Model-Driven Vision for In-Door Navigation

Henrik I. Christensen, Niels O. Kirkeby, Steen Kristensen, Lars Knudsen

Laboratory of Image Analysis,
Institute of Electronic Systems,
Aalborg University
Fr. Bajers Vej 7, Bldg D1
DK-9220 Aalborg Denmark

(*Running title*: Model-Based Navigation)

Abstract

For navigation in a partially known environment it is possible to provide a model that may e used or guidance in the navigation and as a basis for selective sensing. In this paper a navigation system for an autonomous mobile robot is presented. Both navigation and sensing is build around a graphics model, which enables prediction of the expected scene content. The model is used directly for prediction of line segments which through matching allow estimation of position and orientation. In addition the model is used as a basis for a hierarchical stereo matching that enables dynamic updating of the model with unmodelled objects in the environment. For short-terms path planning a set of reactive behaviours are used. The reactive behaviours include use of inverse perspective mapping for generation of occupancy grids, a sonar system and simple gaze holding for monitoring of dynamic obstacles. The full system and its component processes are described and initial experiments with the system are briefly outlined.

Keywords: Model Based Navigation, Active Vision, Reactive Control

1. Introduction

For interaction with a partially known environment it is necessary to equip machines with sensory modalities that allow dynamic (and on-line) modelling of the environment to enable planning and motion. In control of stationary robots it is common to use specially engineered vision systems in combination with tactile sensors [7]. For mobile robots short range navigation is often based on use of ultrasound sonars [6], while long range navigation is based on use of specialised robot vision sensors (triangulation with respect to landmarks), or dense depth maps obtained from laser range cameras [1,2]. To enable a more versatile use of mobile robots for exploration, flexible transportation, etc. it is desirable to equip these robots with flexible vision facilities for modelling of the environment.

In this work, in-door navigation in a quasi-structured but dynamically changing environment has been selected for experiments. In this domain it is assumed that major structures, such as walls remain static, while small structures such as tables, chairs, humans, etc. may change position over time.

For modelling of the environment (for a mobile robot) the requirements will depend on the task that is to be carried out. The tasks/requirements may be divided into the following categories:

1) Maintenance of the robot position with respect to the environment
2) Detection of obstacles & control of the avoidance process
3) Image driven navigation (e.g., servoing on a landmark)
4) Geometric modelling of major obstacles for recognition and/or planning.
5) Recognition of objects/structures in the environment

These categories are not mutually exclusive but indicate tasks that could be carried out by a mobile robot. These five tasks each have different requirements with respect to the need for modelling of the environment, maintenance of acquired information, use of qualitative and quantitative information. Experiments using this set of tasks can thus provide insight into various

methods for control at a task level. The experiments that have been carried out have mainly been related to description of static phenomena perceived by a moving observer. Special care has been taken to enable investigation of the same methods in dynamically changing scenes.

Initially the testbed which forms the basis for the system is described. Each of the component vision sub-systems are then described together with associated experimental result. Finally a summary is provided.

2. Outline of Demonstrator System

The experiments are carried out using a commercial mobile platform equipped with a binocular camera head. The platform is a ROBUTER 20 platform from Robosoft. The platform is equipped with a set of 24 ultrasound sonars (which are not exploited in the reported work), and has a built-in computer for handling of simple motion commands.

The on-board binocular camera head is based on 4 commercial rotational stages which implement pan and tilt for the complete rig and independent vergence for each of the two cameras. In addition the cameras are fitted with motorised lenses that enable change of focus, zoom, and aperture.

For control of the combined system, the platform is equipped with a VME rack that contains a UNIX-V computer and an OS-9 computer (MC 68030), the OS-9 computer is used for control of the different degrees of freedom in real-time. The rack also contains a commercial PID controller board that interfaces to the rotational stages, while dedicated hardware is used for control of the lenses. Software allows control of the individual degrees of freedom or coupled control of all (i.e., fixation on a scene object driven with slaved accommodation). The robot is connected to stationary computers through serial and Ethernet links. Presently image processing is carried out off-board due to considerations related to power consumption.

The platform with the on-board binocular camera head and the computers for real-time control of both are shown in figure 1.

**** FIGURE 1 ABOUT HERE ****

Control of the vehicle and planning (navigation) based on visual information is carried out using a structure as shown in figure 2.

**** FIGURE 2 ABOUT HERE ****

Control of the platform, the camera head and the action selection will not be described in detail here. Each of the four visual modules to the left in figure 2 will, however, be outlined in the following sections.

