REAL-TIME OPTICAL POSITION MONITORING USING A REFERENCE BAR

CHARLES BEUMIER, PASCAL DRUYTS, YANN YVINEC AND MARC ACHEROY
Signal and Image Center, Royal Military Academy, Belgium

ABSTRACT

This paper presents a position monitoring system designed for a hand-held mine detector and based on a camera and an accelerometer. Real-time positioning during scanning allows the image reconstruction of the captured signals. Image representation enables object shape analysis and simplifies target localization. The system reports positions relatively to the bar with reference points placed on the ground by the deminer to indicate the safe limit. The present paper focuses on image processing techniques developed to localize the reference points. The required accuracy in positioning for our application has been evaluated to +/- 0.5 cm in each direction.

KEYWORDS

Positioning system, humanitarian demining, distortion compensation.

1. INTRODUCTION

Humanitarian mine clearance is still a tedious task due to the human risk, and the large number of false alarms a standard metal detector issues. Current approaches consider other sensors (IR, radiometer, chemical, ...) to get more information, possibly combining them, to reduce the false alarm rate (see HOPE [1]) while keeping the high probability of detection. In the HOPE project, we also consider the image reconstruction of acquired signals thanks to position estimation of the sensors during scanning. These images will help analysing the shapes of detected objects.

Several solutions can be envisaged to monitor the sensor position.

A gyroscope and accelerometers [2] offer a high precision, although computing the position from the acceleration implies frequent calibrations and the existing hardware solutions are heavy, expensive and require much power.

A DGPS, or differential global positioning system, is light and small, requiring little power. But it is still expensive and not accurate enough for our application.

Ultrasound solutions are based on a sensor and different sources. Although promising in controlled environment, we doubt that the resolution will be robust enough against wind influences. The placement of the sources also leads to practical problems.

The adopted solution in the context of the HOPE project consists of a camera and an accelerometer and is adapted to the practical conditions of a typical demining procedure. The security bar used by the deminer (see Fig. 1) as the limit of the safe area is painted with a specific pattern full of reference points. A camera attached to the sensor head tracks in real time the reference bar and extracts reference points. These allow the determination of the camera position and orientation up to a rotation around the bar. The accelerometer solves this indetermination.

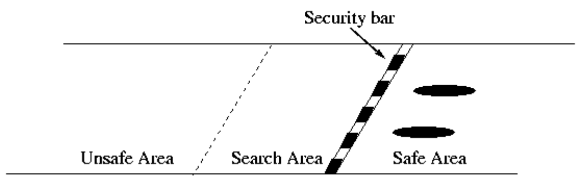

Figure 1: Use of the security bar in a typical demining procedure

The system has been simulated (see [3]). With a normal CCD camera (768x576) and a small focal length (4mm), a sufficiently large number of points of the reference bar are visible during a normal scan of the area (typically 100x50 cm). In those conditions, and with a reasonable error about reference point localization, distortion compensation and accelerometer measurements, the precision requirement of +/- 0.5 cm in all directions appeared possible. We address in this paper the precision in the localization of reference points and the real-time aspect of the related image processing tasks.

2. SYSTEM DESCRIPTION

The optical positioning system developed for the needs of the HOPE project is based on a camera and an accelerometer. The camera tracks reference points of a pattern drawn on the deminer's security bar. The accelerometer gives the gravity direction to solve the rotation indetermination around the bar.

The proposed solution consists of four elements.

a) Reference point localization

Points from the pattern printed on the bar must be localized in the image. The description of the developed image processing techniques is the main topic of the present paper and is given in the next section.

b) Distortion compensation

In order to capture enough points from the bar, at a rather short distance (down to 40 cm), the focal length of the lens must be small (e.g. 4 mm). The implied distortion has been modeled and compensated as summarized in section 4. A detailed presentation is given in [4].

c) Gravity reference

Because the rotation angle around the 1-Dimensional bar cannot be determined by the camera alone (a 2-D bar would be cumbersome and would not provide enough precision in the second direction), a reference orientation is needed. An accelerometer, small and cheap, estimates the gravity direction. However, this supposes that other acceleration sources are known or compensated for. The integration of the accelerometer in the system has not started yet.

