

# Camera-based platform and sensor motion tracking for data fusion in a landmine detection system

Wannes van der Mark<sup>a,b</sup> Johan C. van den Heuvel<sup>a</sup> Eric den Breejen<sup>a</sup> Frans C.A. Groen<sup>b</sup>

<sup>a</sup>TNO Physics and Electronics Laboratory, Oude Waalsdorperweg 63, The Hague

<sup>b</sup>University of Amsterdam, Kruislaan 403, Amsterdam,  
The Netherlands

## ABSTRACT

Vehicles that serve in the role as landmine detection robots could be an important tool for demining former conflict areas. On the LOTUS platform for humanitarian demining, different sensors are used to detect a wide range of landmine types. Reliable and accurate detection depends on correctly combining the observations from the different sensors on the moving platform. Currently a method based on odometry is used to merge the readings from the sensors. In this paper a vision based approach is presented which can estimate the relative sensor pose and position together with the vehicle motion. To estimate the relative position and orientation of sensors, techniques from camera calibration are used. The platform motion is estimated from tracked features on the ground. A new approach is presented which can reduce the influence of tracking errors or other outliers on the accuracy of the ego-motion estimate. Overall, the new vision based approach for sensor localization leads to better estimates than the current odometry based method.

**Keywords:** Motion Estimation, Landmine Detection, Stereo Vision

## 1. INTRODUCTION

Anti-personnel mines form a dangerous legacy in former conflict areas. Several millions of landmines still remain in former conflict areas such as Cambodia, Angola and Bosnia. Cultivation of farmland, transport and economic recovery is obstructed due to the presence of the landmines. The greatest challenge of landmine clearance is finding the buried mines. This is a large problem because often no accurate mine maps were kept during the conflict. Furthermore, landmines can move through terrain by flooding and landslides. Currently, locating landmines is mainly done by hand. Humanitarian demining personnel use a stick to probe every 2.5 square cm in a suspect area. Using this technique a great amount of time and manpower is required to clear a relatively small area. An other drawback is the high risk to the demining personnel.

In the LOTUS project several European partners have developed technology for automated landmine detection in humanitarian demining operations. The developed technology has resulted in the LOTUS platform which was used to demonstrate capabilities of autonomous landmine detection and marking in Bosnia. The potential advantages are clear, automated landmine detection reduces the risks and speeds up the landmine clearance process.

---

Further author information: (Send correspondence to W. van der Mark)

W. van der Mark, J.C. van den Heuvel, E. den Breejen:

E-mail: {vandermark, vandenheuvel, breejen}@fel.tno.nl, Telephone: +31 70 374 0375,

Address: TNO Physics and Electronics Laboratory, Electro-Optical Systems,

Oude Waalsdorperweg 63, P.O. Box 96864, 2509 JG The Hague, The Netherlands.

F.C.A. Groen: E-mail: groen@science.uva.nl, Telephone: +31 20 525 7461, Address:

Informatics Institute, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands.



**Figure 1.** Picture of the LOTUS platform. A Land Rover vehicle has been equipped with a large aluminum frame, all the sensors are attached to this frame.

## 2. THE LOTUS LANDMINE DETECTION SYSTEM

A functional automated system should be able to detect a wide variety of landmines with a high degree of reliability. To detect properties of the different landmine types multiple sensors are used on the platform. The LOTUS system is based on a modified Land Rover vehicle. This vehicle is equipped with a large metal frame on which the different sensors are mounted. A picture of the LOTUS platform is shown in figure 1. The following sensors are used on the LOTUS to detect landmine properties:

**Metal detector** This is a high sensitivity metal detector (MD) by Foerster (Germany), which is mounted at the front of the frame. The device is embedded with an array of 7 detector coils which measure metal content for a rectangular area beneath the sensor.

**Ground penetrating radar** The ground penetrating radar (GPR) was developed by EMRAD (United Kingdom). The radar provides data describing the density of the underlying ground for different depths. It uses 16 antenna's in a 4 by 4 arrangement.

