Stereo Vision-Based Navigation in Unknown Indoor Environment

O. Schreer

Dipl.-Ing. Oliver Schreer, Sekr. EN11, Einsteinufer 17
D-10587 Berlin
e-mail: schreer@rtws18.ee.TU-Berlin.DE

Abstract. Different applications in the field of vision-based navigation of autonomous mobile robots depend on the degree of knowledge of the environment. Indoor environment applications often use landmarks or maps for navigation. Others have only knowledge of known and expected objects. In such applications, parts of the scene are classified in these objects, e.g. road junctions, doors, walls, furniture, and a possible path will be estimated. In case of a lack of *a priori* knowledge of the environment, we propose an approach for vision-based navigation, considering any reconstructed 3D point of the scene. A line segment stereo algorithm and a reconstruction procedure lead to uncertain 3D points of the scene in front of the mobile system. All these 3D points are regarded as obstacles and a following trace estimation will be applied on this 3D data. In order to increase the reliability of reconstructed 3D points, a validation step excludes impossible 3D points, exploiting the stereo geometry of the vision system. After validation a two-step analysis is applied, which contains the minimum distance method and point distribution analysis method. This analysis leads to a possible trace for the mobile robot, resulting in values for steering angle and velocity given to the mobile system. The method has been implemented on the experimental system MOVILAR (MObile VIsion and LAser based Robot) which is based on a multi-processor network. It achieves actually process cycles of approximately 1.5s and a velocity of about 10 cm/s, i.e. a slow walking speed. Experimental results of the above mentioned methods are presented.

1 Introduction

Different applications in the field of vision-based navigation of autonomous mobile robots depend on the degree of knowledge of the environment. Indoor environment applications often use landmarks or maps to navigate. In [9] a method is proposed which recognizes objects like doors, using them as landmarks. In applications like navigating in the middle of the road or in a corridor, the scene is analysed for road junctions [3] or bounds of a hallway [7]. With these methods, known objects are expected in the scene. In case of a lack of *a priori* knowledge of

the environment or expected objects we propose an approach for vision-based navigation, considering any reconstructed 3D point as an obstacle.

The basis of our investigation is a line segment stereo algorithm and a 3D reconstruction procedure which is illustrated in the next chapter. This step leads to uncertain 3D points of the scene in front of the mobile system which are the basis for the trace estimation. In order to increase the reliability of reconstructed 3D points, a validation step is applied which is explained in chapter 3. The stereo geometry of the vision system is exploited to exclude impossible 3D points, i.e. false matches. In chapter 4, the two-step trace estimation procedure is proposed which contains the minimum distance method and the point distribution analysis method. The assumption for this algorithm is a local navigation task, i.e. the search for the closest possible path in front of the mobile system with respect to the dimension of the system while avoiding obstacles. Both methods work with all reconstructed 3D points. The analysis of the distribution of 3D points in space leads to a possible trace for the mobile robot, resulting in values for steering angle and velocity given to the mobile system.

The method is implemented on the experimental system MOVILAR (MObile VIsion and LAser based Robot) which is based on a multi processor network. This experimental system achieves actually process cycles of approximately 1.5s and a velocity of about 10 cm/s, i.e. a slow walking speed and is briefly described in chapter 5. Experimental results of the trace estimation procedure in different real scenes are presented in chapter 6.

2 Reconstruction of 3D Points

To get 3D information from images, two different views of the same scene are required. The two views can be obtained from one camera at two different times, called axial motion stereo, or by two or more cameras at the same time. The advantage of polyocular vision systems is the reconstruction of depth information without moving the mobile system in dynamic scenes. On the other hand more hardware equipment and complex algorithms are necessary.

The classical pinhole camera model is the basis for the mapping of the 3D world into a 2D image. This mapping is described with the projective matrix P_i which contains the intrinsic and extrinsic camera parameters. The reconstruction of the 3D point M for two corresponding points I_1 and I_2 in the left and right image is possible, if the projective matrix is known for both cameras as the result of a calibration procedure [5], [6]. The index i refers to cameras 1 and 2. In figure 1, the scheme of a general stereo rig is shown.