d) Mathematical solution

Let x_c, y_c, z_c be the coordinates in the camera axis system; x_i, y_i the image coordinates and x, y, z the coordinates in the absolute axis system linked to the bar (see Fig. 2). We have:

Absolute to Camera $$\begin{pmatrix} xc \\ yc \\ zc \end{pmatrix} = R \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} + T$$

Camera to Image $$\begin{pmatrix} xi \\ yi \end{pmatrix} = \begin{pmatrix} fex.\dfrac{xc}{zc} \\ fey.\dfrac{yc}{zc} \end{pmatrix}$$

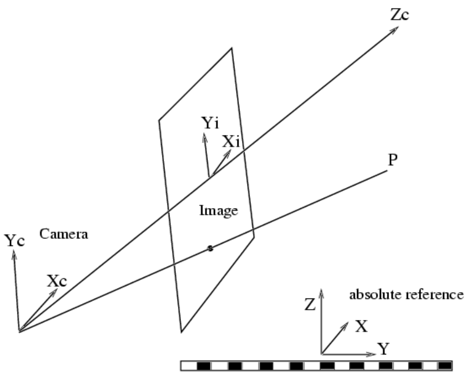

Figure 2 : Coordinate systems

For each reference point along the bar $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ \alpha \\ 0 \end{pmatrix}$, we have, with $R = \left(\overrightarrow{1x} \quad \overrightarrow{1y} \quad \overrightarrow{1z}\right)$, $\overrightarrow{1y} = \begin{pmatrix} \theta 1 \\ \theta 2 \\ \theta 3 \end{pmatrix}$ and $T = \begin{pmatrix} \theta 4 \\ \theta 5 \\ \theta 6 \end{pmatrix}$:

$$\begin{cases} xi.(\alpha.\theta 3 + \theta 6) = fex.(\alpha.\theta 1 + \theta 4) \\ yi.(\alpha\theta 3 + \theta 6) = fey.(\alpha.\theta 2 + \theta 5) \end{cases}$$

This is an overdetermined system in θ_i, with solution up to a multiplicative factor. Dividing the θ_i by $\frac{\theta 6}{\lambda}$, we obtain $\theta = \lambda(\theta'1, \theta'2, \theta'3, \theta'4, \theta'5, 1)$, with λ chosen to have $\theta_1^2 + \theta_2^2 + \theta_3^2 = 1$.

Once the θ_i are identified, $\overrightarrow{1y}$ is known. A second direction is obtained from the vertical thanks to the accelerometer. The third direction completes the trihedral.

3. REFERENCE LOCALIZATION

a) Bar pattern

A special pattern has been designed to be printed on the bar in order to provide for reference points. It consists of two parallel lines (the 'rails') crossed by linear segments showing 'teeth' (see Fig. 3). The crossings are

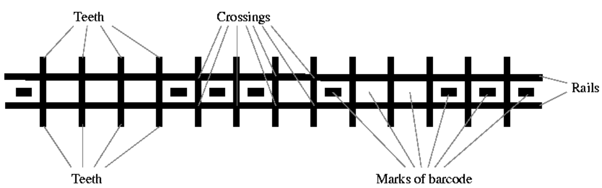

Figure 3: Description of the bar pattern

good candidates for accurate reference points since they can be precisely localized from the intersection of linear segments. Internal rectangular areas are either filled or not, following a binary code which allows the labeling of each crossing. Labeling is necessary to know the coordinates of the reference points in the bar coordinate system.

b) Localizing a tooth

To localize the bar and its reference points as quickly as possible, the teeth are first looked for. Thanks to the pattern shape and the high contrast of the white/black transitions in comparison with the normal contrast of the terrain, a simple procedure has been used to localize a first tooth. It consists in scanning the image and assessing we visit a tooth when the contrast is high in three cardinal directions and small for the fourth one. We distinguish two cases: a top tooth and a bottom tooth (see Fig. 4). Of course, this detection scheme supposes that the bar is rather horizontal in the image but real-time tests showed that a rotation up to 60 degrees from horizontal is allowed, which is much more than the normal rotation of the mine detector during scanning.

