# ROBVISION
# Vision Based Navigation for Mobile Robots[1)]

*M. Ayromlou*[1], *C. Beltran*[2], *A. Gasteratos*[2], *O. Madsen*[3], *W. Ponweiser*[1], *M. Vincze*[1]

[1] Institute of Flexible Automation, Vienna University of Technology

[2] Laboratory for Integrated Advanced Robotics, University of Genoa

[3] Department of Production, Aalborg University

{ma,wp,vm}@infa.tuwien.ac.at, robvision.infa.tuwien.ac.at

*Abstract:*

*We introduce a system developed during the Esprit project RobVision (**ROB**ust **VI**sion for **S**ensing in **I**ndustrial **O**perations and **N**eeds), to navigate a walking robot through a known industrial environment - in this case a ship section. The basic idea is to continuously generate a robot position and orientation (pose) signal by combining visual sensing information from the environment with predetermined CAD-information. These pose signals are used by the trajectory planner in the robot controller. Robust feature detection and real-time capability are important issues to reach an acceptable performance of the overall system. Our pose update cycle time achieved was 120ms but could be enhanced. Due to jerks of the walking robot accelerometers are used for stabilization. Experiments show that our approach is feasible and reaches the positioning accuracies demanded.*

## 1 Introduction and System Overview

Quality assurance and intelligent products are key roads to success in global competition. Supervising and automatically measuring the quality of parts in the production of large structures can reduce work costs by up to 20% (expected by Odense Steel Shipyard Ltd. (OSS), the end user of this project). The inspection of these structures needs automated systems to position the tools necessary in a large environment (e.g. mobile robots). For both, quality measurement and positioning of inspection tools, the key is to generate the 3D pose of objects.

The goal of the RobVision system is to guide a robot through a ship construction. That means also in this application the localization of the robot, i.e. 3D pose calculation along a predetermined path is necessary. For calculating the pose of the robot, features of a known

[1)] This work has been mainly supported by the RobVision project Esprit 28867 and is partly supported by the Austrian Science Foundation (FWF) under grant P13167-MAT and Project GRD1-1999-10693 FlexPaint.

model are searched. The result of the searching process provide finally the basis for the pose determination.

Since industries are mostly using CAD systems to design parts or working areas we apply this model knowledge to initialize the feature searching process. The CAD system provides features to the vision module that tries to find these features in the images. Cameras of a stereo head mounted on the robot deliver these images (Figure 1a). We use two alternative image-processing techniques, a monocular and a binocular, to take advantage of the arising redundancy. The emphasis is here to enhance the robustness of the vision process to make the overall system reliable. To ensure correct feature detection, image and model cues (e.g. geometrical order of features or feature intensity) are integrated [8] [9]. After feature extraction, the pose estimation algorithm can use both 3D and 2D feature information found in the images to calculate the current pose and send it to the robot. Figure 2 illustrates an example for a reference robot trajectory through the ship structure, which is determined prior to the operation of the RobVision system.

(a) (b) (c)

**Figure 1: (a) the 8-legged pneumatic walking robot (b) the Eurohead: mechanical design and (c) implementation**

The following section gives a more detailed overview of the Robvision components and their functionalities. In Section 4 an outline of the system demonstration is given and the summary of the results is presented in section 5.