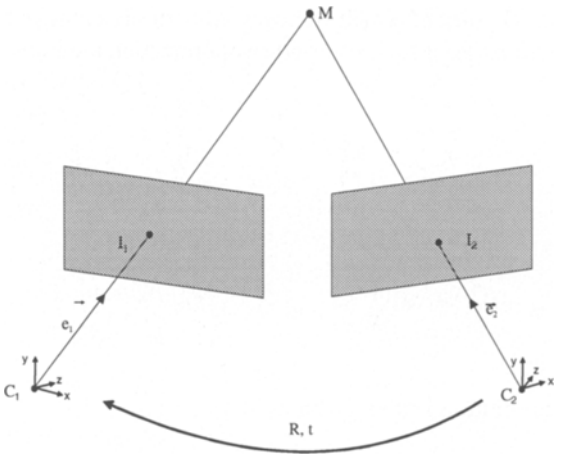

Fig. 1. Stereo geometry for the general case

The 4×3 projective matrix P_i contains the matrix A_i with the intrinsic camera parameters, the rotation matrix R_i and translation vector t_i from camera coordinate system to world coordinate system. In the figure above, the world coordinate system is defined in the optical centre of the left camera. The elements of matrix A_i are the focal length measured in horizontal and vertical pixels α_u, α_v ; the angle θ which describes a non-orthogonality of the retinal coordinate system; and the principal point (μ_0, v_0).

$$A_i = \begin{bmatrix} a_u & -a_u \cdot \cot\theta & u_0 \\ 0 & \frac{a_v}{\sin\theta} & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{1}$$

$$P_i = A_i \cdot \begin{bmatrix} R_i & t_i \end{bmatrix} \tag{2}$$

In the following equations p^i_{jk} denotes the (j,k)-element of the projective matrix of camera i . The first three elements of each row of P_i are composed to the vector $p^i_j = \left(p^i_{j1}, p^i_{j2}, p^i_{j3}\right)^T$. If the two image points $I_i = (u_i, v_i)^T$ for camera 1 and 2 are known, four equations in three unknowns $M = (x, y, z)^T$ are obtained.

$$\begin{aligned} \left(p^1_1 - u_1 \cdot p^1_3\right)^T \cdot M + p^1_{14} - u_1 \cdot p^1_{34} &= 0 \\ \left(p^1_2 - v_1 \cdot p^1_3\right)^T \cdot M + p^1_{24} - v_1 \cdot p^1_{34} &= 0 \\ \left(p^2_1 - u_2 \cdot p^2_3\right)^T \cdot M + p^2_{14} - u_2 \cdot p^2_{34} &= 0 \\ \left(p^2_2 - v_2 \cdot p^2_3\right)^T \cdot M + p^2_{24} - v_2 \cdot p^2_{34} &= 0 \end{aligned} \tag{3}$$

The optimum 3D point M which is closest to both optical rays of each image point has to be estimated by least squares or other optimization methods [1], [6], [10].

For two cameras, set

$$A \cdot a = b, \quad \text{with} \quad A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, \quad a = (x, y, z)^T \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \tag{4}$$

and

$$A_i = \begin{pmatrix} \left(p_1^i - u_i \cdot p_3^i\right)^T \\ \left(p_2^i - v_i \cdot p_3^i\right)^T \end{pmatrix} \quad \text{and} \quad b_i = \begin{pmatrix} u_i \cdot p_{34}^i - p_{14}^i \\ v_i \cdot p_{34}^i - p_{24}^i \end{pmatrix} \tag{5}$$

The least squares solution is then given by

$$a = \left(A^T \cdot A\right)^{-1} \cdot A^T \cdot b \tag{6}$$

The problem in stereo vision is to solve the correspondence problem between distinct tokens of the left and right image [6]. In our approach, a line segment stereo algorithm is used, based on a prediction, propagation and validation strategy [1]. We have choosen line segments as tokens because they describe scenes of man-made indoor environment in an adequate manner, and supply geometrical and structural features for a succesful match. The algorithm was extended by additional grey-level features like the gradient and the mean grey value at each edge. These additional features increase the amount of correct matches and the reliability of the solution [3]. It is important to say that the kind of token, e.g. points or lines is no prerequisite for the subsequent trace estimation algorithm.

The matching and reconstruction algorithm assumes a parallel stereo geometry which results in faster algorithms, e.g. there is no necessity for rectification. On the other hand, the reconstruction results are less accurate, yet a small uncertainty in relation to the dimension of the mobile system is acceptable.

Finally, the 3D points will be obtained by considering the starting and end points of each corresponding segment pair. Unfortunately, the matching result is not entirely correct, due to errors of the previous line segmentation algorithm and the occurence of many parallel straight line segments. In order to reduce the amount of incorrect matches, the result of 3D reconstruction is compared to the stereo geometry.

3 Validation by Using Stereo Geometry

The stereo geometry defines boundaries which have to be fulfilled by the reconstructed 3D points. These boundaries are based on physical limitations, but also on assumptions regarding the navigation process. In the following explanations, the parallel stereo geometry is considered. The notation in the subsequent chapters is sometimes similiar to the previous one, but is selfexplaining in the context.