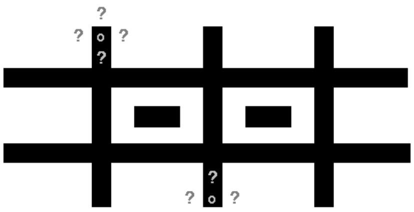

Figure 4: Checking a tooth

The distance between the questionable tooth point and the four cardinal neighbors '?' has been chosen equal to 3 pixels, which is larger than half of the line width, but smaller than any distance between lines of the pattern for the range of scales in normal use.

The test about the contrast is implemented by comparing the gray levels between the expected tooth pixel and the neighbor values: to validate the hypothesis of a tooth, the difference must be large enough (a non-critical threshold value of 50 has been used) for three neighbors and lower than the threshold for the fourth neighbor.

c) Finding chains of teeth

Once a tooth has been found, a chain of teeth is looked for to find other teeth, regularly spaced, to the right of the original tooth. For this, a set of candidate directions covering the half plane is considered, stopping as soon as a tooth is missing around the expected position. During the search, the steps in X and Y directions are updated softly to compensate for the warping effect of the distortion. Because the scale is confined to a range of values and the bar pattern has a fixed size (1 cm between consecutive teeth), a small 5x5 window around the expected position is scanned. A chain is validated when at least 5 consecutive teeth have been found.

For each chain developed on the right side of the originally detected tooth, the left part is looked for by searching in the opposite direction. The chain is stopped when a tooth is missing around the expected position.

d) **Grouping chains**

Chains may be incomplete, due to wrong detection or because the bar is partially occluded. In order to get the best precision, the maximal number of points have to be localized. The different tooth chains have to be merged. Due to the specificity and uniqueness of the bar, this task is easily done by checking that direction and spacing of tooth chains are compatible.

e) **Pairing chains**

Normally, there exists one chain of grouped elements of top teeth and one for bottom teeth so that the pairing is trivial. In case there would be a second bar (to better limit the area to scan and to avoid the use of the accelerometer), the top and bottom chains would be paired based on proximity and compatibility of direction.

Once the chains are paired, the association between top and bottom teeth is obtained by considering the more orthogonal direction between the chains and the crossing segments. With this offset, each top tooth is coupled with the corresponding bottom tooth to form the crossing segments.

f) **Localizing rails**

Several points are extracted to track the rails. They are localized between the crossing segments. The point in the middle of two consecutive top teeth and the point in the middle of the two corresponding bottom teeth define a segment which crosses the two rails. These crossings, called 'rail points', are obtained as the points with maximal gray level transition along the segment.

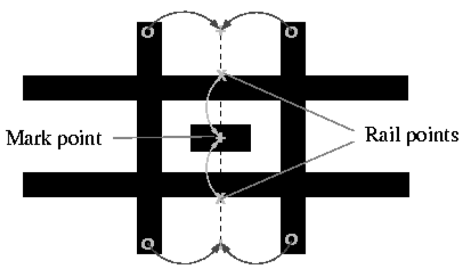

Figure 5 : Localizing rail points

g) **Extracting and checking the barcode**

The labeling of the crossing segments is done indirectly by the presence or absence of dark rectangles in the pattern (see 'Marks of barcode' in Fig. 3). With a sufficient number of consecutive marks (1 if dark, 0 else), one is able to give the position of the marks in the complete barcode sequence. This position will be used as label for the corresponding crossing segment.

The extraction of the barcode is performed by measuring the average gray level around the mark point '+' situated in the middle of the two rail points 'x' (see Fig. 5). This gray value is compared with an average of the gray values of the other marks to decide for a 0 (bright) or 1 (dark).

The sequence of 0 and 1 is then compared with the reference barcode to identify the label of each crossing. The adequacy of the sequence with the reference is a way to check that the localized chains belong to the bar. Chains with less than 7 teeth can normally not solve their labeling, except if neighboring chains have a correct labeling.

h) **Reference points**

The reference points '*' are situated at the highest bright/dark transition around the middle points between two consecutive mark points '+'.