**Infrared and a multi spectral cameras** These cameras are provided by TNO (The Netherlands), they are mounted high in the frame to look down towards the ground. The field of view covers the area between the MD and the GPR. Differences in ground surface temperature are observed by these cameras.

During operation the platform moves in a straight line over the minefield. A terrain area will first be observed by the MD sensor. Then, it is viewed by the infrared (IR) and multi spectral cameras. Finally, the area will be observed by the GPR. At the back of the vehicle there is a marking unit. This unit consists of several spray paint cans which can be triggered to mark a specific location on the ground to indicate detected landmines.

The configuration of the sensors on the LOTUS is a necessity. Because of the high sensitivity of the MD, it has to be mounted far away from any metal containing parts. The construction between the MD and the metal frame is completely made from syntectic materials. For reliable operation the MD and the GPR must be close to the ground. In order to give the cameras a clear view, the MD and GPR were separated in the frame by some distance.

An odometer, a small wheel at the back vehicle, is used on the LOTUS platform to measure the travelled distance in the minefield. For each 2.5 cm of displacement, the wheel sends out a TTL type trigger pulse. Sensors which have a high sample rate, such as the MD and GPR, use this signal to start an observation. The infrared and multi spectral cameras

have a lower sample rate. After image capture on the same trigger pulse they skip a number of pulses, before taking new images.

Fusion of data obtained from the sensors is used to decide which locations contain landmines. Sensor-fusion on the LOTUS uses a virtual grid, which is related to the ground, to combine measurements. The size of square grid cells is equal to the 2.5 cm distance travelled between each trigger pulse. As input, each sensor has to deliver detection confidence levels for the grid squares. In the paper by Cremer<sup>1</sup> et al. several methods are applied and compared to determine which of the grid cells may contain (parts of) mines given inputs from different sensors.

At the moment of a synchronization pulse the different sensors do not survey the same grid cells. Currently, fixed offsets are used to define the relative positions of the sensors. Given the offset, the data of the camera systems can be incorporated into the grid by distributing the confidence levels at each pixel over grid cells. Dispensing the measurements of the MD and GPR into the grid is more difficult because these sensors consist of several detectors or antennas. Beside the offsets between the sensors also the spatial arrangement of elements in the sensors is used.

A problem of this approach is that it only considers the one dimensional translation of the platform. Many of the minefields are in rough and unstructured terrain. In this situation the vehicle will also have roll, pitch and yaw motions. Another problem is caused by the odometry. Through wheel slippage, the travelled distance between the synchronization pulses does sometimes not equal 2.5 cm. The problem is complicated by the fact that the sensors do not have a fixed position in the frame. A flexible trapezium construction allows the MD to follow the contours of the surface. This minimizes the distance between the sensor and the surface, which ensures that optimal sensor readings are obtained. However, obstructions on the ground will force a change in position and orientation of the sensor.

The above influences cause motions of the sensors which do not comply with the one dimensional translation measured by the odometry. Accuracy of sensor alignment with the data fusion grid will be reduced. This increases the number of grid cells which may belong to a value from a sensor. In such a situation, the localization reliability can only be ensured by enlarging the suspect area around the detection. The larger area still has to be searched by hand by the mine clearance personnel.

Recent field tests in Bosnia have shown that the current approach is sufficient on flat terrain. If the vehicle and sensor rotation and translation motions could also be measured, the accuracy of the sensor data alignment would be improved. This would enable deployment of the system in much more difficult terrain conditions.

### 3. VISUAL MOTION ESTIMATION

The current approach for merging sensor data based on only odometry readings and predefined offsets is not sufficiently accurate in rough terrain. To improve accuracy two types of motion have to be estimated on the LOTUS platform:

#### 1. **Sensor position and orientation**

Because of the independent motion of the MD, its position and orientation changes must be monitored during operation.

#### 2. **Displacement of the platform**

To incorporate measurements correctly, requires the alignment of a sensor with the virtual grid used for data-fusion. The grid is related to the ground over which the vehicle and the attached sensors move. Therefore, in order to align the sensors correctly the vehicle motion must also be measured.