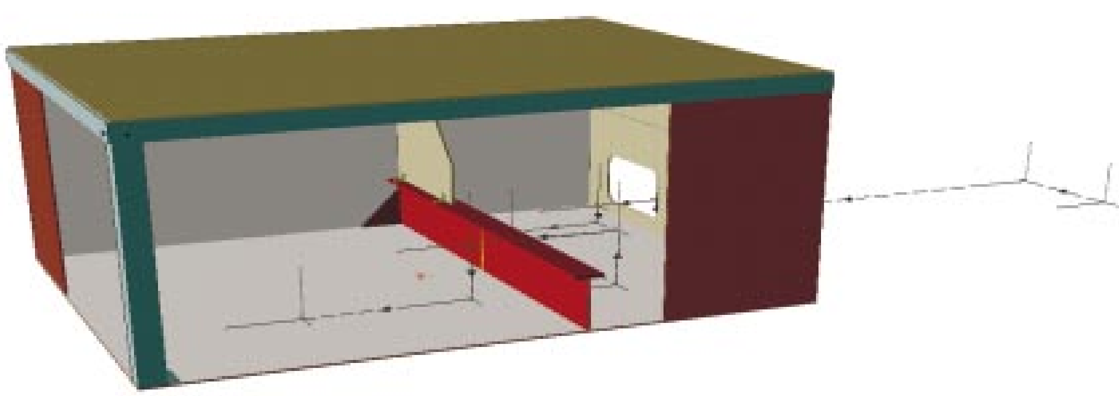


**Figure 2: A CAD-model and the Reference Robot Trajectory**

## 2 The System Components

Figure 3 illustrates the main functionalities of the entire system with model and trajectory data as inputs. The model of the ship environment is converted to the internal format by the *CAD System* and are in the following called model features. These model data are input to the *Vision* and *Pose Calculation* function. The CAD tool is also responsible for the continuous generation of image feature list, which is the actual appearance of the model from the camera viewpoint, and the according viewpoint itself which specifies a 3D point (x,y,z) towards which the camera ought to look at.

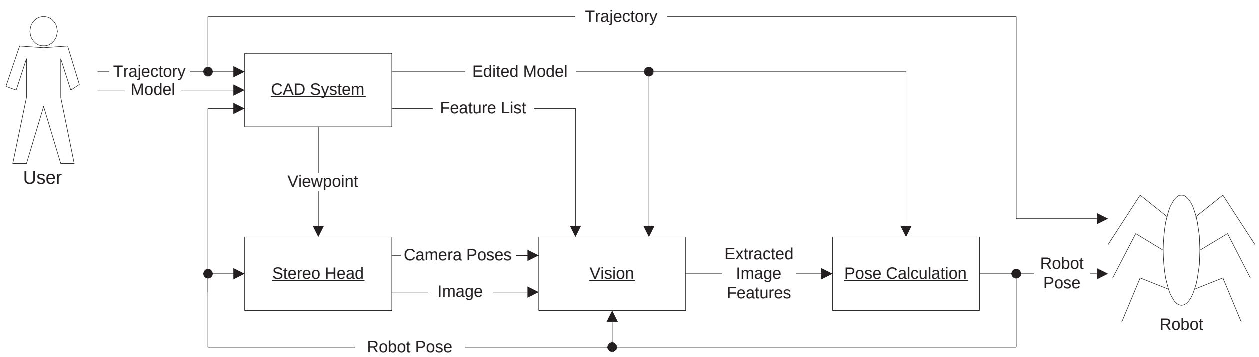


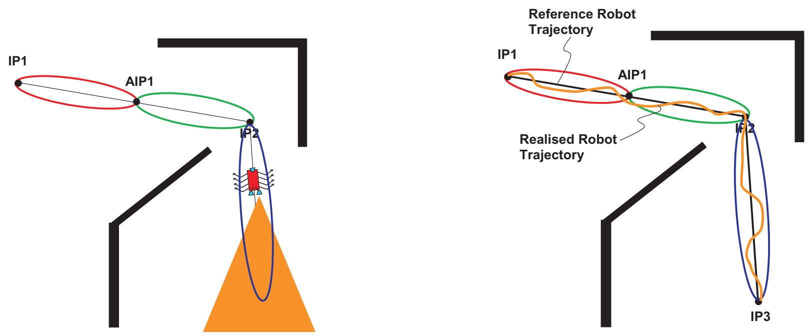

**Figure 3: System Functions**

The direction to the selected viewpoint is realized by the *Stereo Camera Head*. This camera pose and the images of the cameras are send to the vision process. Vision uses its inputs (edited model, feature list, camera pose and an estimated robot pose) to find the actual position of the features in the image. These feature positions are matched to the model features in the pose calculation function to generate the actual robot pose. This pose is the basis to calculate a robot motion to fulfil the trajectory predefined. In the following subsections the internal functions are presented in more detail.

### 2.1 CAD System

The main function of the CAD system *C2V* (Cad to Vision), developed by Department of Production, Aalborg University is to provide the vision and pose calculation component with geometrical features. These features specify the shape and location of geometrical entities such as lines, junctions and regions that the cameras can expect to see in the camera images while the robot is moving inside a steel structure.