For the planning a rule based system is used. A set of primitive actions (determine position, servo landmark, detect obstacle, avoid obstacle, and resume path) has been defined, and the planning selects visual behaviours to accommodate actions. Coordination/planning of action is performed through use of a rule based system which exploits discrete event dynamic system (DEDS) theory as described by Kosecka et al. [14].

3. Model Based Navigation

For in-door navigation static structures like walls are modelled a priori. When such structures are modelled by means of a 3-D CAD model, it is possible to make projections corresponding to a prediction of what a camera is expected to see from a given view-point. The projection can then be used for matching against line segments extracted from images. Once a set of matched line segments has been obtained it is possible to perform least square fitting, which in turn allows estimation of the position and the orientation of the robot [15]. A system based on similar ideas has recently been presented by [13].

To test the methodology outlined above a system has been implemented. Here a section of the Laboratory of Image Analysis has been modelled in a home grown CAD system that allows extraction of both 2-D and 3-D line structures after use of a hidden line removal algorithm. An example projection is shown in figure 3.

**** FIGURE 3 ABOUT HERE ****

To limit the computational demands for image processing line segments are only extracted in the proximity of predicted line segments. Presently fixed sized windows are used, but Kalman filtering can be included for updating. Lines are located through selection of a set of reference points along the predicted line structure. At each reference point a search perpendicular to the predicted line is carried out. If a zero crossing is found in the second order derivative at a location where the

gradient is above a threshold *T,* it is assumed to represent the location of the edge at the reference point. Once edges have been detected, at the reference points, a least square optimisation is used for estimation of coefficients for the image lines. The matching of line structures between an image and the model is based on simple proximity. Other more advanced methods, such as relaxation could have been used, but due to speed considerations a simple method was chosen. Kumar [15] has described a more elaborate matching scheme based on graph search, but this method has proven computationally demanding. An example image with extracted line structures superimposed is shown in figure 4.

**** FIGURE 4 ABOUT HERE ****

The method has been implemented in experimental software. The sub-system is implemented as separate modules that are build into a complete system through use of the VIPWOB software package [9]. The system is capable of doing matching once the position is known with an accuracy of ±5 cm and an angular accuracy of ±2°. The need for accurate initial estimates is due the perspective effects, where changes in position will results in major images translations for structures close to the robot. For handling of motion between subsequent images, which is larger than 5 cm, odometry information from the robot is fed into the system as part of the prediction.

The implementation demonstrates that it is possible to simplify image processing through use of predictions, and the exploitation of model information allows simple reconstruction of the 3-D position and orientation of the robot. The accuracy of the estimate is within ±2cm and ±0.5°, which is considered adequate.

4. Cooperative Stereo

Estimation of distance from a camera to structures in the scene is typically performed using stereo techniques [5]. A well-known problem in stereo is correspondence, where features, in two or more views, that represent the same physical object, are matched. Once a list of matching features has been produced it is possible to estimate disparity and depth using simple triangulation combined with information about the camera system (calibration data). To achieve a good accuracy in the reconstruction it is necessary to use low-level features such as edgels, but use of such primitives results in a large number of features that must be matched and the process consequently becomes

computationally demanding, and matching errors will frequently result in a poor reconstruction. Use of coarse features simplifies the matching but accuracy is sacrificed. To achieve both robust and efficient matching, and an accurate reconstruction hierarchical multi-primitive matching methods may be exploited.

Marapane & Trividi [16] have presented a method for hierarchical multi-primitive matching. In this method images are initially segmented into regions. Regions from left and right images are then matched. This provides a coarse estimate of scene disparity. For the area around the perimeter of regions an edge detection is carried out to allow detection of line structures. The line structures within a pair of matched regions participates then in a correspondence analysis. The results of this process enable estimation of planar polyhedral structures. If a more accurate estimate is necessary it is further possible to match corner points and use these for reconstruction of depth.

This above approach has been adopted in a stereo system. A simple queue based segmentation algorithm, where similarity is based on intensity differences within a 4 neighbourhood, is used for the pre-processing. Once regions have been obtained a correspondance analysis between image regions and predicted regions is carried out. The predicted regions are generated by the CAD system mentioned in the previous section. All regions matched to the prediction are interpreted as being part of the prior model. Image regions associated with static structures are used by the position/pose estimation sub-system outlined in section 3. The method is illustrated in figure 5.

**** FIGURE 5 ABOUT HERE ****

A problem in stereo reconstruction is occlusion boundaries, which may result in incorrect reconstruction around such contours. A simple example is shown in figure 6. To cope with such phenomena one may detect occlusion cues (i.e., T and λ junctions). For each of the positions where an occlusion might be present the system performs a fixation, using the binocular head, and a verification of depth is initiated through analysis of the response during controlled change of focus. The verification is carried out across both of the lines that make up the occlusion cue. If both edge segments have a high frequency peak at the same focus setting the junction is related to a physical corner while peaks at different locations indicate that an occlusion is present. For this situation the depth estimate obtained from focusing is used.

