

Camera Based Motion Tracking for Data Fusion in a Landmine Detection System

Wannes van der Mark, Johan C. van den Heuvel, Eric den Breejen

TNO Physics and Electronic Laboratory

Electro-Optical Systems, Oude Waalsdorperweg 63, The Hague, The Netherlands

Phone: +31-70-374-0375, Email: {vandermark,vandenheuvel,breejen}@fel.tno.nl

Frans C.A. Groen

University of Amsterdam

Informatics Institute, Kruislaan 403, Amsterdam, The Netherlands

Phone: +31-20-525-7461, Email: groen@science.uva.nl

***Abstract** – We present a method, based on stereo vision, for estimating the position and orientation of the LOTUS platform and its sensors. The LOTUS platform was developed to demonstrate the capabilities of automated landmine detection. Motion of the platform and pose and position of its sensors have to be measured to relate different sensors observations accurately to each other. Techniques from camera calibration are used for sensor pose and position estimation. The platform motion is estimated from tracked features on the ground. A special estimator was developed to deal with problems related to estimating rotation from coplanar surface points. This estimator also uses weights in order to remove outlier points caused by tracking errors and other influences. Simulation experiments shows that the weights can safeguard the estimator against a limited amount of outliers. Experiments with real stereo images from the LOTUS platform show that the relative pose and position of sensor can be estimated with high accuracy. When combined with the ego-motion of the cameras the position of the sensors can be related to fixed points to the ground. The results show that the vision based approach provides more useful position estimates when compared to an odometry based approach.*

I. INTRODUCTION

For safety reasons, there are dangerous tasks which can be better performed by machines than by men. An example of such a task is clearance of landmines in former conflict areas. The main problem of demining has always been finding the buried mines. Currently this is done mainly by hand by the demining personal. Because of the sheer number and areas in which the landmines still remain, this is a very tedious and dangerous task.

The LOTUS project has been a cooperation between several European partners to develop technology for automated landmine detection. Result of their collaboration has been the LOTUS platform, which was used to demonstrate the capabilities of automatic landmine detection and marking for humanitarian demining operations in Bosnia.

On the platform multiple sensors are present which are able to detect properties which may belong to landmines. This approach was chosen because a practical system must be able to find a large variety of landmine types. As can be seen in Fig. 1. the Lotus consists of a all-terrain vehicle with a large metal

frame on which all the sensors have been mounted. Three types of sensors are used to detect landmines; ground penetrating radar, metal detector and cameras.

The ground penetrating radar (GPR) has been mounted near the vehicle. This radar was developed by EMRAD (United Kingdom). It uses 16 antennas to measure the density of the ground for different depth layers. The metal detector (MD) is mounted at the front of the frame. This sensor which was produced by Froestner (Germany) contains 7 detector coils. The coils measure if metal containing parts are near. High in the frame, between the MD and the GPR, downward looking infrared and multi spectral (IR) cameras are mounted. These cameras observe temperature differences in ground temperature. In order to indicate detected landmines on the ground a marking unit is used at the back of the vehicle. This is a device which can mark a part of the terrain with spray paint.

Data from the different sensors needs to be collected for the same location on the ground in order to decide if it contains a mine. The LOTUS drives over a minefield in a straight line. In this way, the MD will first pass over a location, then the IR cameras and finally the GPR.

A virtual grid aligned with the ground plane is used to combine the different sensor measurements. The size of each grid cell is 25 mm square. In [1] several techniques for landmine detection with sensor fusion are described which are based on the grid approach. Currently, odometry is used to measure the progress of the platform over the lane. Together with the fixed distances between sensors this is used to incorporate observations into the stationary grid.

Field tests have shown that the current approach is sufficient on flat terrain. However, in more difficult terrain the motion platform is more complicated than a single translation. Also changes in orientation have to be considered because of the uneven surface. Localization is further complicated by the way the MD is mounted on the platform. To ensure good sensor readings, the sensor can follow the contours of the surface. However, in unstructured terrain the MD will be displaced by



Fig. 1. The LOTUS platform and its main components.