C2V consists of two components, namely: "Determine Feature Table" and "Perform Feature Table Look Up". The first component "Determine Feature Table" generates features for the entire trajectory and stores this in a "Feature Table". The functionality is performed off-line in advance because of the large and time consuming algorithms needed. In this off-line phase of C2V a model of the robot is placed in the starting location of the reference robot trajectory. A viewpoint is determined and the features associated with this view are derived and generated. The robot modelled is then moved forward along the reference robot trajectory while looking




(a) (b)

**Figure 4: (a) the same features are visible along areas of the trajectory, (b) the Realised- and the Reference Robot Trajectory**

at the viewpoint until new robust features become visible and the old robust features become invisible. A new viewpoint is then selected together with a new set of visible features. The robot modelled then continues along the trajectory looking towards the new viewpoint. This approach divides the trajectory into areas where the same viewpoint and features can be used (Figure 4a).

In the online phase of C2V "Perform Feature Table Look Up" the estimated robot pose generated by the pose calculation component is used to identify in which of the areas generated by the off-line system the robot is presently located (Figure 4b). The according viewpoint and the features associated with this area are sent to the stereo head, vision and pose calculation components. This approach requires that the deviations of the robot trajectory realized are less than the areas generated by C2VoffLine. Simulations show that large deviation on 500 mm or more, depending on conditions, can be acceptable.

## 2.2 Vision

The basic goal of the vision component is to extract the positions of geometrical features from the camera image robustly. To fulfil this vision task two different approaches are applied.

The first method build in the *PRONTO* component, developed by Laboratory for Integrated Advanced Robotics, University of Genoa uses stereovision. The main idea of this approach is that of measuring the depths of edge junctions. A Hough technique is implemented to extract the lines on the image planes of the stereo pair. A weighted LMS (Least Mean Square) method is used to relate them to the features provided by the CAD system [4]. Then a closed loop method is followed, so that by moving simultaneously the three degrees of freedom of the head the junction is put at the principal point of the image in both images. When this is the case the two cameras are verging on the certain junction and the direct kinematics of the head are applied, in order to determine the 3D position of the junction relatively to the head.

The second vision method build in the *V4R* (Vision for Robotics) component, developed by Institute of Flexible Automation, Vienna University of Technology is a monocular search and

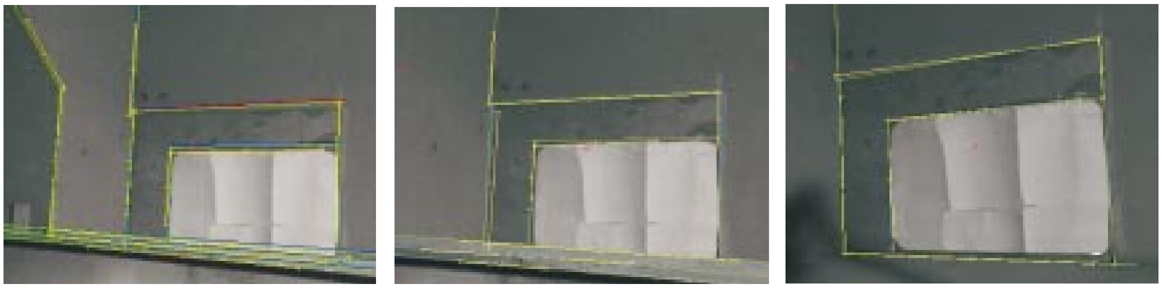


**Figure 5: A tracking sequence along the robot trajectory: Projection of the model and result of feature finding**

tracking strategy that is based on a local windowing technique. Local methods have the advantage to save computing time and therefore to keep the cycle-time lower than the frame rate of the camera (40ms) [5]. For a local method it is important to know the initial area to search for a feature. The pose calculation of the last step enables the projection of the model features into the image to get a clue of the rough feature location. Around this location a small area of interest is searched for the feature specified by the CAD system.

V4R is presently able to track features like lines, junctions, ellipses and arcs. The method used is a combination of cue integration strategies (with cues like intensity values, color, texture) to get a representative collection of image edgels and the RANSAC-method (Random Sample Consensus) for Edge finding [1] [9]. Once a feature is found for the first time, this feature is then tracked in the next cycle. Additional information of the features found in the image (image cues like intensity) can be stored to facilitate the search for the same feature in the next tracking cycle. An example for a tracking sequence is given in Figure 5.

The outcome of the vision task is the measurements of the features (both the distance measurements and the feature positions in the image) that are afterwards the input of the next functional unit - the pose calculation.