****** FIGURE 6 ABOUT HERE *******

The method has been tested with success on a number of in-door scenes. The strategy of using accommodation (focus) for verification at selected scene locations is only useful when a small depth of field of view can be obtained with the optical system. To achieve this in an in-door setting where the depth range typically is 1-3 meters (for the obstacles of interest) the motorised zoom is exploited once fixation has been obtained. The focal length is changed to a maximum, which results in a minimum depth of field of view. As a criteria function the maximum gradient magnitude described by Krotkow [12] is used. The verification of depth is achieved through small scale changes of focus, which allow estimation of the presence of a local extrema in the criteria function. For the estimation of depth the full focusing range is traversed in a search for a global extrema, corresponding to the best depth estimate, the procedure is described in detail in [12].

When a set of planar surfaces has been detected in an image it is added to the CAD model. In the prediction of the next image both old and new information is used. Now if the surfaces which have been detected earlier are detected again the parametric data are updated using a Kalman filter in which X, Y, and Z coordinates for corners are maintained using a common covariance matrix. An example of processing using the stereo system is shown in figure 7. The obtained results are described in further detail in [11].

**** Figure 7 ABOUT HERE ****

The experiments involving model based stereo and navigation has demonstrated how the two sub-systems complement one another. The stereo system continuously provides updates to the model, while the latter exploits the model for estimation of its position and the information is in addition used for global path planning.

5. RECOGNITION OF LANDMARKS

For navigation through the environment it is necessary to provide facilities that allow recognition of a priori defined landmarks. Such landmarks may represent locations that are used for bootstrapping (coarse estimation of the initial position of the robot), recognition of goal locations and homing. In

this work doorways have been defined as landmarks. The location of a doorway is used for estimation of the orientation of the robot with respect to a room. In addition by having access to information about the physical size of a doorway estimation of the distance to a doorway is possible, as for example described by Kanatani [10]

When the robot has to move from one room to another, the doorways are in addition used as the reference position for a visual servoing operation in which the robot tries to move towards the door while maintaining the structure in the centre of the image. Motion of the camera head is here slaved to the robot. I.e., initially the robot tries to change its orientation so that the camera points directly ahead. For each camera fixation on the landmark is maintained and any difference in the angles between the two cameras is used for a change in orientation of the platform.

Landmarks are represented as a rectangle where the height/width ratio of approximate 2.5. Objects such as racks (box like), etc. may also be recognised with this procedure, if other parameters are provided.

For the recognition of rectangular structures, such as a doorway, a line extraction process is used. Presently a morphological edge detector [8] is used, a better performance in terms of edge quality may be obtained with the Deriche operator [4], but at the expense of speed. Once a set of line structures has been detected, pairs of co-linear lines are detected. Such lines might represent the edge on both sides of a door opening (figure 8b). In parallel a search for pairs of parallel lines is carried out. These lines might represent each side of the door frame (figure 8c). The set of parallel lines is afterwards grouped with the set of colinear lines.

***** FIGURE 8 ABOUT HERE *****

For groupings where the parallel lines are end-to-end with the colinear lines a door hypothesis is generated. The remaining part of the door frame is subsequently verified and used in a ranking of the hypotheses. The best hypotheses are asserted as positions of landmarks, which subsequently are used for triangulation (in combination with the prior model) for estimation of the approximate position of the robot. When a landmark has been recognised it may also be used for visual servoing in navigation (the selection will depend on the present task). The landmarks are used for navigation

in non-cluttered environments, where it can be expected that landmarks are easy to recognise and there is little chance of hitting obstacles.

For maintenance of the description of a landmark a simple line based verification driven search is used. In this search the line segments, which define the doorways, indicate the estimated location of the door and a search perpendicular to these lines is used for relocalisation in subsequent images.

**** FIGURE 9 ABOUT HERE ****

The landmark recognition procedure has been applied in the in-door scenario described earlier. Example images from a sequence are shown in figure 9. The left most illustration shows the original grey-scale image, while the next illustrates some of the lines extracted. The next image has a superimposed representation of all the hypothesised landmarks, and the right most image has a superimposed structure representing the landmark hypothesis, which has received the most support.

In the extraction of landmarks, like doorways, it has been noticed that for close-by structures the fine structure of the door frame will result in a large number (>20) of line segments. Often four to five parallel line structures are extracted, which in turn results in a number of coincident landmark hypotheses (and a redundant search). To avoid sets of overlapping door hypotheses a coincidence test is used for removal of hypotheses which overlap (in image space) with the best hypothesis.