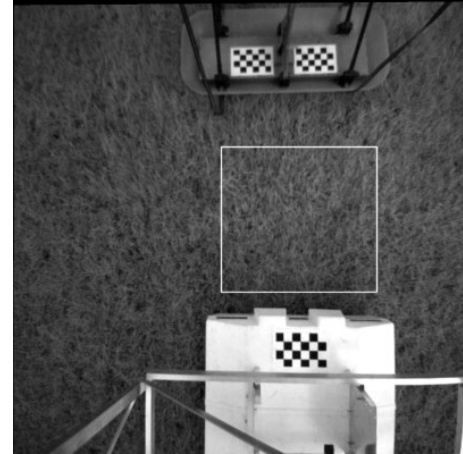
A possible solution to the first problem is adding sensors for measuring the stance of the MD. However any additions to the MD can not contain metals which interfere with the detection capabilities. Also changes in position must be measured. Active sensors such as laser range finders are often employed to measure distances to objects. They have however a limited field of view and ambiguity in distinguishing between the MD and objects on the ground.

Solution to the second problem requires measuring the vehicle motion. Using odometry complemented with a gyroscope<sup>2</sup> or an internal navigation system is a possible solution to improve vehicle localization. However, gyroscopes and INS systems are expensive.

To avoid cost and complexity by adding multiple sensors for measuring motion of the MD and the vehicle, a camera-based approach was investigated. Stereo cameras were added to the LOTUS platform near the infrared and multi spectral



**Figure 2.** The picture shows the stereo cameras used for motion estimation. They were mounted near the other cameras which are used for gathering infrared and multi spectral data.



**Figure 3.** An example image from the left stereo camera shows the MD and GPR sensor below. The white square indicates the field of view of the infrared and multi spectral camera.

cameras. Just like the other cameras, they were mounted facing down to view the passing terrain. Wide angle lenses were used to view both MD and GPR in one image. Image capture by the cameras is synchronized with data capture of the other sensors via the trigger signal from the odometry. The picture of figure 2 shows all the cameras on the LOTUS system. The two large cameras are the multi spectral and infrared cameras. On the right are the two stereo cameras used for motion estimation. In figure 3 an example left image of the stereo cameras is shown.

#### 4. ESTIMATING THE POSE AND POSITION OF SENSORS

Figure 3 shows an example image from the left stereo camera; both the MD and GPR are visible. We attached checkerboard patterns on the surface of these sensors. The intersections of these patterns are easily extractable via image processing. An example of a detector for checker junctions, based on image derivatives, can be found in ter Haar Romeny.<sup>3</sup>

The detected junctions in the image and the geometry of the attached pattern can be used to reconstruct the pose and position of the sensor. Figure 4 shows a drawing in which a checkerboard pattern is projected onto a camera image. The camera has a three dimensional frame of reference indicated by  $x$ ,  $y$  and  $z$ . An intersection point on the pattern is indicated by vector  $A$ . This point is projected onto the camera image point indicated by vector  $a$ . Relating the space vector  $A$  to the homogenous image vector  $\tilde{a}$  can be done via a projection matrix  $P$  :

$$\lambda \tilde{a} = PA \quad \text{with} \quad P = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}, \quad \tilde{a} = \begin{pmatrix} a_x \\ a_y \\ 1 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} \quad (1)$$

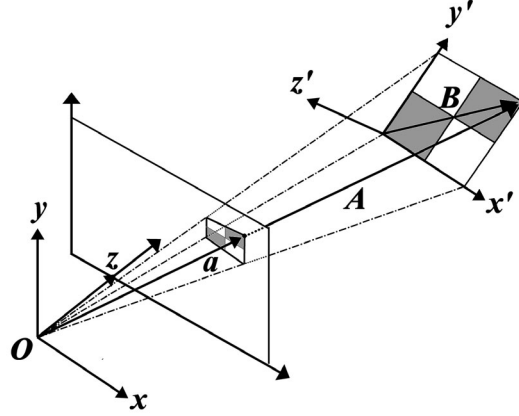
Where  $\lambda$  is a scale factor,  $f_x$  and  $f_y$  describe the focal length and  $(c_x, c_y)$  is the principal point of the camera in pixels.

