \ \frac{\sum_{i=1}^{N_j} k_i^j}{\sum_{i=1}^{N_j} \frac{1}{2}(l_{i-1}^j + l_i^j)} \quad , \tag{5}$$

with $l_0^j = l_0, j = 1 \ldots F$ the step-length corresponding to the node from which the search was started. The costs k of a walking primitive are heuristically defined in such a way, that the execution of walking-primitives with step-lengths close to a nominal step-length l_n is favored as inspired by human locomotion [9]. Alternatively, costs k based on the electrical energy needed for the execution of a walking primitive might be used to reduce the power consumption of the system.

As long as there is no obstacle in a distance x^Q closer than $D_s \approx 3 \, l_n$ the biped walks with a nominal step length l_n. Otherwise the graph is searched once per step using current vision data until the obstacle has been passed. From the resulting step sequence P^ν the parameters $\boldsymbol{p}^{\zeta_1^\nu}$ of the first planned primitive are sent to the robot locomotion controller while the parameters of $\zeta_2^\nu, \zeta_3^\nu, \ldots \zeta_{N_\nu}^\nu$ are discarded. This strategy takes into account that results of visual perception are getting more precise the shorter the distance to the obstacle. In addition it allows to react to sudden changes in the obstacle situation.

To ensure real-time capability the parameters $\boldsymbol{p}^{\zeta_1}$ of the next step to be executed must be known well before the current step ends. Therefore a graph search must be aborted, if it is not completed when new parameters $\boldsymbol{p}^{\zeta_1}$ are requested by the robot locomotion controller. If feasible solutions have been found, the currently best solution is transmitted. Otherwise, the biped is stopped and the search is restarted allowing a longer time period.

The biped is also stopped if a graph search is completed without finding a step sequence for overcoming the next obstacle e.g. if the obstacle is too high. The graph search is restarted periodically and availability of a solution will indicate that the obstacle has been removed. In such a case the biped resumes locomotion demonstrating its reactive behaviour.

5   Experimental Results

To test the suitability and efficiency of our approach the stereo camera head developed at our laboratory was mounted on top of the robot BARt-UH. The stereo-vision system comprises two cameras with view-angles of $55°$ in horizontal and $42°$ in vertical direction and a stereo base-line of $240\,\mathrm{mm}$. Image

resolution is 640×480 pixels. Some additional details about the experimental setup can be found in [10].

Due to overlapping feet, cf. Fig. 5, the range of step-lengths which BARt can execute [1] is decomposed into two continuous intervals $l \in [0.08\,\text{m}; 0.12\,\text{m}]$ and $l \in [0.29\,\text{m}, 0.38\,\text{m}]$. The set of walking primitives used for step sequence planning was obtained by discretization of these intervals with $\Delta l = 0.01\,\text{m}$.
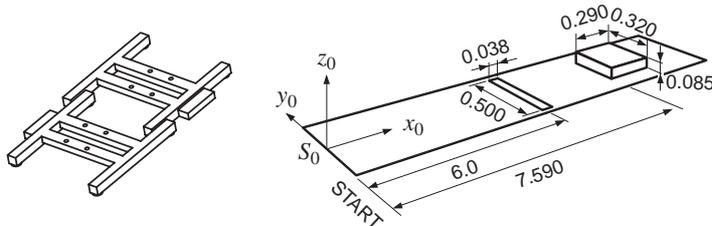
Fig. 5. Overlapping feet of BARt (**left**). Experimental setup (**right**). Two obstacles are put on the walking trail: a ribbon-type obstacle at 6 m and a box at 7.59 m

Fig. 5 shows the scenario for a typical experiment in which the perception system guided BARt to the goal position. Two obstacles are put on the walking trail: a ribbon-type obstacle at 6 m and a box at 7.59 m. During walking scene analysis was performed on pairs of images acquired once per step. The size and distance of the obstacles relative to the current standing foot of the robot were estimated using stereo vision.

In Fig. 6 the distance and size estimation errors for both obstacles are plotted versus the position of the camera system relative to the world coordinate system S_0, cf. Fig. 5. The performance of size estimation is depicted in Fig. 6 using as examples the width of the ribbon-type obstacle and the height of the box. In the presented experiment precise estimation of these parameters is critical because they determine if it is possible for BARt to step over the obstacles.

As expected, the estimation of obstacle parameters gets more precise as the robot approaches the obstacle. At distances $< 2\,\text{m}$ the distance estimation error is $< 5\,\text{cm}$ and the size estimation error is $< 2\,\text{cm}$.
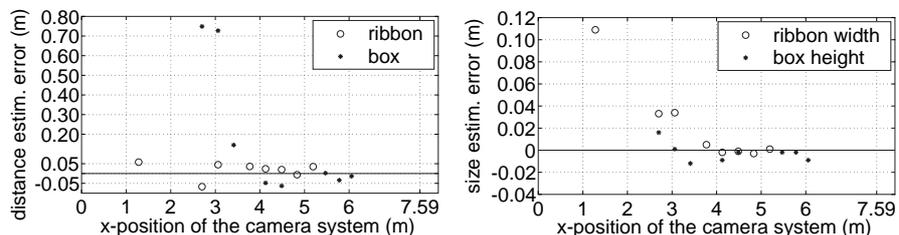
Fig. 6. Estimation precision for distance (**left**) and size (**right**)

The capability of the biped to react to changes in the environment was validated by putting a ribbon, in front of the robot during walking, as shown in Fig. 7. When the obstacle came inside the camera system's field of view, the ribbon was identified as a rectangular obstacle on the ground and its location and dimensions were estimated. The local environment map was updated with the obtained information. Consequently, the step sequence was adapted such that the left foot was placed in such a position in front of the ribbon that BARt was capable to successfully overcome the obstacle with the next step.