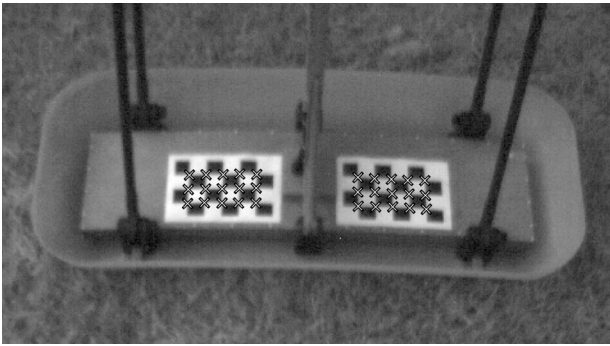


Fig. 2. A detail of a left image of the stereo camera in which the MD is visible. The crosses indicate the detected checker intersections.

obstacles on the ground. In order to ensure that the data fusion of LOTUS also works correctly in more difficult terrain an alternative for the odometry was investigated. Because of the independently moving MD, techniques were needed which could measure the orientation and position of sensors and the motion of the vehicle. To measure both, LOTUS was equipped with two cameras in a stereo configuration. They were mounted near the IR cameras and have wide angle lenses in order to view both the MD and GPR. A camera can provide information about the distance and pose of an object when its precise dimensions are known. Stereo vision can measure distance to arbitrary objects if corresponding points can be found in the stereo images. Motion of a camera can be observed by tracking image points which remain fixed in the world. This is known as the ego-motion of the camera.

A more detailed description of LOTUS and preliminary results of the vision system can be found in our earlier paper [2]. In this paper we present how these techniques have been adapted especially for the conditions under which the LOTUS platform operates. Also the accuracy of the developed methods is investigated.

II. SENSOR POSE AND POSITION

Known geometry is used to estimate the orientation and position of sensors. Sheets with checkered patterns were added to the surface of the MD and the GPR. The distance between intersections of the checkers was measured beforehand.

Intersections of checkers are easily extractable via image processing. We used the junction detector described by ter Haar Romeny [4], which is based on image derivatives, to find the intersection points. Fig. 2 shows an example image in which junctions were extracted with this detector.

Points on the detected junctions in the image and points on the real pattern both lay in their own plane. A transformation between points in two planes can be described by a homography matrix H . This is a 3 by 3 matrix which relates the homogenous points $\tilde{y} = (y_1 \ y_2 \ 1)^T$ on the pattern to the image points $\tilde{x} = (x_1 \ x_2 \ 1)^T$ up to scale factor κ :

$$\kappa \tilde{x} = H \tilde{y} \quad (1)$$

In camera calibration the homography matrix is used to estimate intrinsic and extrinsic parameters. Intrinsic parameters are properties associated with the camera projection, such as focal length, principal point and skew. All these intrinsic parameters are contained in the 3 by 3 projection matrix P . Extrinsic parameters are the orientation and position difference between the camera reference frame and a reference frame in the world. This difference is defined by a Euclidian transformation in 3D space, consisting of a rotation matrix R and a translation vector T . A frame of reference is defined by the points on the pattern. With the transformation and the projection matrix they are projected onto the camera image. Their product is equal to the homography between the camera image and the pattern plane:

$$P[R \ T] = H \quad (2)$$

In a camera calibration procedure, the projection matrix is first estimated from H . Calibration software, such as the toolbox by [5], can be used to estimate P of a camera. In the paper by Zhang [6] a complete camera calibration method is explained. It also shows how to extract the rotation matrix and translation vector from the homography when P is known. If h_1 , h_2 and h_3 are the column vectors of H then:

$$\begin{aligned} \hat{r}_1 &= \kappa P^{-1} h_1 \\ \hat{r}_2 &= \kappa P^{-1} h_2 \\ \hat{r}_3 &= h_1 \times h_2 \\ T &= \kappa P^{-1} h_3 \end{aligned} \quad \text{with} \quad \kappa = \frac{1}{\|P^{-1} h_1\|} = \frac{1}{\|P^{-1} h_2\|} \quad (3)$$