The vector  $A$  lives in the reference frame of the camera. For our application we would like to transform a reference to a point from one sensor reference frame to another. Suppose there is a vector  $B$  which belongs to another reference frame  $(x', y', z')$  but indicates the same point. A Euclidian transformation, applying a rotation by matrix  $R$  and a translation  $t$ , can be used to map vector  $B$  onto vector  $A$ :

$$A = RB + t \quad (2)$$

The orientation and translation difference between the two reference frames describe the pose and position of a sensor. By embedding the rotation matrix and translation vector into a matrix  $M$ , the projection model can be expanded to include the transformation:

$$\lambda \tilde{a} = PMB = P \begin{pmatrix} R & t \end{pmatrix} \tilde{B} \quad \text{with} \quad \tilde{B} = (B_{x'} B_{y'} B_{z'} 1)^T \quad (3)$$



**Figure 4.** Homography relationship between a plane in 3D space and the camera image plane.

The formula relates a 3D point in an arbitrary reference frame to a point in the camera image. However, the points on the pattern lie on a single plane. Only two dimensions are needed to describe point positions uniquely if the pattern is aligned with the  $x'$  and  $y'$  axis of the local reference frame. Formula 3 can be reduced likewise. If  $r_1$ ,  $r_2$  and  $r_3$  are the column vectors of  $R$  and a point on the pattern is indicated by a homogenous 2D vector  $\tilde{b}$ , then:

$$\lambda \tilde{a} = P \begin{pmatrix} r_1 & r_2 & r_3 & t \end{pmatrix} \tilde{B} \rightarrow \lambda \tilde{a} = P \begin{pmatrix} r_1 & r_2 & t \end{pmatrix} \tilde{b} \quad \text{with} \quad \tilde{b} = \begin{pmatrix} B_{x'} \\ B_{y'} \\ 1 \end{pmatrix} \quad (4)$$

By multiplying the reduced transformation matrix with the projection matrix the homography matrix  $H$  is obtained:

$$P \begin{pmatrix} r_1 & r_2 & t \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & h_3 \end{pmatrix} = H \quad (5)$$

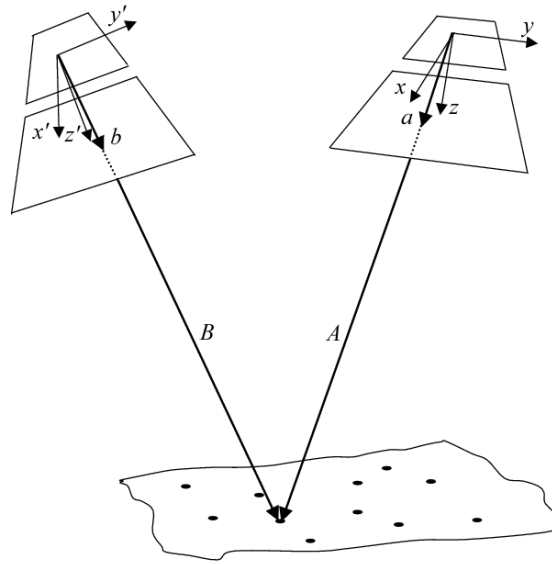
This matrix can be used to relate points on a plane directly to image projection points. Because the geometry of the points on the pattern is known or measured in advance, real distances can now be related to pixel distances in the image. Camera calibration software such as the calibration toolbox of Jean-Yves Bouget<sup>4</sup> exploits this fact in order to estimate parameters such as the focal length and the principal point.