6. Image Driven Navigation.

The purposive modules presented in section 3-5 can also be used for close range navigation. For such application it is, however, more efficient to use a simple technique such as an inverse perspective projection method [17]. The basic geometry of a binocular set-up is shown in figure 10.

**** FIGURE 10 ABOUT HERE ****

When calibration data are available it is possible to perform an affine transformation of the left and right images to provide a map corresponding to a view directly from above. If the transformed left and right images are subtracted from one another all structures located in the ground plane will be at the same location and the difference image will at such locations contain zero values. For structures elevated above the ground plane a positive difference will be detected. In an implementation, proper consideration of noise and inaccurate calibration, will result in non-zero values even at locations where structure in the images is located in the ground plane. Through use

of global thresholding it is, however, possible to reliably detect obstacles. An example doorway is shown in figure 11a and c. In figure 11b is shown the corresponding difference image obtained after the affine transformation.

**** FIGURE 11 ABOUT HERE ****

The difference image in figure 11 illustrates that it is possible to detect the card board box in front of the robot (the small black region at the bottom of the image) and the doorway (the separation in the middle of the image). The method thus provides a simple method for generating a map of the immediate surroundings of a platform. This method is an example of a purposive module which allows fast computation of a (low abstraction) map, but where it is faster to recompute the map rather than storing it for later use. I.e., an example of a purposive module where contextual information is of little utility.

7. Summary

A set of different modules for a purposive vision system has been outlined. The different components are all designed for application in the domain of mobile robot navigation in an in-door environment, but they have been carefully implemented to allow use in other domains as well. The component processes are integrated using the VIPWOB package [9] available in public domain. For the integration and control of the processes a discrete event formalism is adopted. This formalism has been described in detail in[14].

Bibliography

[1] F. Arman, and J. Aggarwal, Model Based Object recognition in Dense Range Images - A Review, , ACM Comput. Surveys, 25 (1) (1993) 5-44.

[2] C.S. Andersen, C.S., C.B. Madsen, J.J. Sørensen, N.O.S. Kirkeby, J. Jones, and H.I. Christensen, Navigation using range images on a mobile robot, Robotics & Autonomous Systems, 10 (12) (1992), 147-160.

[3] M. Andersson and A.Lundquist. Tracking lines in a stereo image sequence, Proc. 7th Scan. Conf. on Image Anal. , Aalborg, August (1991).

[4] R. Deriche, A Recursive Implementation of the Canny Edge Detector, Intl Jour. Computer Vision, 2 (1) (1988).

[5] Dhong and Aggarwal J, A Review of Stereo Matching Techniques, IEEE Trans. on SMC, XX (12), (1989).

[6] H. Durrant-Whyte H., Sonar Based Navigation, (Kluwer Press, Boston, MA. 1991).

[7] K. Fu, R.C. Gonzales and C.S.G. Lee. Robotics: Control, Planning and Sensing, (Prentice-Hall Inc., New York, NY. 1987).

[8] R. Haralick and L. Shapiro, Computer and Robot Vision, Vol. 1, (Prentice Hall Inc., Los Angeles, MA, 1992).

[9] Kirkeby N.O.S & Christensen H.I. The Vision Programmers Workbench, In: H.I. Christensen and J.L. Crowley, eds., Experimental Environments for Computer Vision, (World Scientific Press, Singapore,1994).

[10] K. Kanatani, Geometric Computation for Machine Vision (Oxford University Press, Oxford, 1993)

[11] S. Kristensen , H. Møller-Nielsen and H.I. Christensen,, A Cooperative Depth Extraction, In Proc 8th Scan. Conf. Image Anal., Tromsö, Norway, (1993).

[12] E. Krotkow, Active Computer Vision by Cooperative Focus and Stereo (Springer Verlag, Berlin 1991)

[13] Kosaka A & Kak A, Fast Vision-Guided Mobile Robot Navigation Using Model-based Reasoning and Prediction of Uncertainties, CVGIP. 56(3), (1992). 271-329.

[14] Kosêcká J, Christensen H I, & Bajcsy R, Discrete Modelling of Gaze Control and Navigation, Intl. Jour Comput Vision (March 1993)(submitted).

[15] R. Kumar, Model Dependent Inference of 3D Information From a Sequence of 2D Images, Ph.D Dissertation, COINS-Dept. Univ. of Mass. Amhurst, MA, 1990.

[16] S.B. Marapane and M.M. Trivedi, Region Based Stereo Analysis for Robotic Applications, IEEE Trans on SMC, 19(6), (1989) 1447-1464.