### 2.3 Pose Calculation

The pose calculation function is a part of the V4R component and computes the position and orientation of the robot related to the environment object, i.e. the ship. It is possible to use 2D image feature positions extracted in V4R itself and 3D distance measurements delivered by PRONTO stereo vision component. The method implemented fits these image features extracted to the according model received from the CAD system. The algorithm is based on a method proposed by Wunsch [10] and was modified for providing accuracy estimation of the calculated pose as well as outlier detection [6]. The resulting robot pose together with the predefined robot trajectory is then the basis for the motion planing of the robot.

### 2.4 Stereo Head

The task of the stereo head is the realization of the correct camera orientation, image acquisition and image stabilization [7]. It is a part of the PRONTO component. The hardware

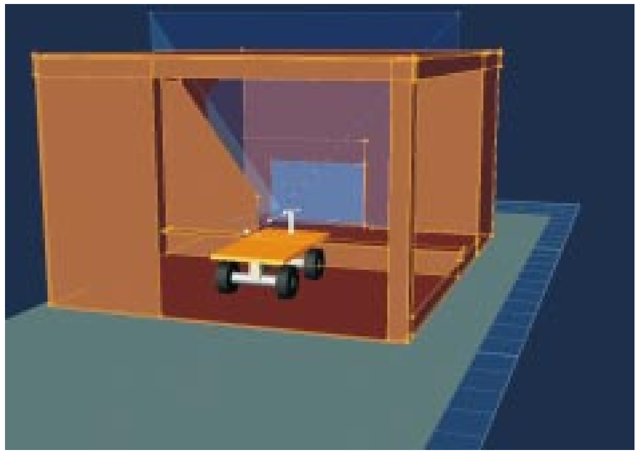


**Figure 6: Model of the mockup where the cameras are looking at a viewpoint**

module (the Eurohead of LIRA-Lab) is shown in Figure 1b,c. It has been designed and implemented to be an accurate vision-based measuring device. For the control of the pan, tilt and vergence four harmonic drive actuators are used with harmonic drive gearing and zero backlash. To initialize the head a calibration process is needed using hard-stops included in the head [4].

During the movement of the robot, PRONTO concentrates its computational efforts to stabilize the head and to keep the gaze of the cameras fixated on the view point recommended by the CAD system. The stabilization is performed using angular accelerometers that react to the movements of the robot. This consents PRONTO to calculate the angular velocity by integrating the acceleration in a 40 ms cycle to produce a compensation movement.

The gaze of the head is maintained in the direction of the viewpoint in a 240 ms cycle using the robot pose feedback generated by V4R. Both, stabilization and gaze orientation have the purpose to guarantee a good image quality to increase the robustness of the tracking algorithm described in section 2.2.

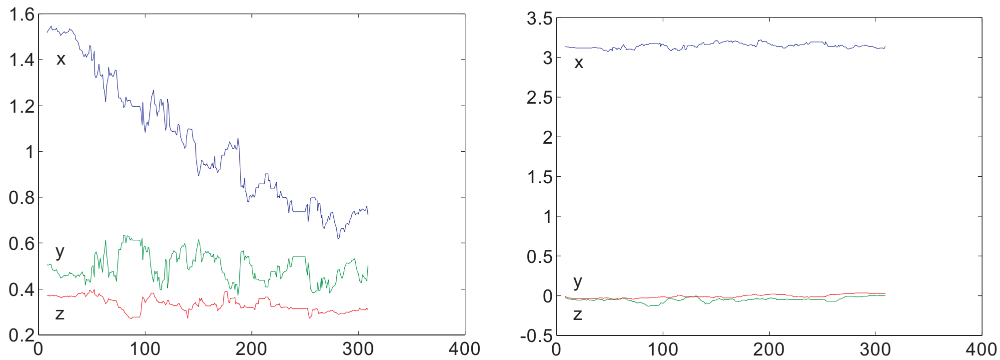
## 3 Demonstration

This section gives an overview of the tests for evaluating the whole system. As seen in Figure 6 a mock-up - a typical test environment - was built with structures that can be found in a big ship environment.