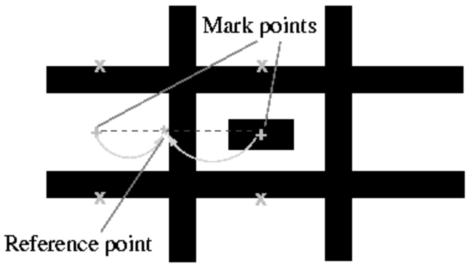

Figure 6: Reference point

4. DISTORTION COMPENSATION

In order to capture large areas (typically 1.0 x 0.5 m) at short distances (down to 40 cm), the camera lens must have a short focal length. This induces important distortion effects which impair the image reconstruction fidelity and limit the accuracy for image positioning.

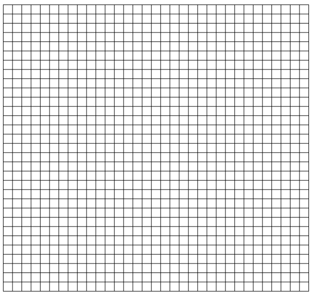

Figure 7: reference grid

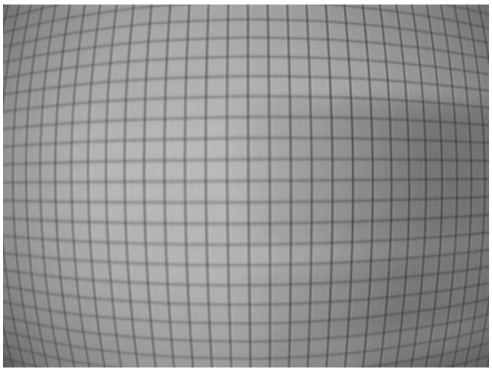

Figure 8 : Distorted image

The compensation of distortion is based on the capture of a grid (see Fig. 7 and 8) offering the crossings as reference points. These points of known coordinates in the grid are detected in the image to measure the deformation. With the assumption that the grid is perfectly perpendicular to the optical axis of the camera and that the image deformation is mainly due to radial distortion (with rotational symmetry) and de-centering [5], we modeled the deformation with the polynomial:

$$
\begin{aligned}
corr &= 1 + K1_x X + K1_y Y + K2_{xx} X^2 + K2_{xy} XY + K2_{yy} Y^2 \\
&+ K3_{xxx} X^3 + K3_{xxy} X^2 Y + K3_{xyy} XY^2 + K3_{yyy} Y^3 \\
&+ K4_{xxxx} X^4 + K4_{xxxy} X^3 Y + K4_{xxyy} X^2 Y^2 + K4_{xyyy} XY^3 + K4_{yyyy} Y^4 \\
&+ \ldots
\end{aligned}
$$

$$
\begin{aligned}
xG - xG_c &= SC_x \times X \times corr \\
yG - yG_c &= SC_y \times Y \times corr \\
X &= x - x_c \\
Y &= y - y_c
\end{aligned}
$$

where x, y are the image coordinates, xG, yG are the grid coordinates, SCx and SCy are scale factors and 'c' subscript relates to the image or grid center. The coefficients ($K1_x$, $K1_y$, $K2_{xx}$, $K2_{xy}$, …) and parameters (SC_x, SC_y, xG_c, yG_c, x_c, y_c) are updated to reduce the average error between the image positions transformed by the model and the theoretical grid positions. With coefficients up to the 4^{th} power and a focal length of 4 mm, the average localization error after correction is 0.5 pixel. This error reflects the model limitation and the imprecision in the automatic crossing localization.

Crossing point extraction and model fitting is normally done once since the parameters remain valid as long as the camera lens is not modified. Compensated positions are obtained by applying the above model, a quick procedure involving few computations and to be repeated for a few points (maximum 100 reference points).

5. RESULTS

The image processing techniques presented above were tested relatively to time consumption and accuracy. The software development followed time optimization so that the current solution takes about 5 ms to extract the reference points from the image. At the video speed of 25 images per second, this consumes only a part of the available 40 ms so that other processing concerning the accelerometer and the mathematical solution can be done.