3.1 The Horizontal Case

First, the horizontal x/z-plane is shown in figure 2. The cameras are placed on the x-axis, the left one at *B*/2 and the right at -*B*/2, at which *B* is the base length of the stereo system. The cameras are arranged parallel to each other and oriented towards positive z-direction. The origin of the x/z-plane is equal to the origin of the local platform coordinate system and placed at the front side of the mobile system.

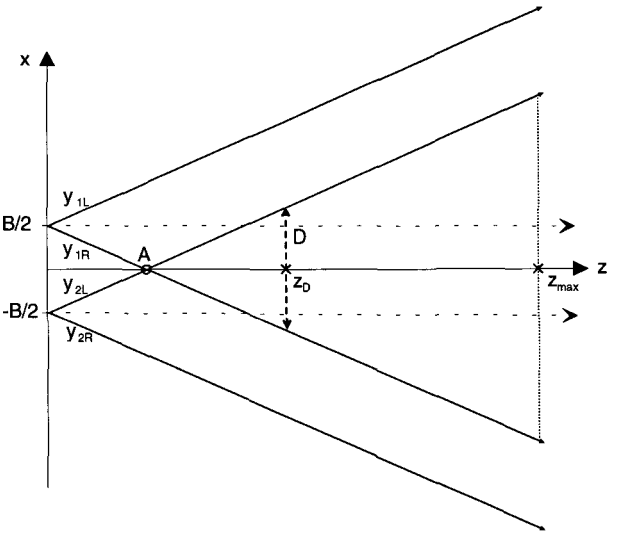

Fig. 2. The visual area of a parallel stereo geometry in top view (x/z-plane)

In x-direction, the area of view from both cameras is bounded on the left side by the left ray of the right camera and on the right side vice versa. It is expressed in the following equation, at which m_χ is the slope of left rays, depending on the focal distance f and the CCD-chip dimension Δx:

$$y_{2L}(z) = m_\chi \cdot z - \frac{B}{2}, \quad y_{1R}(z) = -m_\chi \cdot z + \frac{B}{2}, \quad \text{with } m_\chi = \frac{\Delta x}{2 \cdot f} \tag{7}$$

Observing this area of view, all reconstructed 3D points outside this area have to be excluded. In z-direction, the closest point to the vision system is *A*. If we consider the width *D* of the mobile system, a resulting minimum distance z_D is obtained.

$$z_D = \frac{D + B}{2 \cdot m_\chi} \tag{8}$$

The area in front of the mobile system up to the distance z_D is not completly observable. Consequently, 3D points lying in this security area are also excluded.

3.2 The Vertical Case

In the vertical x/y-plane, 3D points should lie above the ground plane and under a maximum height, that depends on the height of the mobile system and a security distance. This assumption is realistic for indoor environment. In natural outdoor environment obstacles below the ground plane may occur and the assumption will be no longer valid.

This validation procedure leads to a set of 3D points which lie within the 3D boundaries explained before. There is no guarantee that all reconstructed 3D points correspond to real 3D points of the scene, and the trace estimation algorithm described below has to take this fact into account.

4 Trace Estimation

The task of the mobile system is to move straight ahead or search the next possible path in the area of view and to avoid obstacles. In order to solve this problem, the trace estimation procedure works with all valid 3D points. The movement is only horizontal in the x/z-plane, therefore the x- and z-coordinates are observed. The orientation of the vehicle is in positive z-direction.

The uncertainty in 3D point reconstruction prohibits a trace estimation from single points. The main goal of the proposed methods is to deal with this uncertainty by means of a statistical analysis of the whole 3D point distribution in order to reduce the influence of false 3D points.

4.1 Minimum Distance Method

The following assumption is made:

> *If there is a possible path between two obstacles, and if edges of these obstacles are found, there are two pairs of 3D points with a distance greater than a maximum width* d_w.

The value of d_w is determined by the dimension of the mobile system. The line between these two points is called *trace segment*, and the point in the middle of the line is called *trace point* Tr_k. 3D points of the same 3D edge are excluded from the analysis in pairs, because a path through the 3D edge is impossible. Since the mobile system has to move straight ahead, the angle β between each trace segment and the x-axis should be less than $\pm 45°$. This boundary is no restriction to any given path, and will be explained at the end of the chapter.