[17] H.A. Mallot, H.H. Bulthoff, J.J. Little, and S. Bohrer, Inverse Perspective Mapping simplifies optical flow computation and obstacle detection, Biol. Cybernetics, 64 (1991) 177-185.

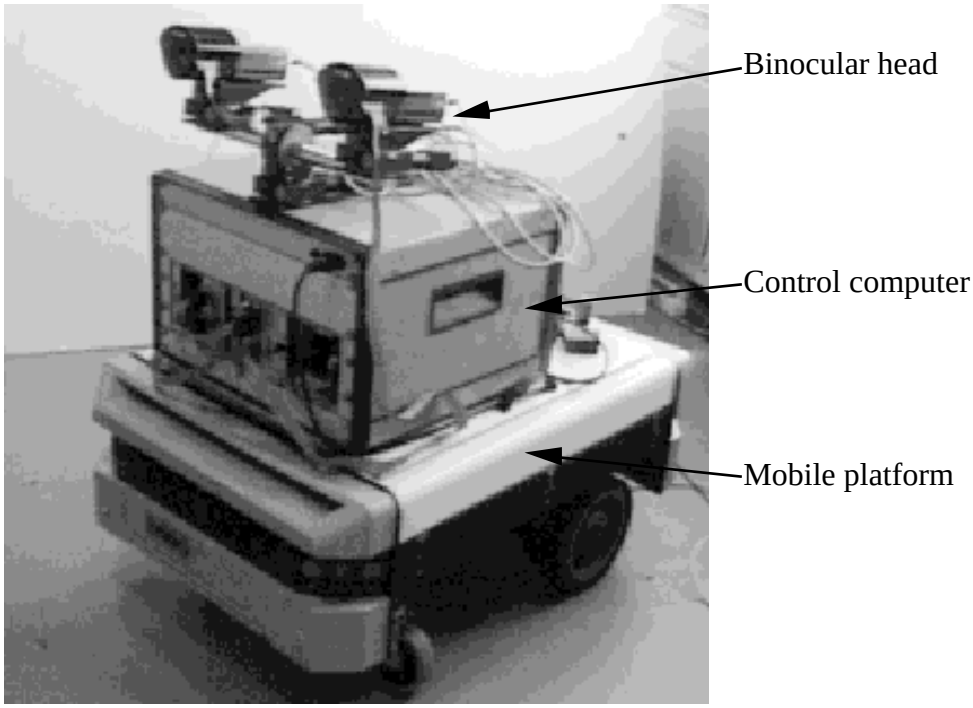


Figure 1. The AUC demonstrator system (ARVID).

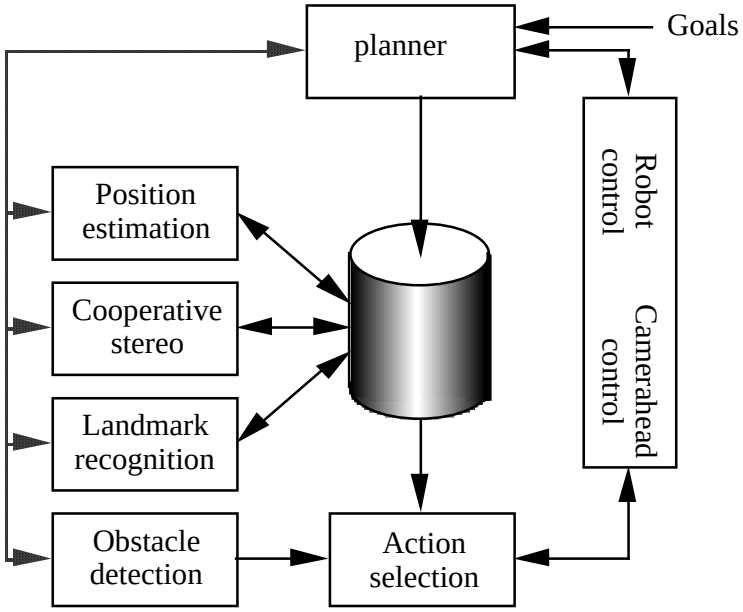


Figure 2. The vision sub-systems and their relation to the navigation system.