During the simulations confirming the possibility to achieve the precision of +/- 0.5 cm, we modeled the error for the localization of the reference points with a uniform distribution between 0 and 1 pixel. To see if this is realistic, we need to compare the obtained positions with some references. To obtain references, we placed the bar perpendicularly to the camera so that the reference points should be regularly spaced on a line when the distortion is compensated. We thus tried to fit the localized reference points with a set of equidistant points on a line. The error measure is the average distance of the localized points with the model points. A few trials with the bar at different positions and orientations gave the same order of error: 0.6 pixel of average distance. Considering that the distortion was not completely compensated (the border errors were larger) and the point localization was not subpixel accurate, we are confident to offer a precision better than what the simulations assumed.

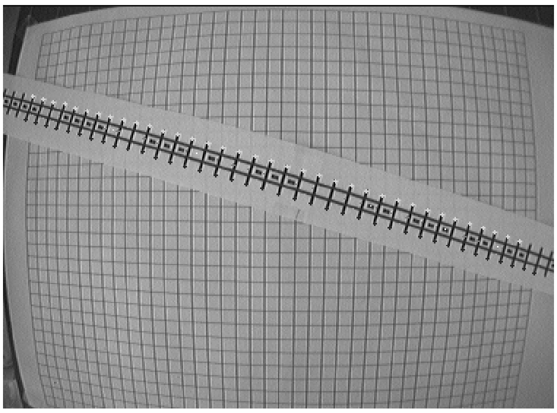

Figure 9 : Localization of reference points

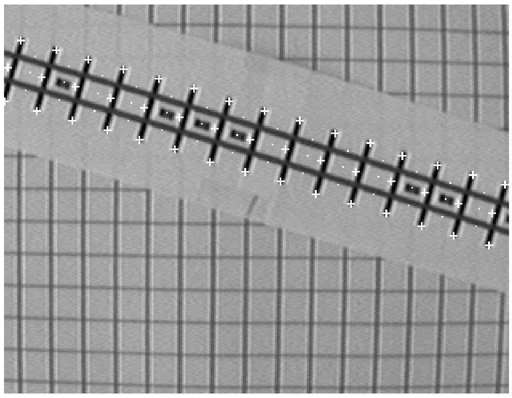

Figure 10 : Zoom of a part of figure 9

Fig. 9 shows the localization of the reference points with rail and mark points and teeth for an image containing a bar (here a band of paper) on a grid (to check the robustness of tooth detection). Fig. 10 displays a zoom area of Fig. 9.

6. CONCLUSIONS

A position monitoring system for hand-held mine detector has been presented. It consists of a camera and an accelerometer and reports positions and orientations relatively to the security bar used by the deminers to delimitate the security area and the search area.

The whole system has first been simulated to estimate the possible precision of the approach with a common (cheap) camera and accelerometer. The present paper has described the processing techniques developed to localize automatically the reference points printed on the security bar. The solution is fully automatic, fast enough to operate in real-time at video speed and offers a precision which is in accordance with the simulation hypothesis about image localization to achieve the desired accuracy of +/- 0.5 cm in all directions.

Future work will concern subpixel localization of reference points and the integration of the accelerometer with compensation of the acceleration of the detector.

7. ACKNOWLEDGEMENT

This work has been supported by the HOPE project of the European ESPRIT programme.

Bibliography

[1] HOPE, “Hand-held OPErational demining system (HOPE)”, Esprit Project 29870, DG III.

[2] I. Popova, “Gyrostabiliser for Mobile Antenna Post”, PhD thesis, Technische Universiteit Delft, The Netherlands, 1996.

[3] C. Beumier, P. Druyts, Y. Yvinec, M. Acheroy, “Motion estimation of a hand-held mine detector”, *Signal Processing Symposium*, Hilvarenbeek, NL, 23-24th March 2000.

[4] C. Beumier, M. Acheroy, “Compensation of camera distortion”, Submitted to BMVC2000.

[5] M. Sonka, V. Hlavac, R. Boyle, *Image Processing Analysis, and machine Vision*, PWS Publishing, p. 457.