The previous considerations are expressed in the following manner:

$$\left|P_i(x,z)-P_j(x,z)\right| > d_w \quad \text{and} \quad |\beta| < 45°$$
$$\forall i, j \in N, \quad i \neq j, \quad i \text{ and } j \text{ from different edges}, \quad N = \text{number of 3D - points} \tag{9}$$

$$Tr_k(x,z) = \left(P_i(x,z) - P_j(x,z)\right)/2 \tag{10}$$

In [2], a confidence measure for 3D line segments is proposed which depends on the distance to the robot and some other elements of the processing chain. The uncertainty in the distance depends on less resolution of the disparity by computing 3D edges far away from the vision system. This weighting is introduced in our 3D point-approach in a similiar way. We compute the distance-dependent weighted sum of all trace points, which leads to the resulting mean trace.

$$Tr_{mean}(x,z) = \sum_{k=1}^{M} Tr_k(x,z) \cdot w(d_k), \quad d_k : \text{ distance of } Tr_k(x,z)$$
$$\text{w(r)}: \text{weighting - function}$$
$$\text{M}: \quad \text{number of trace points} \tag{11}$$

The polar coordinates of the mean trace point contain the steering angle α for the mobile system. The figure below illustrates the idea.

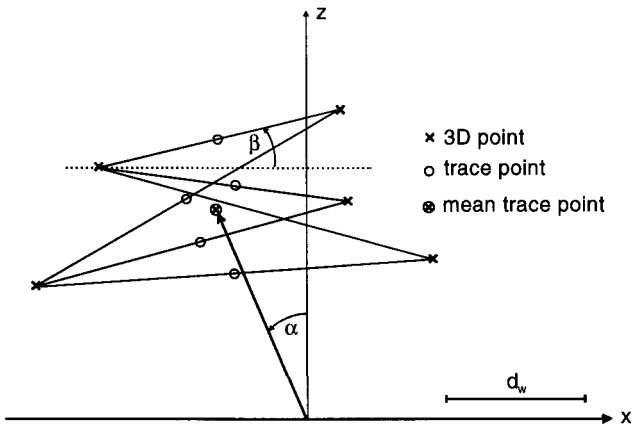

Fig. 3. The minimum distance method applied on five 3D points

This criterion works well if the obstacles are distributed in two parts which the mobile system is able to pass. In the case of aligned 3D points resulting from e.g. a wall, an additional criterion is introduced by analysing the neighbourhood of the 3D points. If a trace segment is found that fulfils the criteria in (9), an area around the trace segment is observed, marked grey-shaded in figure 4 . The width of the area is

determined by a constant d_v, describing a minimum distance of trace points in depth of x/z-space.

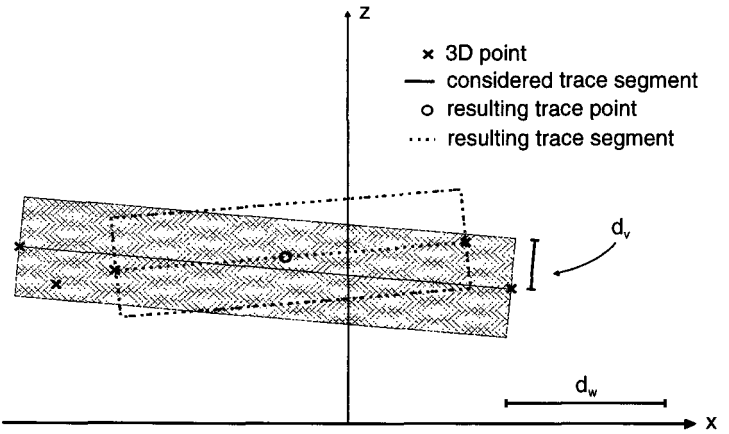

Fig. 4. The case of aligned 3D points

If one or more other 3D points are found within the area, the current observed trace segment and resulting trace point will be rejected. In the example above, one possible trace point is found at the end of the procedure. Now, we define a minimum amount of trace points required for estimating a mean trace. In the example above no trace estimation is possible. Nevertheless, the trace estimation problem will be solved by applying a second analysis, explained in the next chapter.

The boundary for the angle β between each trace segment and the x-axis is no restriction for paths with a larger angle, because it is applied locally at time t_i. At time t_{i+1}, the mobile system was moved in a new position with a new orientation, observing new 3D points with a new resulting mean trace. The behaviour is shown in the figures below.