Because of image noise, the matrix A formed by the column vectors \hat{r}_1 , \hat{r}_2 and \hat{r}_3 is not always a good rotation matrix. Single value decomposition can be applied to approximate the true rotation matrix better:

$$\text{SVD}(A) = USV^T \quad R = UV^T \quad (4)$$

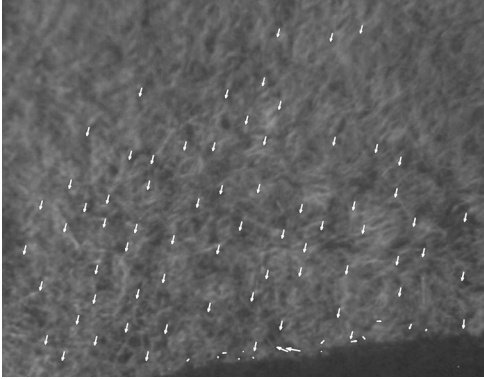


Fig. 3. Features tracked on the ground.

III. EGO-MOTION ESTIMATION

As seen in the previous section, the orientation and translation difference between a calibrated camera and a sensor can be estimated if a pattern with known geometry is attached. With the measured transformations, all sensor observations can be put in one frame of reference; that of the camera. However, the sensors do not survey the same pieces of terrain simultaneously. One by one they will pass over the same area because of the vehicle motion. A second transformation is needed to move the sensor observations from the camera to the reference frame of the ground. This transformation is equal to concatenating successive ego-motions of the camera.

A camera which moves freely through space changes its position and orientation. Two successive images of this camera will be separated by a rotation and a translation in the 3D space. For the cameras on LOTUS the distance to and the aspect angle with the surface can change. In order to estimate the ego-motion, known geometry cannot be exploited because of the random surface structure.

We use stereo vision to estimate the ego-motion of the cameras. It offers the possibility of estimating the distance of image points. In this section, a method is explained for estimating the rotation and translation from succeeding stereo images.

A. Feature Tracking

The method used for ego-motion estimation from the cameras is feature based. Feature points are extracted with the Harris [7] corner detector. The advantage of using corner like features is that the uncertainty of their image position is isotropic [9]. Locations of detected features and corresponding points in other images are therefore unbiased.

For two successive stereo image pairs, feature correspondences are searched between the left images of the first and second stereo pair. The correspondences are searched for feature locations in the first left image, which have been extracted with the corner detector or remain after tracking from the previous stereo pair. We use the tracker algorithm of the Intel OpenCV

[8] computer vision programming library in order to find these correspondences.

After the features have been tracked, they are matched in stereo. Stereo matches for the first left image are searched in the first right image. Normalized cross correlation (NCC) is used to compare possible corresponding points in the right image within a fixed disparity interval. To prevent bad matches, stereo correspondences are rejected if the NCC score is too low. Ambiguous matches are prevented by rejecting correspondences if there is a number of best points with similar NCC scores. The same technique is applied to match features in the second stereo pair. Rejected stereo matches from both stereo pairs are also used to remove tracks. In Fig.3 an example is shown of stereo tracked features on the ground.

B. Motion Estimation with Coplanar Surface Points

The points x_i in the left image are now matched with points x'_i in the right image. Both x_i and x'_i are projections of a single point on the ground surface. Using stereo reconstruction, the vector X_i can be computed which indicates this point. In the left image of the second stereo pair the corresponding features, obtained by tracking, are indicated by y_i . They also have been matched in stereo with the features y'_i in the right image. Stereo reconstruction for both pairs provides two sets of space vectors: X_i and Y_i .

Because of the rotation and translation differences between the two camera positions the vector sets are separated by a Euclidian transformation:

$$X_i = RY_i + T \quad (5)$$

In order to find the ego-motion of the stereo cameras the rotation R and translation T have to be estimated. On first sight, this looks quite straightforward. Rotation can be made independent from translation by centering the vectors in X_i and Y_i around their centroids \bar{X} and \bar{Y} :

$$\hat{X}_i = X_i - \bar{X} \quad \hat{Y}_i = Y_i - \bar{Y} \quad (6)$$

We seek the best approximation of rotation matrix R , which rotates \hat{Y} into \hat{X} . This can be done via single value decomposition of Eq. 4 with $A = \hat{X}\hat{Y}^T$.