The capability of the system to distinguish between obstacles which can be overcome and those which cannot, was demonstrated by placing a box in the walking trail. This obstacle was too high for the robot to step over it. When the vision system recognized the box and estimated its dimensions, the step sequence planner could not find a step sequence to step over this obstacle and the robot stopped. As soon as the obstacle was removed from the walking trail, the vision system updated the local environment map. Accordingly, a step sequence guiding the robot towards the goal position was found and BARt continued walking.
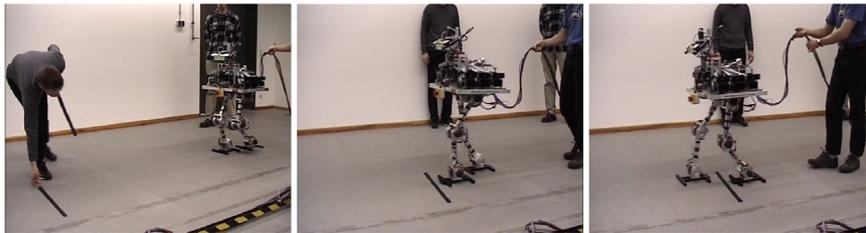
Fig. 7. A ribbon-type obstacle is inserted in the walking trail during operation of BARt-UH. The obstacle is recognized and localized and the robot adapts its steps overcoming the ribbon in a smooth and stable way.

6   Conclusions

In this article an approach to vision-guided robot walking is described. A line-based stereo vision algorithm is applied to provide the information about rectangular obstacles on the walking trail. This information is used by a predictive step-sequence planner which selects appropriate steps alowing the robot to overcome the obstacles and reach the goal position.

The presented experiments with the biped robot BARt-UH were conducted to examine the performance of the developed methods. The obstacle recognition and precision of visual estimation were not degraded significantly by the effects of camera movements caused by walking as the images were acquired at an instant during step execution when camera movement is slow.

The dynamic updating of the local environment map and continuous step-sequence planning allowed the robot to react to changes in the environment.

Future research will focus on improvement of the robustness of visual estimation using EKF techniques, advanced gaze control strategies and extension of reactive step-sequence planning to curve walking.

References

1. Albert, A., Gerth, W. (2001) New Path Planning Algorithms for Higher Gait Stability of a Bipedal Robot. Proc. of the Int. Conf. on Climbing and Walking Robots (CLAWAR 2001), Karlsruhe, Germany, 521–528
2. Cupec, R., Lorch, O., Schmidt, G. (2001) Object Recognition and Estimation of Camera Orientation for Walking Machines. Tagungsband zum 17. Fachgespräch Autonome Mobile Systeme, Stuttgart, Germany, Springer, 1–10
3. Denk, J., Schmidt, G. (2001) Synthesis of a Walking Primitive Database for a Humanoid Robot using Optimal Control Techniques. Proc. of IEEE-RAS Int. Conf. on Humanoid Robots (HUMANOIDS 2001), Tokyo, Japan, 319–326
4. Eberst, C. et al. (2000) Robust Video-Based Object Recognition Integrating Highly Redundant Cues for Indexing and Verification. Proc. of the IEEE Int. Conf. on Robotics and Automation, San Francisco, California, 3757–3764
5. Kagami, S. et al. (2001) Design and Implementation of Software Research Platform for Humanoid Robotics: H7. Proc. of the IEEE-RAS Int. Conf. on Humanoid Robots (HUMANOIDS 2001), Tokyo, Japan, 245–258
6. Kajita, S., Tani, K. (1996) Adaptive Gait Control of a Biped Robot based on Realtime Sensing of the Ground Profile. Proc. of the IEEE Int. Conf. on Robotics and Automation, Minneapolis, Minnesota, 570–577
7. Kuffner, J. J. Jr. et al. (2001) Footstep Planning Among Obstacles for Biped Robots. Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Maui, Hawaii, USA, 500–505
8. Lebègue, X., Aggarwal, J. K. (1993) Significant Line Segments for an Indoor Mobile Robot. IEEE Transactions on Robotics and Automation, Vol. 9, No. 6, December 1993, 801–815
9. Lorch, O., Denk, J., Seara, J. F., Buss, M., Freyberger, F., Schmidt, G. (2000) ViGWaM — An Emulation Environment for a Vision Guided Virtual Walking Machine. Proc. of IEEE-RAS Int. Conf. on Humanoid Robots (HUMANOIDS 2000), Cambridge, Massachusetts
10. Lorch, O., Albert, A., Denk, J., Gerecke, M., Cupec, R., Seara, J. F., Gerth, W., Schimdt, G. (2002) Experiments in Vision-Guided Biped Walking. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (submitted), Lausanne, Switzerland
11. Pack, D. J. (1996) Perception-Based Control for a Quadruped Walking Robot. Proc. of the IEEE Int. Conf. on Robotics and Automation, Minneapolis, Minnesota, 2994–3001
12. Shin, D. et al. (1989) Realization of Obstacle Avoidance by Biped Walking Robot Equiped with Vision System. Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Tsukuba, Japan, 268–275
13. Yagi, M., Lumelsky, V. (1999) Biped Robot Locomotion in Scenes with Unknown Obstacles. Proceedings of the IEEE International Conference on Robotics and Automation, Detroit, Michigan, 375–380