When a camera is calibrated with such a technique it is possible to form matrix  $P$ . With this matrix the orientation and position difference can be extracted from the homography. Zhang<sup>5</sup> showed how to extract the rotation matrix and translation vector from the homography:

$$\begin{aligned} \hat{r}_1 &= \lambda P^{-1} h_1 \\ \hat{r}_2 &= \lambda P^{-1} h_2 \\ \hat{r}_3 &= h_1 \times h_2 \\ t &= \lambda P^{-1} h_3 \end{aligned} \quad \text{with} \quad \lambda = \frac{1}{\|P^{-1} h_1\|} = \frac{1}{\|P^{-1} h_2\|} \quad (6)$$

Due to measurement noise in the image the rotation matrix  $\hat{R}$  formed by the column vectors  $\hat{r}_1$ ,  $\hat{r}_2$  and  $\hat{r}_3$  is not always a true orthogonal matrix with determinant 1. To satisfy the constraints of a true rotation better, single value decomposition can be applied:

$$\text{SVD}(\hat{R}) = USV^T \quad R = UV^T \quad (7)$$



**Figure 5.** Motion of the stereo camera.

## 5. MOTION ESTIMATION OF THE PLATFORM

To relate the estimated pose and position of sensors to specific locations on the ground, the progression of the vehicle has to be measured. We therefore use a second technique which estimates the vehicle motion from the passing terrain in the stereo camera images.

### 5.1. Feature Tracking

Estimating the ego-motion from stereo images requires a set of fixed points in the environment which are tracked between succeeding images and matched between stereo images.

Corner like features can be extracted with the Harris<sup>6</sup> corner detector. Features are first detected in the left image. Their locations are matched between the left and right to find the stereo disparities. For a feature in the left image, normalized cross correlation is used to compare possible corresponding points in the right image within a fixed disparity interval. The best scoring points are selected as the stereo matches.

In order to compute the ego-motion of the cameras, feature correspondences will also have to be found between succeeding stereo images. We use the tracker algorithm of the Intel OpenCV<sup>7</sup> computer vision programming library. It uses the Lucas Kanade method for local optic flow computation to track features.

### 5.2. Motion Estimation

To correctly combine the sensor with vehicle motion, the latter must also be estimated in 3D. In figure 5 the position of the stereo camera is shown at two successive time steps. The cameras see the same ground points, which have been tracked from the first stereo camera to the second. This relates image projection points  $a$  and  $b$  of the tracked features. Stereo vision enables the estimation of distance for image points through triangulation. This provides estimates of the vectors  $A$  and  $B$  which indicate the points on the ground. With the Euclidian transformation of formula 2 vector  $B$  can be transformed into vector  $A$ . The rotation and translation difference between these two vectors describes the vehicle induced motion between the two stereo cameras.

A method for estimating the rotation and translation, given a set of  $A$  and  $B$  vectors, would be a Least Squares (LS) estimate based on formula 2. LS assumes that the error is normally distributed over the vectors. The errors in image feature locations caused by noise can be assumed isotropic because a corner detector is used.<sup>8</sup> However, it is well known that the error of a stereo reconstruction based on noisy image positions is non-isotropic. For a 3D reconstruction point the

error in distance ( $z$ ) will be larger than the other directions. This leads to suboptimal estimates for rotation and translation with LS on only reconstructed vectors from stereo.

Another problem for LS motion estimation is presence of outliers; features with a motion which does not correspond to the ego-motion of the cameras. Outliers can be caused by errors in the tracking and matching of features. In our application we also observed that shadows cast by the sensor frame are a problem. The edges of the shadows remain stationary in the images. Features tracked on these edges do not move correctly with the motion of the vehicle.

A way of incorporating projection into ego-motion estimation was shown by Lasenby<sup>9</sup> et al. The original calibration algorithm was developed to estimate the rotation and translation difference between a number of cameras looking at the same scene. It uses iteration to improve the estimation of the rotations, translations and the distances of points in the scene. The algorithm starts out with a initial estimate of the rotation, translation and the reconstructed feature locations in space. In order to simplify the problem of projection equation 1 is rewritten as:

$$\lambda_i a_i = A_i \quad (8)$$

Similar, the point for the second camera is indicated by the space vector  $B_i$  and is projected onto  $b_i$  with scalar  $\mu_i$ . For all points, the algorithm tries to find the best estimates for  $R$ ,  $t$ ,  $\lambda_i$  and  $\mu_i$  by minimizing the error:

$$E = \sum_{i=0}^n \|\lambda_i a_i - R(\mu_i b_i) + t\| \quad (9)$$

The outline of the iteration in the algorithm is as follows:

1. Update the translation estimate for  $t$  given  $R$ ,  $A_i$ ,  $a_i$ ,  $b_i$ ,  $\lambda_i$  and  $\mu_i$ .
2. Update the rotation estimate for  $R$  given  $t$ ,  $A_i$ ,  $a_i$ ,  $b_i$  and  $\lambda_i$ .
3. Update  $\lambda_i$  and  $\mu_i$  given  $R$ ,  $t$ ,  $A_i$ ,  $a_i$  and  $b_i$ .
4. Update  $A_i$  given  $R$ ,  $t$ ,  $\lambda_i$ ,  $\mu_i$ ,  $a_i$  and  $b_i$ .

When the vectors  $a$  and  $b$  only contain normally distributed image noise and move under the motion defined by  $R$  and  $t$  the minimum is found when the position indicated by  $A$  and  $B$  matches. If outliers are present, some of vectors will not match. The algorithm will then converge to a suboptimal estimate. We expanded the original algorithm with weights. The weights can be used to cancel some of the points from the estimation process. By recomputing the weights per iteration, the outliers can be identified. This enables the algorithm to be robust against the presence of a limited percentage of outliers.

### 5.3. A weighted estimator for translation

A weighted estimator for translation is straightforward. This is simply a weighted average of the differences between the rotated reconstructed vectors and the space vectors.

$$t = \frac{1}{w_{sum}} \sum_{i=0}^n w_i (A_i - \mu_i R b_i) \quad \text{with} \quad w_{sum} = \sum_{i=0}^n w_i \quad (10)$$

### 5.4. A weighted Least Squares estimator for rotation

Rotation can be made independent of translation by centering the space vectors around their centroids. Because each space vector now has its own weight the centroids also have to be computed using a weighted average:

$$\bar{A} = \frac{1}{w_{sum}} \sum_{i=0}^n w_i A_i \quad \bar{B} = \frac{1}{w_{sum}} \sum_{i=0}^n w_i B_i \quad (11)$$

For numerical stability it is always wise to normalize vectors involved in LS fitting. This can be achieved by dividing them through their norm. Because each vector now has equal length the weights can easily be incorporated.

$$\hat{A}_i = w_i \frac{A_i - \bar{A}}{\|A_i - \bar{A}\|} \quad \hat{B}_i = w_i \frac{B_i - \bar{B}}{\|B_i - \bar{B}\|} \quad (12)$$

We seek the best approximation of rotation matrix  $R$ , which rotates  $\hat{A}$  into  $\hat{B}$ . By using the single value decomposition for estimating rotation from formula 7 this is achieved.

### 5.5. Remaining steps, update of the scalars and the space vectors

In the remaining steps the scalars and the space vectors are updated:

$$\lambda_i = \frac{A_i \cdot a_i}{a_i \cdot a_i} \quad \mu_i = \frac{B_i \cdot b_i}{b_i \cdot b_i} \quad (13)$$

$$A_i = \frac{1}{2} (\lambda_i a_i + \mu_i R b_i + t) \quad B_i = \frac{1}{2} (R^{-1} (\lambda_i a_i - t) + \mu_i b_i) \quad (14)$$

### 5.6. Assigning the weights

The weights are used to exclude outliers from the estimation process. At every iteration they are recomputed. Each weight is based on the corresponding error between the reconstruction of  $a_i$  and  $b_i$  under the transformation of  $R$  and  $t$ :

$$E_i = \|\lambda_i a_i - \mu_i R b_i + t\| \quad (15)$$

The variance  $\sigma^2 = var(E)$  indicates the magnitude of the differences in  $E_i$  for the current estimation. When the variance is relatively high, outliers could be present. In this case the weights are assigned as follows:

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

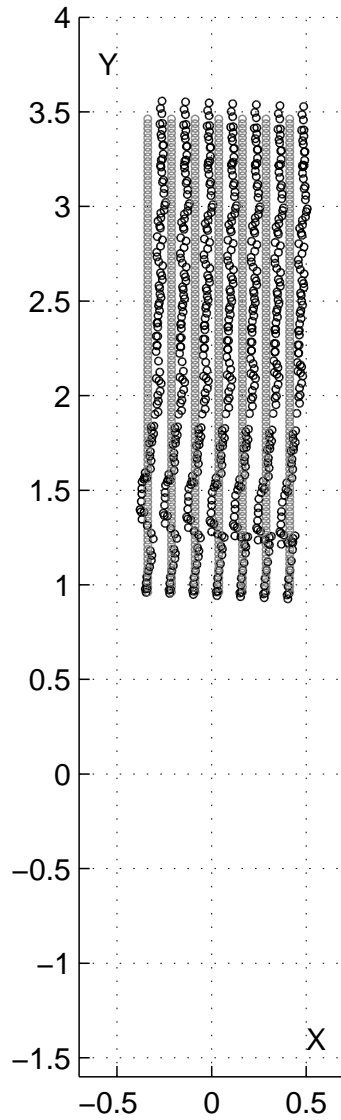
## 6. RESULTS

By concatenating the ego-motion estimates, the vehicle position and orientation can be calculated for each successive stereo pair. Sensors can be aligned with the virtual data-fusion grid by combining the vehicle localization with their estimated pose and position. In the figure 6 to 8 the results are shown for motion estimation with the current odometry based approach and the new vision based method.

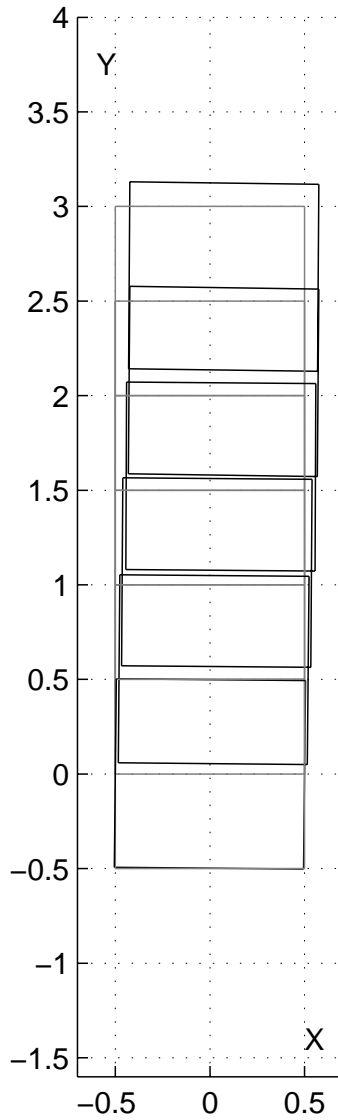
Figure 6 shows the track of the 7 detector coils in the MD sensors. With the color gray the estimated position are shown in the grid based on only odometry. Because this approach assumes that the vehicle moves a fixed distance between every reading only a translation motion is observable. The color black shows the position and orientation by combining estimates using the new vision approach. In the example sequence, the MD is disturbed by some obstructions on the ground. The effect of these disturbances is visible in the figure. The detector coils will be moved a considerable distance to the left or right when the MD strikes against an obstruction. Some parts of the terrain are actually not observed well by the coils, when the MD swings back from the top of an obstacle to the ground. This is visible as the white space between 1.5 and 2.0 meters in the figure.

Figure 7 shows the fields of view of the infrared and multi spectral cameras. Because of the lower sample rate, a new image is only captured by both cameras every 20 trigger pulses. This number was chosen to have an overlap of half a image for every consecutive observation. Gray indicates the estimated position with odometry. For this track the overlap of images seems to be correct. Black shows the estimated positions and orientations with the new vision based approach. In this track we can see that the vehicle makes a slight turn to the right. This has consequences for the overlap of the images and the alignment with the ground plane.