Several trajectories were examined to test the robustness of the system and its accuracy. Malfunction can happen since the mock-up is made of big metal plates welded together, that is a big defiance for the vision. Some welding causes irregular edge features, contrast is sometimes poor and additional features on plate surfaces can affect pose calculation results negatively. Due to these problems feature redundancy is a basic issue. Figure 7 shows an example for a view. C2V generates the features seen in the image to deliver it to the vision systems (Figure 7a). PRONTO is able to extract the 3D coordinates of the junctions (Figure 7b) and

(a) (b) (c)

**Figure 7: (a) View with features generated by C2V, (b) 3D junctions detected by PRONTO, (c) lines in their searching windows generated by V4R**




(a) (b)

**Figure 8: Robot pose output along a trajectory, (a) position, (b) orientation**

V4R the 2D position of edges and junctions in the image (Figure 7c). From all the features extracted a pose is calculated. Figure 8 gives the poses calculated along a path.

As can be seen the pneumatic walking robot produces many jerks. Especially fast changes in the orientation of the robot cause big deviations of the feature positions in the image. Because of the fast but spatially restricted windowing technique such features are then lost. This fact reduces the reliability of the system. The head stabilization implemented lower the optical flow and increases the robustness of feature finding.

The tests executed show the system benefits from the redundancies implemented. As long as enough features can be tracked, the system is able to re-find lost features. Wrong detected image features can be filtered out since there are enough features which indicate the right pose. Finally the stabilization of the pose increases by including additional measurements like depth information of the image features.

## 4 Summary

During the RobVision project a prototype of a visual navigation tool has been built to guide a walking robot through a big steel construction. It can be shown that applying CAD-model information in combination with visual sensing input enables the extraction of continuous robot poses. The model delivered by the user is adapted to allow a selection of features desired. Two separate vision systems utilizing different approaches guarantee enhancement through redundancies. The jerks of the pneumatic walking robot is a big challenge for the vision but is partly compensated by including accelerometer data. The average position accuracies reached are 5.49mm in X-direction 3.29mm in Y-direction and 8.48mm in Z-direction. That means that the overall 3D position accuracy reached on average was 10.62mm.

Future work includes the integration of linear accelerometers and inclinometers to the existing gyros, in order to perform a more efficient and accurate image stabilization. Moreover, since the robustness of the pose estimation method relies on the redundancy of the features measured, more 3D features extracted by stereoscopic vision, such as lines in 3D [2] and the normal of surfaces, are under investigation. Extensions of feature validation techniques to reach more robustness of the vision is another important issue. Together with the improvements of the vision capabilities, future work is also necessary to enhance error recovery and fault tolerance behaviours.

As a conclusion it can be stated that CAD based vision is a potential position-sensor in normal industrial environments. To reach the reliability demanded multiple views, redundancies in image processing strategies and feature validation techniques are necessary. Still enhancements are necessary to achieve an overall reliable system behaviour.

## References


[1] Fischler M.A., Bolles R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography; Communications of the ACM Vol.24(6), pp.381-395 (1981).

[2] Gasteratos A., Beltran C., Finding the 3D Orientation of a Line Using Hough Transform and a Stereo Pair, Technical Report, TR 3/00, LIRA-Lab, DIST, University of Genoa, Genoa, 2000.

[3] Gasteratos A., Martinotti R., Metta G., Sandini G., Precise 3D Measurements with a High Resolution Stereo Head. IWISPA 2000, Pula, Croatia, 2000, 171-176.

[4] Gasteratos A. , Sandini G., On the Accuracy of the Eurohead, LIRA - TR 2/00, July 2000.

[5] Hager G. and Toyama K. (1998): The XVision-System: A Portable Substrate for Real-Time Vision Applications, Computer Vision and Image Understanding, 69 (1), 23-37.

[6] Kraus K.: Photogrammetry Volume 2, Advanced Methods and Applications, Dümmler/Bonn, pp.205-215, 1997.

[7] Panerai F., Metta G., Sandini G., Visuo-inertial Stabilization in Space-variant Binocular Systems, Robotics and Autonomous Systems, 30 (2000) 195-214.

[8] M. Vincze: Robust Tracking of Ellipses at Frame Rate; Pattern Recognition 34(2), 487-498, 2001.

[9] M. Vincze, M. Ayromlou, W. Kubinger: An Integrating Framework for Robust Real-Time 3D Object Tracking; Int. Conf. on Vision Systems, Gran Canaria, S. 135-150, January 13-15, 1999.

[10] Wunsch P., Modellbasierte 3-D Objektlageschätzung für visuell geregelte Greifvorgänge in der Robotik, PhD Thesis, Munich University of Technology, pp.37-47, 145-149, 1997.