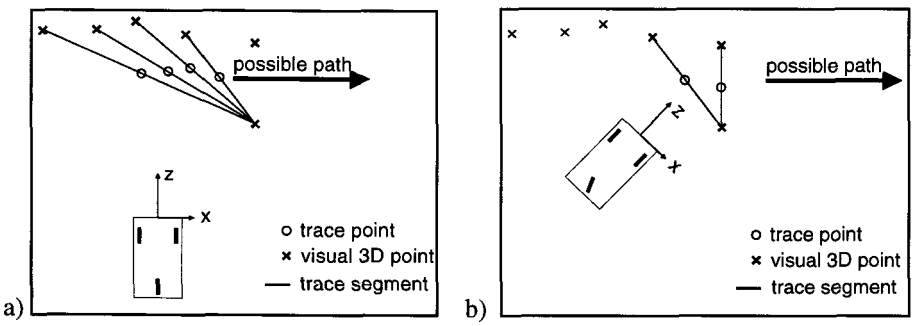

Fig. 5. a) Position and orientation at time t_i, b) Position and orientation at time t_{i+1}

4.2 Point Distribution Method

In case of rejection by the previous method, the distribution of the 3D points in the x/z-plane will be analysed. The 2D covariance matrix of the points is calculated by

$$\vec{p}_i = P_i(x,z) \tag{12}$$

$$\underline{C} = \sum_{i=1}^{N} (\vec{p}_i - \vec{m}_0)\cdot(\vec{p}_i - \vec{m}_0)', \vec{m}_0 = \text{mean of the point distribution} \tag{13}$$

The eigenvectors represent the orientation of the distribution and the roots of the eigenvalues correspond to the standard deviation in each main axis. In order to extract trace information of different point distribution ellipsoids, classifying of position and orientation is necessary.

For a reliable analysis, small and concentrated point distributions should be excluded, as no main direction is detectable. The position and orientation of the point distribution ellipsoid are determined by the centre of gravity m_0 and the angle φ between the main axis of the distribution and the z-axis. The steering angle α depends on the end-point of the main axis adding a security distance d_s perpendicular to the main axis or in horizontal direction. The different possibilities of orientation and position on the left hand side of the z-axis with the resulting steering angle are shown in figure 6. The coherence for point distributions with a centre of gravity on the right hand side of the z-axis is equal.

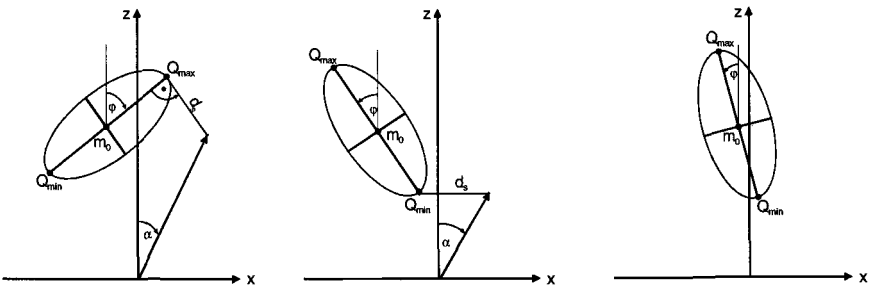

Fig. 6. Different cases of point distribution ellipsoids

We observe three different cases depending on the orientation, the centre of gravity m_0, the point Q_{min} with smaller z-coordinate and Q_{max} with larger z-coordinate related to the mobile system:

a) If the angle is positive, the aiming point is found next to Q_{max}, taking into account the security distance d_s perpendicular to the main axis.

b) If the angle is negativ, the aiming point is found next to Q_{min} , taking into account the security distance d_s as a horizontal shift.

c) If the centre of gravity m_0 and the point Q_{min} are located on different sides of the z-axis and the absolut value of the angle φ is less then a pre-set threshold, it is difficult to decide whether the obstacles are on the left or right side of the mobile system. In this case, no steering angle will be estimated, the mobile system takes the previous value and reduces its speed.

5 The Experimental System

The above mentioned methods have been investigated on our experimental mobile system MOVILAR (MObile VIsion and LAser based Robot). It is a three-wheel vehicle with rear drive. The steering motor and the drive motor are DC-motors, each of them is driven by a PWM power module (Pulse Width Modulation). A controller-card, based on a microcontroller 80C537, provides the link between the vehicle hardware and the control computer.

Fig. 7. The mobile system MOVILAR

A stereo vision head with two CCD-cameras provides a resolution of 512 x 604 pixels. Both cameras are mounted on rotating discs on a splint at the head of the vehicle. They can be shifted along this splint which enables different stereo geometries, with regard to the distance between the cameras (baselength) and the angle between the axes of the two cameras. The image processing runs on a multi-transputer system. The kernel of the system consists of two transputer frame grabbers (TFG), digitizing the images in real time and providing a digital image in the video memory. Furthermore, for algorithms with high computational expense, e.g. image preprocessing, two PowerPC-moduls with an MPC 601 processor are available. For further investigations, a laser scanner PLS-200 is installed in front of the mobile platform.