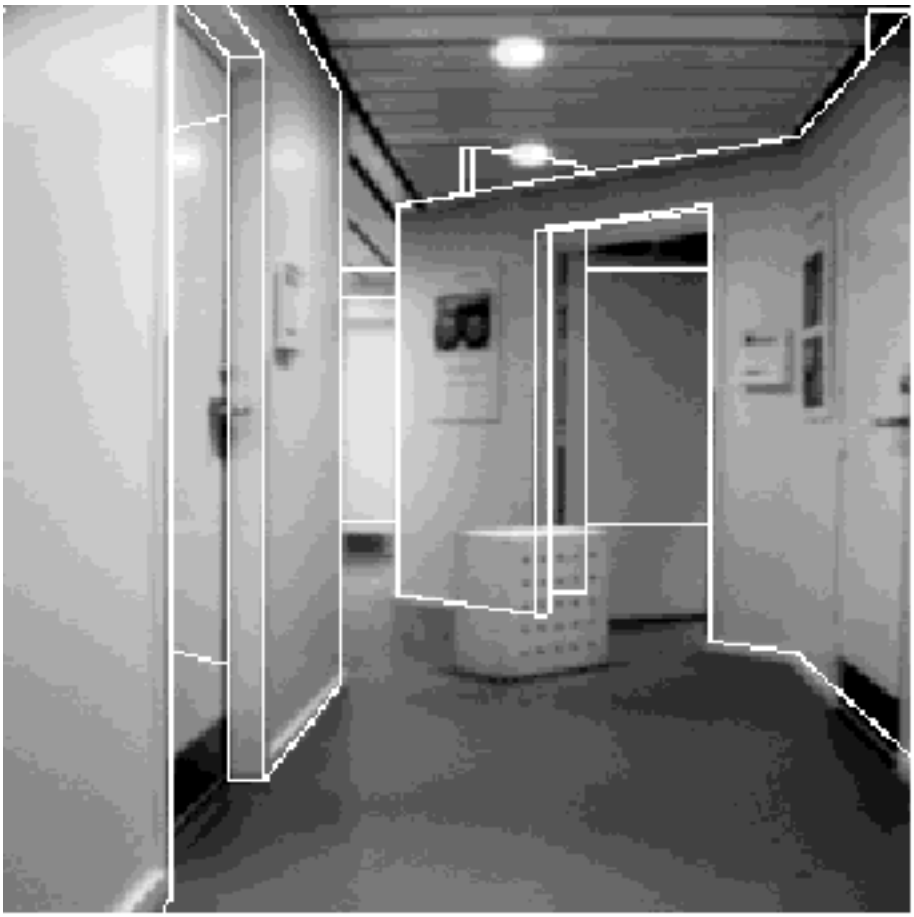


Figure 3. A projection from the geometric model.

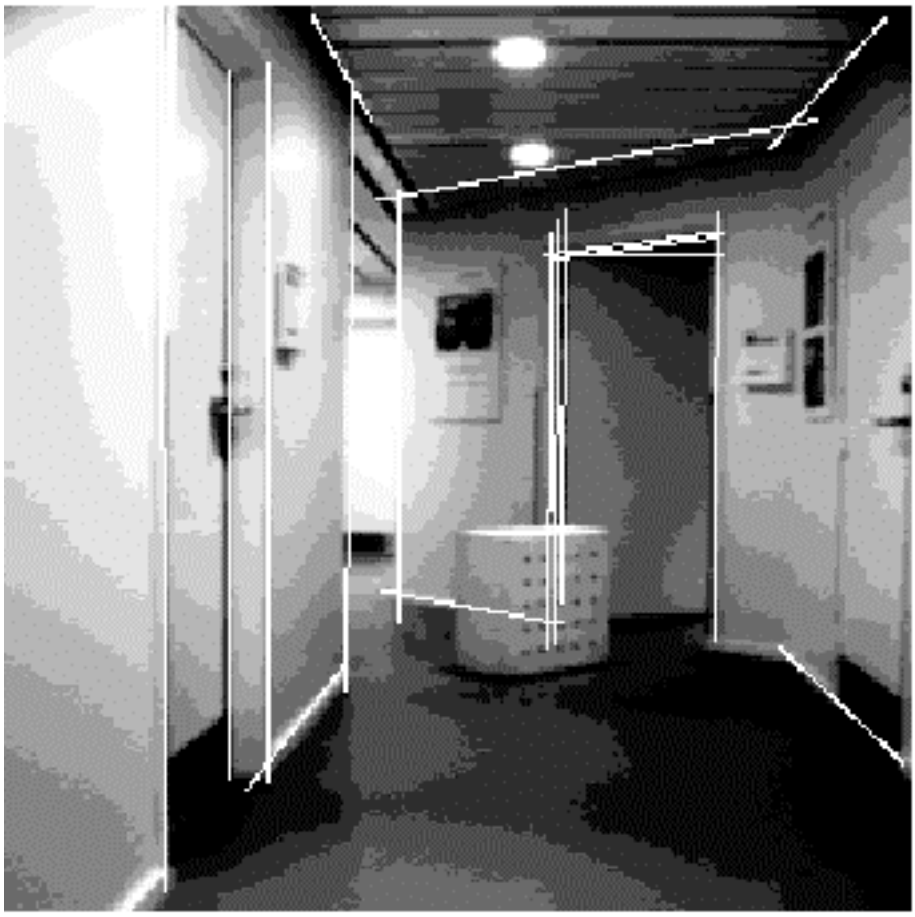


Figure 4. Image from the laboratory section with superimposed representation of extracted lines.