On the LOTUS platform the stereo cameras track points which lie on the ground plane. In principal, the reconstructed points will be coplanar. A 3D rotation estimation in this situation can be troublesome. Because of the coplanarity of points, there are multiple solutions for R which turn \hat{X}_i into \hat{Y}_i . These other solutions are mirrored versions of real rotation in the plane in which the points lie.

In order to prevent the occurrence of mirrored rotations, a special rotation estimator is used which exploits the fact that the points will be mainly coplanar. The first step of this estimator is a principal component analysis of the points in \hat{X}_i . A local reference frame for a set of points is defined by the eigenvectors of their covariance. The two eigenvectors associated with

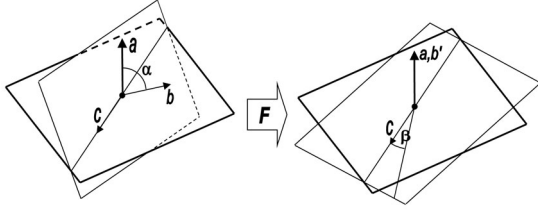


Fig. 4. Estimating rotation from coplanar points.

the largest eigenvalues will be aligned with the plane in which the points lie. These two vectors are taken as the first columns of matrix D , the remaining orthogonal eigenvector is taken as the third column. By rotating the vectors of \hat{X}_i with the inverse of D , the first two dimensions of the reference frame will be aligned with the plane and the third will be perpendicular to it. Also the vectors of \hat{Y}_i are rotated by D^{-1} :

$$\hat{X}'_i = D^{-1}\hat{X}_i \quad \hat{Y}'_i = D^{-1}\hat{Y}_i \quad (7)$$

Because of the transformation to the reference frame related to \hat{X} the normal of the plane is equal to vector $a = (0 \ 0 \ 1)^T$. The normal of the plane through the points in \hat{Y}' can be estimated with a single value decomposition of these values. The right singular vector belonging to the smallest singular value is the normal b to the plane. A drawing of the resulting situation can be seen on the left in Fig. 4. In order to avoid ambiguity, b is forced to point in the same direction as a .

The cross product of a and b is a vector c which is orthogonal to both vectors. Together with the angle α of the dot product of a and b this describes a rotation. We use a quaternion to describe the rotation around the axis c with angle α . With the matrix representation F of the quaternion, the points in \hat{Y}' can be rotated into the plane of \hat{X}' .

Now, all points lie in a single plane. This is shown on the right in Fig 4. Only a rotation around the joint normal of the planes with angle β separates the points. Single value decomposition is also used to estimate this rotation, defined by the 2 by 2 matrix G_{2D} . The rotation R can now be composed as:

$$R = (GF)^{-1}D^{-1} \quad \text{with} \quad G = \begin{pmatrix} G_{2D} & 0 \\ 0 & 1 \end{pmatrix} \quad (8)$$

C. Refining the Motion Estimate

With the previously described method an estimate can be made of R and T from the stereo reconstructed vector sets X_i and Y_i . Unfortunately this estimate will not be optimal. This is caused by the fact that the motion estimate is based on reconstructed vectors from noisy image feature. As indicated earlier, the distribution of the error in feature location can be considered isotropic. However, it is well known that errors of a stereo reconstruction based on noisy image features is non-isotropic. The uncertainty of the distance direction will be larger than

in other directions [10]. This fact should be considered in the ego-motion estimate.

There is a second problem with the presented least squares approach. Outliers can be present among the feature tracks. Tracking errors can cause outliers because they generate features with a motion which does not correspond to the ego-motion. On the LOTUS, shadows in the stereo images can also cause outliers. These shadows are cast by the large sensor frame onto the ground. If features are tracked on the shadow edges they remain stationary in the image. This does also not correspond to the ego-motion of the cameras. In Fig.3 a shadow is visible, it caused some outliers among the tracked features. Least squares considers each data point as equal. When outliers are present, this leads to suboptimal estimates.

In order to deal with the larger error in distance and outliers we have adapted the camera calibration algorithm of Lasenby [11] et. al. Originally, this algorithm was intended for estimating the rotation and translation differences between a number of cameras looking at the same scene. It introduced a way of improving the distance estimate of the used feature points in order to get a better motion estimate. The key idea is using an extra scalar λ_i which “extends” the feature vector x_i to the space vector X_i :

$$\lambda_i x_i = X_i \quad (9)$$

This can be seen as a simplification of the projection matrix P . The points y_i in the second stereo pair are extended by their own scalars μ_i to Y_i .