6 Experimental Results

The system moves with a mean velocity of about 10 cm/s in our lab and finds a possible path without collision. In order to illustrate the performance of the estimation procedure, a part of real scenes is presented, showing all steps of processing in the subsequent figures. The basis is a stereo segment image as input for the stereo algorithm, shown in the top left window. The left and right line segments are marked with dashed and solid lines. The homologous segments as a result of matching can be seen in the window top right, where the corresponding segments are equally numbered. On bottom left hand side, the 3D data in the x/z-plane are displayed. The reconstructed 3D points are marked with a cross, the computed trace points with circles, and the estimated trace is displayed with a straight line. The point distribution ellipsoids are also visualized. The length of both main axes is three times the standard deviation of the distribution.

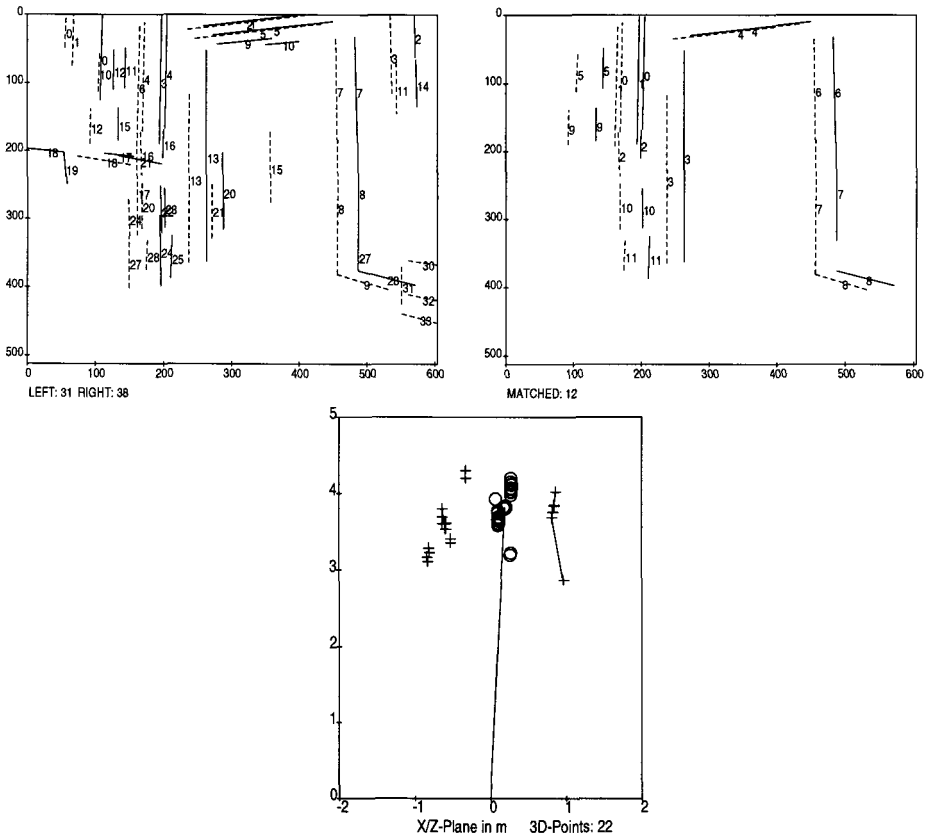

Fig. 8. The minimum distance method

In the figure above, the minimum distance method was applied, since 3D points were found on the left and right hand side, with a possible path between them.