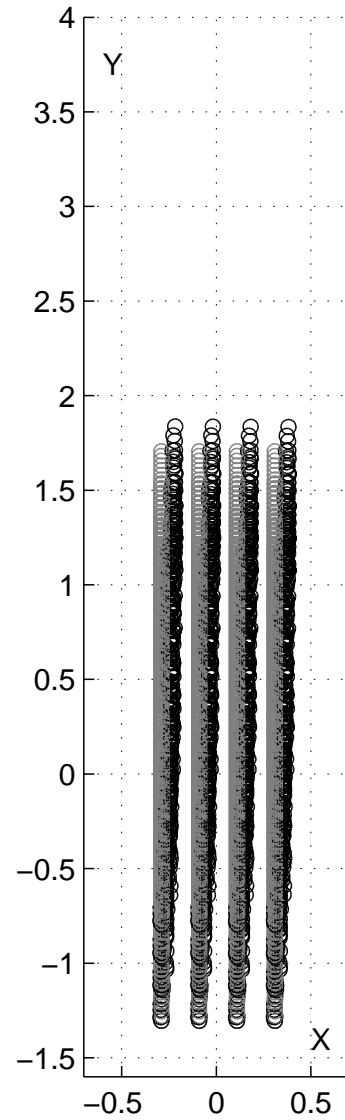
Figure 8 shows the track of the 16 antennas in the GPR sensor. With gray the estimated position of readings using odometry is shown. The result is similar to that of the MD, only translation in one direction is considered. With black the position and orientation is shown given the motion estimates from the vision based approach. In this track we can also see the slight right turn of the vehicle.



**Figure 6.** Estimated alignment of the detector coils in the MD. Gray shows the estimates based on odometry. Black indicates the estimates from stereo vision.



**Figure 7.** Estimated alignment of the IR and multi spectral camera. Gray shows the estimates based on odometry. Black indicates the estimates from stereo vision.



**Figure 8.** Estimated alignment of the antennas in the GPR. Gray shows the estimates based on odometry. Black indicates the estimates from stereo vision.

## 7. CONCLUSION

Landmine detection is a clear example of the importance of motion estimation for vehicles. On the LOTUS several sensors are used to detect different landmine properties. A virtual grid related to the ground is used for data-fusion of the different sensor observations. Accurate incorporation of sensor observations into grid requires the position and orientation of the platform itself and the sensors attached to it. Currently, odometry and fixed offsets between sensors are used for the data-fusion on the LOTUS. The drawback of this approach is that it only considers one dimensional translation of the platform over the minefield. This is not very accurate given the fact that the vehicle will also undergo rotational motions in more rough terrain and that the MD can move independently from the sensor frame.

Our new method uses computer vision to estimate orientation and position of the sensors in the frame and the vehicle in the terrain. The pose and position of sensors is estimated from added patterns with known geometry. An advantage of this approach is that no devices, which could interfere, have to be attached to the sensors.

The motion of the vehicle is estimated from tracked image features on the ground surface. Tracking failures and shadows cast by the sensor frame can cause outliers. By using weights in the estimation process a limited percentage of outliers can be excluded. This enables the robust estimation of three dimensional motion from feature tracks in stereo images.

With a single stereo vision camera we can measure the independent motion of the sensors and keep track of the vehicle position and orientation in three dimensional space. If the two are combined, data from different sensors observed at different time instances can be aligned into the single grid used for sensor-fusion. When compared to the currently used odometry based approach on LOTUS, the new method provides improved sensor localization estimates for the data-fusion in landmine detection.

## 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-personel landmine detection", *Information Fusion 2*, Elsevier, pp. 187, 2001.
2. 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.
3. 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.
4. J-Y Bouguet, *Camera Calibration Toolbox for Matlab*, 2000, <http://www.vision.caltech.edu/bouguetj/calib.doc/>
5. Z. Zhang, "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations", *Proceedings International Conference on Computer Vision*, pp. 666-673, Corfu, Greece, 1999.
6. C. Harris, M. Stephens, "A combined corner and edge detector", *Proceedings 4'th Alvey Vision Conference*, pp. 147-151, 1988.
7. 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.
8. 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.
9. 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.