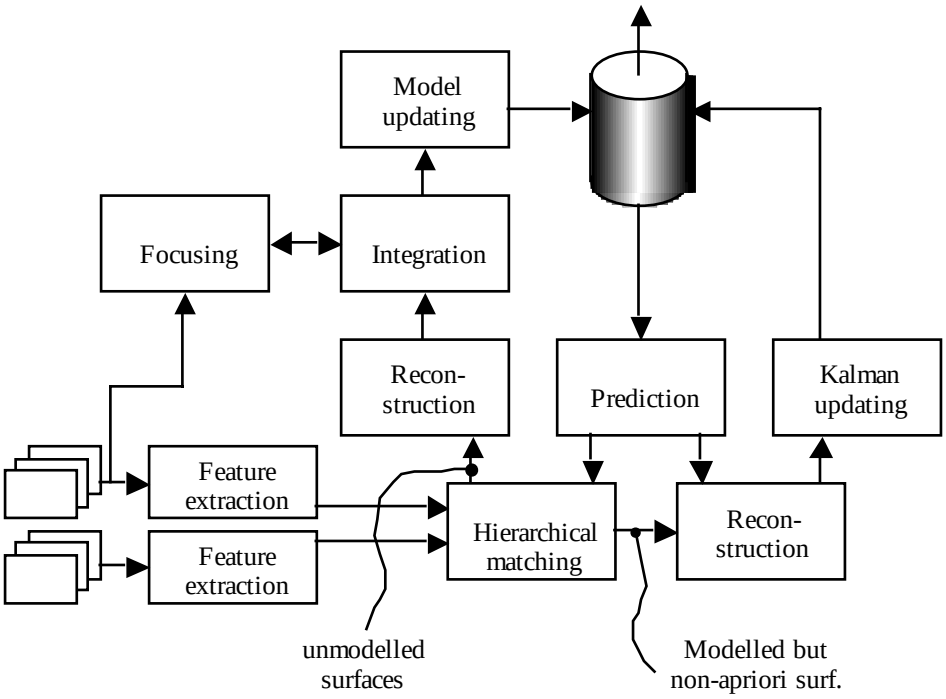
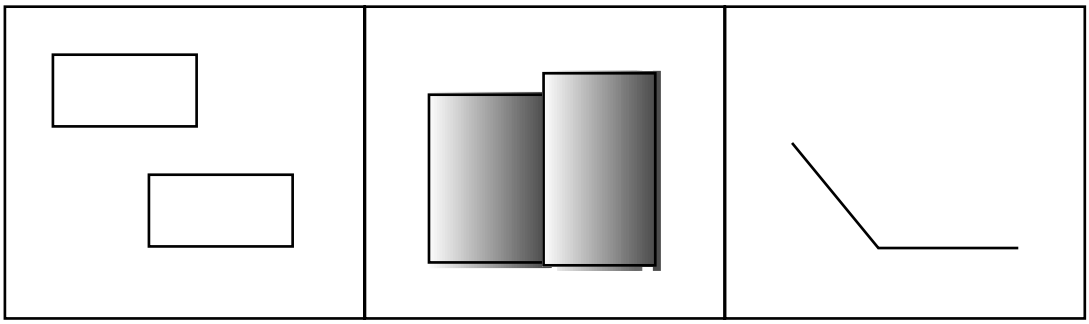


Figure 5. Block diagram for the cooperative stereo system.



Top view of scene Image from front Reconstruction

Figure 6. An example scene where an occlusion is present when view from the front. The left view shows the scene when viewed from the top. The middle view corresponds to the image. The right view to the reconstruction which is obtained when there is no explicit handling of occlusions.

Figure 7. Illustration of the continuous updating of the scene model based on binocular stereo matching. The left image is a grey level image of the scene, while the middle image illustrates the initial model. The right image illustrates the model after information about the obstacle has been added.