After an initial guess is provided for R and T , by the least squares method, iteration is used to improve the estimates for R , T , λ_i and μ_i . The error of the estimate indicated by the norm of the distance between the reconstruction of x_i and y_i under the transformation of R and t is:

$$E_i = \|\lambda_i x_i - \mu_i R y_i + T\| \quad (10)$$

We expanded the original algorithm with weights w_i for each point. The weights are used to exclude outliers from the estimation process. In an iteration of this algorithm the following steps are undertaken in order to update the values and the weights:

C.1 Translation update. A weighted average of the difference between the rotated reconstructed vectors $\mu_i R y_i$ and X_i is computed:

$$T = \frac{1}{w_{sum}} \sum_{i=0}^n w_i (X_i - \mu_i R y_i) \quad \text{with} \quad w_{sum} = \sum_{i=0}^n w_i \quad (11)$$

C.2 Rotation update. The rotation update used, also uses the rotation estimator for planar points explained earlier. Only the way in which the centroids are computed and the vectors are centered around them is different. Just like the translation the weights are now included in a weighted average for the

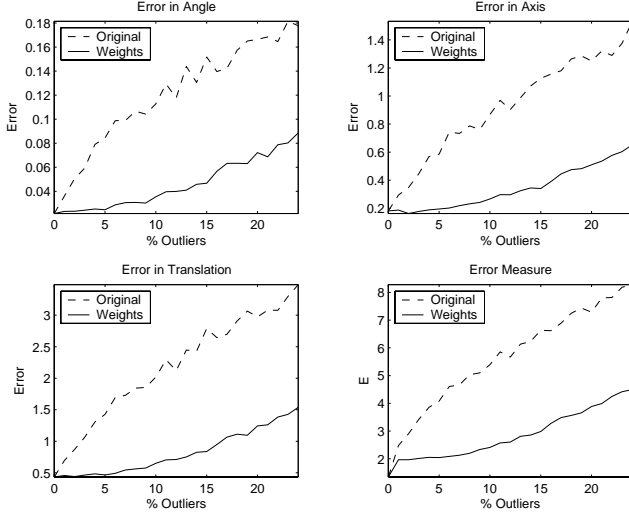


Fig. 5. Accuracy of the estimator.

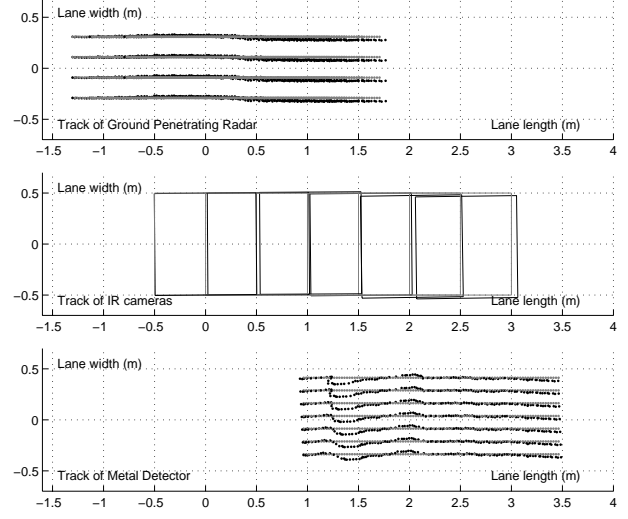


Fig. 6. Results for motion estimation on the LOTUS.

centroid estimate:

$$\bar{X} = \frac{1}{w_{sum}} \sum_{i=0}^n w_i X_i \quad \bar{Y} = \frac{1}{w_{sum}} \sum_{i=0}^n w_i Y_i \quad (12)$$

In order to incorporate weights into the centered vectors, they are normalized to equal length:

$$\hat{X}_i = w_i \frac{X_i - \bar{X}}{\|X_i - \bar{X}\|} \quad \hat{Y}_i = w_i \frac{Y_i - \bar{Y}}{\|Y_i - \bar{Y}\|} \quad (13)$$