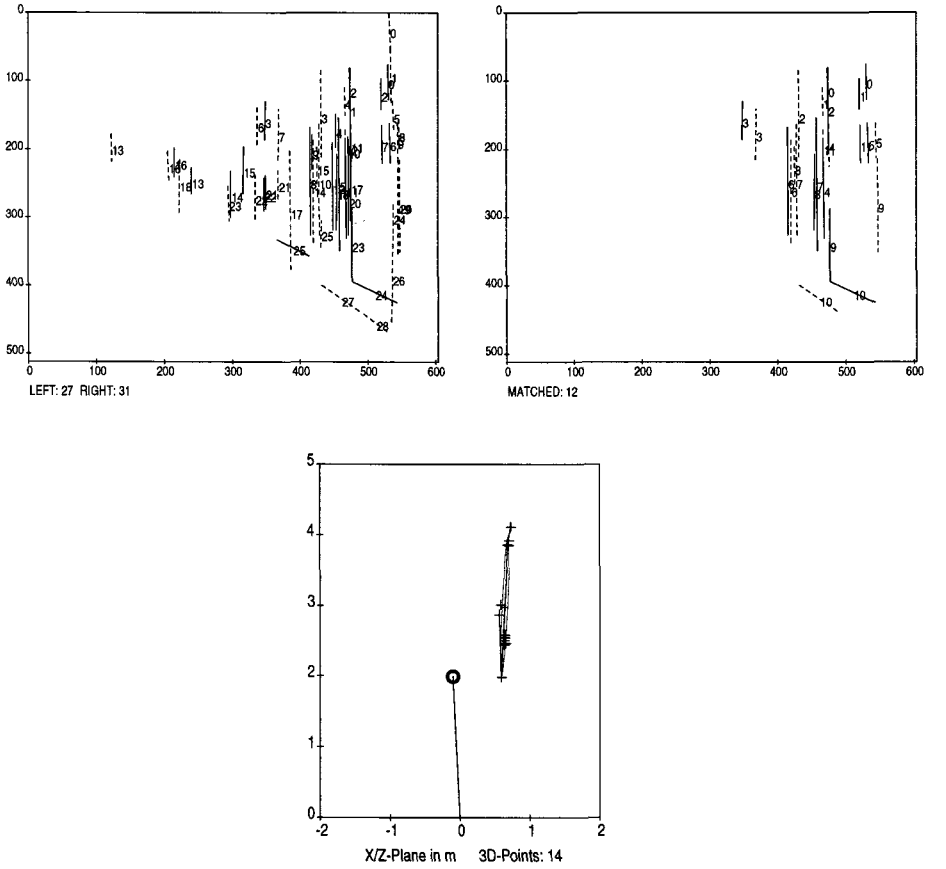

Fig. 9. An example for moving along a wall by applying the point distribution method

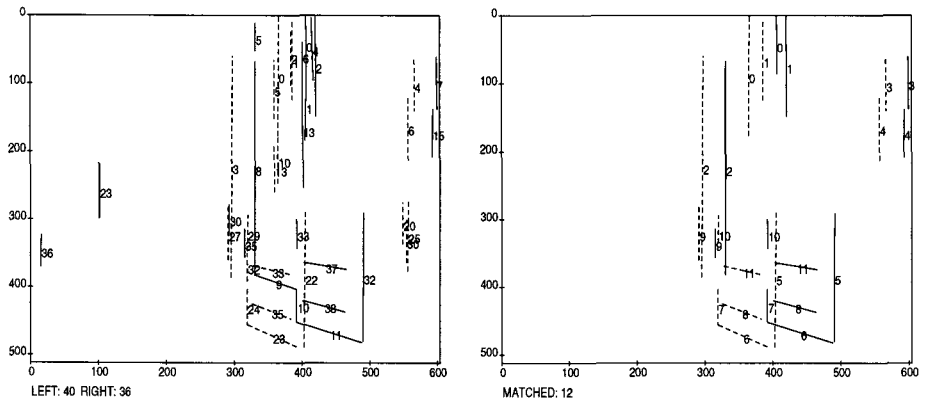

Fig. 10. Stereo segments and matching result for obstacles found on the right side

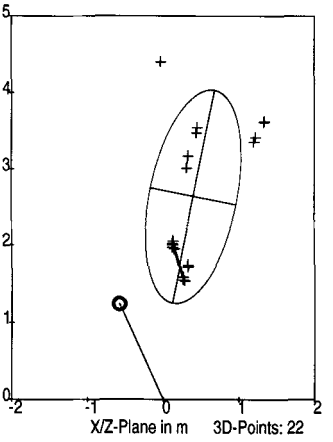

Fig. 11. Obstacles found on the right side by applying the point distribution method based on the stereo segments in figure 10

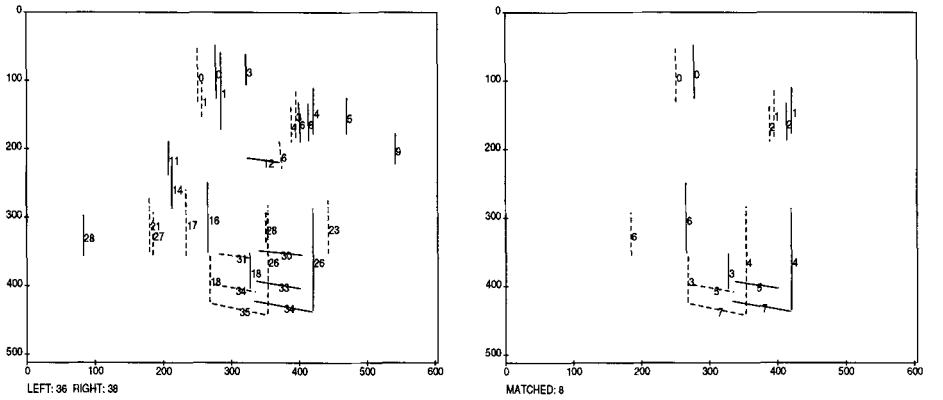

Fig. 12. Stereo segments and matching result for obstacles in front of the mobile system