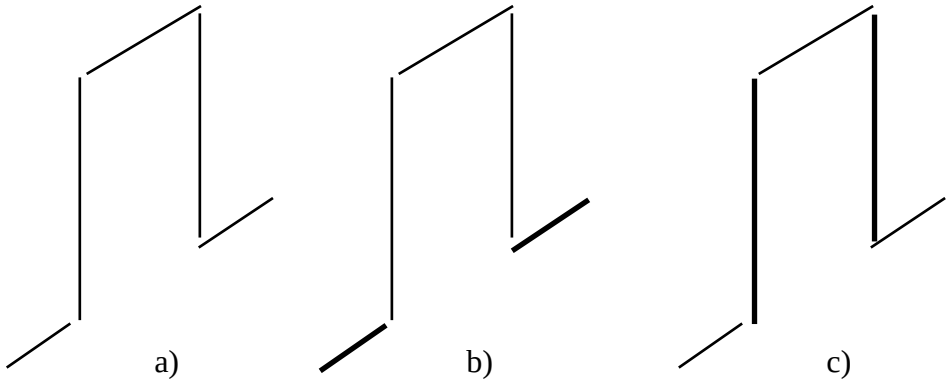


Figure 8. Outline of methods for recognition of landmarks (doorways). In figure a) line segments for the doorway are represented. In figure b) the segments used for detection of colinearity are shown. In figure c) the parallel segments from the landmark are shown.

Figure 9. Three different images where landmarks (doorways) are recognised (see text).

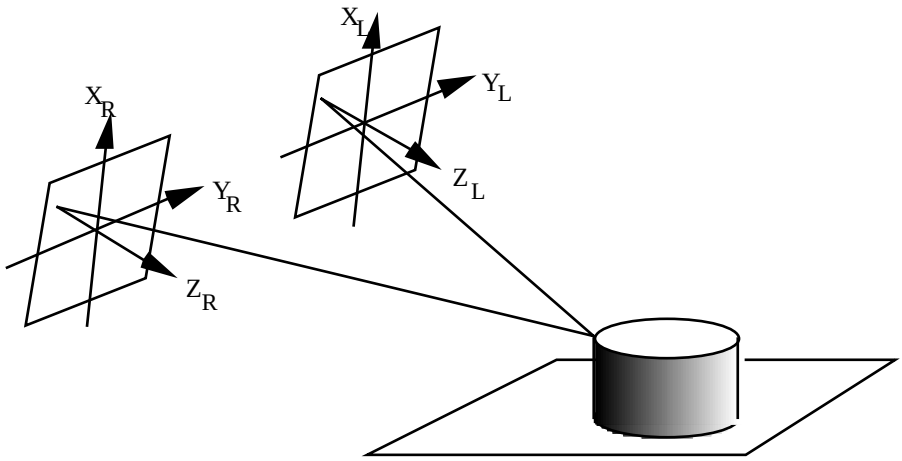


Figure 10. Geometry for a binocular set-up.

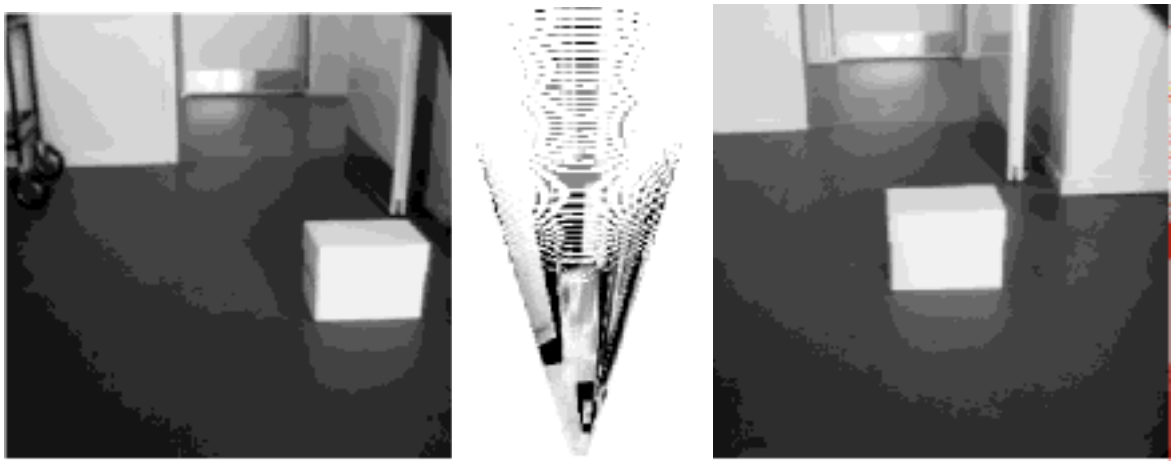


Figure 11. Binocular pair of images (a and c) with the corresponding difference image after transformation into the horizontal plane in figure b (white denote a small difference while black denotes a large difference).