C.3 Scalar update. The scalars λ_i and μ_i are updated using x_i , y_i and the values of X_i , Y_i :

$$\lambda_i = \frac{X_i \cdot x_i}{x_i \cdot x_i} \quad \mu_i = \frac{Y_i \cdot y_i}{y_i \cdot y_i} \quad (14)$$

C.4 Vector update. The vectors X_i and Y_i are updated by the mean of reconstructed vectors from both positions:

$$X_i = \frac{1}{2} (\lambda_i x_i + \mu_i R y_i + T) \quad (15)$$

$$Y_i = \frac{1}{2} (R^{-1} (\lambda_i x_i - T) + \mu_i x_i)$$

C.5 Weights computation. After the update of the values involved in the estimation process, the weights are recomputed. Their values are based on the error distance E_i of Eq. 10. The variance $\sigma^2 = var(E)$ gives an indication how strong the values of E_i differ. It is assumed that if the variance is relatively high, outliers are present and that these outliers form a minority among the inliers. The following formula is therefore used to assign the weights values of the features based on their E_i values and the current variance:

$$w_i = \begin{cases} 1 & E_i < 2\sigma \\ 1 - \frac{E_i - 2\sigma}{\sigma} & 2\sigma \leq E_i < 3\sigma \\ 0 & E_i \geq 3\sigma \end{cases} \quad (16)$$

IV. RESULTS

A. Accuracy of the ego-motion estimator

The robustness of the new estimator was tested using simulation. A cloud of random points was generated in a 3D space. This cloud was given a planer shape to simulate the surface observed by the stereo cameras. A rotation with random angles varying between -10° and 10° and a fixed translation was used to move the points to a second location. In order to simulate outliers, a random vector was added to a certain percentage of the points of the first set. Both resulting clouds of points were projected onto two virtual stereo cameras. Normally distributed noise was added to the image locations.

In the simulation, the original estimator without weights was compared with the new estimator. Fig. 5 shows the averaged results for 500 repetitions of the simulated estimation. The first three plots show the error in motion estimation for the percentage of outliers. For the accuracy of rotation angle and axis estimation and the translation the weighted estimator performed better. In the fourth plot the error measure E is shown against the outliers, it is obvious that the weighted estimator is less burdened by the outliers.

B. Motion estimation with real data

The techniques described were used on a real sequence of 100 stereo images recorded with the LOTUS platform. The position of the MD and GPR was estimated from the images using the known geometry of the attached patterns. For every two succeeding stereo images the ego-motion was estimated. The current position of the stereo camera was determined by concatenating the ego-motions from the starting position. The relative pose and position of sensors was transformed to the world reference frame by combining them with the current po-

sition of the stereo camera.

Fig. 6 shows the results for the three sensors on the LOTUS. The top figure indicates the position estimates for the GPR. The middle figure shows the estimates for the IR cameras, the bottom figure the estimates for the MD. Because the MD and GPR consist of several detector coils and antennas, their individual positions have been drawn. For the IR cameras the field of view is shown. Gray indicates the estimates for sensor positions obtained using the odometry on LOTUS. Black is used to indicate the positions estimated with the new vision based method.

The GPR is fixed on the metal frame. Pose and position estimates for the attached pattern should remain constant. This can be used to check the accuracy of the pose and position estimation with known geometry. For the example sequence, a unit vector was transformed with the estimated rotation and translation between the GPR surface and the stereo cameras. We computed the standard deviation for each of three values of the resulting vectors. The result was (0.1328, 0.2986, 0.2682) mm, which shows that the relative sensor position can be determined with high accuracy.

During the example sequence the MD strikes some obstructions on the ground. This is visible in the track of the MD. In contrast to the track of the GPR and IR sensor the MD clearly moves to the left and right when it hits an obstacle on the ground. The track also shows that some areas are not observed by the MD. This happens when the MD swings back from the top of an obstacle. The interruptions of the track, for example at 1.8 m, are the result of this. Only the vision based approach can observe the occurrence of such situations. The tracks of the GPR and IR show the motion of the platform. When compared with the estimates of the odometry it can be seen that the vehicle does not perfectly drive in a straight line. The difference in final position estimate by the odometry and vision is 60 mm along the length and 31 mm along the width of the lane on the minefield. These numbers are considerable, given the size of the grid cells used in the sensor data-fusion.