In figure 9 and 11, the minimum distance method did not lead to a result, but a reliable trace could be established through the point distribution analysis. The examples correspond to the cases a) and b) of figure 6. An example for rejection by the trace estimation algorithm i.e. the case c), is given in figure 13. The angle φ is less than a pre-set threshold, and the point distribution is closed to the z- axis.

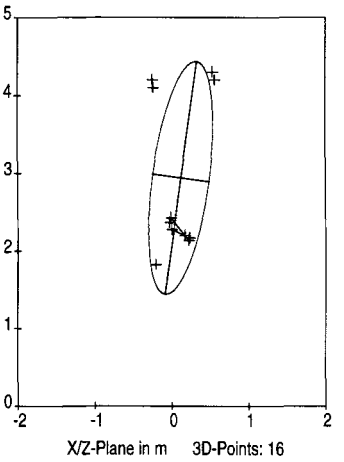

Fig. 13. Rejection by the trace estimation algorithm based on the stereo segments in figure 12

7 Conclusion

We have proposed a method for stereo vision based navigation of autonomous mobile robots. This method does not require any knowledge of the environment (like maps, landmarks or objects). The main idea is to consider every reconstructed 3D point as an obstacle. A validation procedure uses the knowledge of the stereo geometry to eliminate impossible 3D points. The remaining uncertainty of the 3D points is taken into account by two successive statistical methods. In case of point distributions in two parts of the space which the mobile system is able to pass, the minimum distance method leads to a robust trace angle estimation. This method takes into account the uncertainty related to the distance of considered 3D points.

In case of failure of this method, the point distribution analysis method is applied. The method results in an adequate estimation of the steering angle when obstacles are seen only at one side of the mobile system, e.g. in case of moving along a wall. The main goals of this approach was to deal with the uncertainty of 3D point reconstruction, to find a reliable trace, and to avoid collisions while navigating the mobile system.

Acknowledgement

The author would like to thank Prof. Dr. I. Hartmann for his contribution to this research.

References

[1] N. Ayache: „Artificial Vision for Mobile Robots", MIT Press, Cambridge, Massachusetts, 1991.

[2] M. Buffa, O. Faugeras, Z. Zhang: „A Stereovision-Based System for a Mobile Robot", Research Report No. 1895, Institut National de Recherche en Informatique et en Automatique (INRIA), April 1993, Sophia-Antipolis.

[3] M. Ekinci, B.T. Thomas: „Road Junction Recognition and Turn-Offs for Autonomous Road Vehicle Navigation", Int. Conf. on Pattern Recognition, Aug. 1996, Austria.

[4] O. D. Faugeras: „Three-Dimensional Computer Vision", The MIT Press, Cambridge, Massachusetts, London, England, 1993.

[5] O. D. Faugeras and G. Toscanini: „The Calibration Problem for Stereo", Proc. of Conf. Comput. Vision and Pattern Recognition, Florida, 1986.

[6] R. K. Lenz and R. Y. Tsai: „Calibrating a Cartesian Robot with Eye-on-Hand Configuration Independent of Eye-to-Hand Relationship", PAMI Vol.11, No.9, September 1989.

[7] L. Robert, C. Zeller, O. Faugeras: „Applications of Non-metric Vision to Some Visually Guided Robotics Tasks", Research Report No. 2584, Institut National de Recherche en Informatique et en Automatique (INRIA), June 1995, Sophia-Antipolis.

[8] O. Schreer, I. Hartmann and R. Adams: „Analysis of Grey-Level Features for Line Segment Stereo Matching", Int. Conf. on Image Analysis and Processing, Sept. 1997, Florence, Italy.

[9] P. Weckesser, F. Wallner, R. Dillmann: „Position Correction of a Mobile Robot Using Predictive Vision", Int. Conf. on Intelligent Autonomous Systems, IAS-4, 1995.

[10] Z. Zhang, G. Xu: „Epipolar Geometry in Stereo, Motion and Object Recognition", Kluwer Academic Publisher, 1996, Netherlands.