V. CONCLUSION

On the LOTUS platform for humanitarian demining, different sensors are used to detect landmine properties. For accurate and reliable detection it is necessary to combine the different sensors readings correctly. For this purpose, the location of the sensors on the ground needs to be measured constantly. Because some of the sensors can move independently, position and orientation of both the platform and sensors has to be estimated. In this paper a vision based approach is presented which uses image sequences of a single stereo camera.

The stereo camera views both the sensors and the ground simultaneously. By exploiting known geometry added to the sensors, their position and orientation can be estimated. Experiments with real sequences show that the estimate is accurate.

Motion of the platform is estimated from the ego-motion of the cameras. The method is feature based; corner like points

are tracked on the ground surface. Because no known geometry is available on the ground, stereo reconstruction is used to estimate the spatial position of the surface points.

To obtain the rotation motion, a special estimator is used which exploits coplanarity of surface points. This ensures stability of the rotation estimate in flat terrain.

After an initial guess of the three dimensional motion, the estimate is improved via update steps in an iteration. This iteration is also used to improve the estimates for the distance of stereo reconstructed points. With the addition of weights the influence of outliers can be reduced. Simulation shows that the use of weights can make the estimator robust against a limited percentage of outliers.

Motion estimation with real stereo data shows that the position of the vehicle and its sensors can be estimated for more degrees of freedom. This is clearly an advantage in contrast to the odometry method which only measures one dimensional translation.

ACKNOWLEDGMENTS

This project is partly funded by the European Commission as ESPRIT project LOTUS, number 29812. The authors would like to thank the people of EMRAD and Foerster for their assistance and cooperation during the data capture of the stereo image sets used for this research.

REFERENCES

- [1] F. Cremer, K. Schutte, J.G.M. Schavemaker, E. den Breejen, "A comparison of decision level sensor-fusion methods for anti-personnel landmine detection", *Information Fusion* 2, Elsevier, pp. 187, 2001.
- [2] W. van der Mark, J.C. van den Heuvel, E. den Breejen, F.C.A. Groen, "Camera-based platform and sensor motion tracking for data fusion in a landmine detection system", *Unmanned Ground Vehicle Technology V*, Proceedings of SPIE, Vol. #5083, 22-23 April 2003.
- [3] H. Chung, L. Ojeda, J. Born, "Accurate Mobile Robot Dead-Reckoning with a Precision-Calibrated Fiber-Optic Gyroscope", *IEEE Transactions on Robotics and Automation*, Vol. 17, No. 1, February 2001.
- [4] B.M. ter Haar Romeny, *Front-end vision and Multiscale Image Analysis: Introduction to Scale-Space Theory*, Chapter 10, Kluwer Academic Publishers, Dordrecht, the Netherlands, 2002.
- [5] J-Y Bouguet, *Camera Calibration Toolbox for Matlab*, 2000, http://www.vision.caltech.edu/bouguetj/calib_doc/
- [6] Z. Zhang, "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations", *Proceedings International Conference on Computer Vision*, pp. 666-673, Corfu, Greece, 1999.
- [7] C. Harris, M. Stephens, "A combined corner and edge detector", *Proceedings 4'th Alvey Vision Conference*, pp. 147-151, 1988.
- [8] G.R. Bradski, V. Pisarevsky, "Intel's Computer Vision Library: applications in calibration, stereo segmentation, tracking, gesture, face and object recognition", *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 796-797, 2000.
- [9] Y. Kanazawa, K. Kanatani, "Do We Really Have to Consider Covariance Matrices for Image Feature Points?", *Electronics and Communications in Japan*, Part 3, Vol. 86, No. 1, pp. 1-10, 2003.
- [10] N. Molton, *Computer Vision as an Aid for the Visually Impaired*, PhD Thesis, Department of Engineering Science, University of Oxford, 1998.
- [11] J. Lasenby, W.J. Fitzgerald, C.J.L. Doran and A.N. Lasenby, "New Geometric Methods for Computer Vision: an application to structure and motion estimation", *Int. J. Comp. Vision*, 36(3), pp. 191-213, 1998.