Proceedings of the 2005 IEEE
International Conference on Robotics and Automation
Barcelona, Spain, April 2005

Using Laser and Vision to Locate a Robot in an Industrial Environment: A Practical Experience*

Guillem Alenyà*, Josep Escoda†, Antonio B. Martínez† and Carme Torras*
* *Institut de Robòtica i Informàtica Industrial (UPC-CSIC)*
Llorens i Artigas 4-6, 08028 Barcelona, Email: {galenya,torras}@iri.upc.es
† *Department d'Enginyeria de Sistemes, Universitat Politècnica de Catalunya*
C/ Pau Gargallo, 5. 08028 Barcelona, Email: {josep.escoda,antonio.b.martinez}@upc.es

***Abstract*— The fully flexible navigation of autonomous vehicles in industrial environments is still unsolved. It is hard to conciliate strict precision requirements with quick adaptivity to new settings without undergoing costly rearrangements. We are pursuing a research project trying to combine the precision of laser-based local positioning with the flexibility of vision-based robot motion estimation. An enhanced circle approach to dynamic triangulation combining laser and odometric signals has been used to improve positioning accuracy. As regards to vision, a novel technique relating the deformation of contours in an image sequence to the 3D motion underwent by the camera has been developed. Interestingly, contours are fitted to objects already present in the environment, without requiring any presetting. In this paper, we describe a practical experience conducted in the warehouse of a beer production factory in Barcelona. A database containing the laser readings, image sequences and robot odometry along several trajectories was compiled, and subsequently processed off-line in order to assess the accuracies of both techniques under a variety of circumstances. In all, vision-based estimation turned out to be about one order of magnitude less precise than laser-based positioning, which qualifies the vision-based technique as a promising alternative to accomplish robot transfers across long distances, such as those needed in a warehouse, while backing up on laser-based positioning when accurate docking for loading and unloading operations is needed.**

I. Introduction

Mobile robots are increasingly used in flexible manufacturing industry and service environments. The main advantage of these vehicles is that they can operate autonomously in their workspace. To achieve this automation these vehicles must include a positioning -or localization- system in order to provide the robot with position and orientation in the plane as accurately as possible [12].

In the past two decades, several approaches have been proposed to solve the positioning problem. These can be classified into two general groups [5]: absolute and relative positioning. Absolute positioning methods estimate the robot position and orientation in the workspace by detecting some landmarks in the robot environment. Two subgroups can be further distinguished depending on whether they use natural or artificial landmarks. Approaches based on natural landmarks use distinctive features in the environment that have a function other than robot navigation. Conversely, artificial landmarks are placed at known locations in the workspace with the sole purpose of enabling robot navigation. The laser-based technique used in this work falls in the latter category, its main advantage being the high precision that can be attained.

Relative positioning methods, on the other hand, compute the robot position and orientation from an initial configuration, and, consequently, are often referred to as motion estimation methods. A further distinction can also be established here between incremental and non-incremental approaches. Among the former are those based on odometry and inertial sensing, whose main shortcoming is that errors are cumulative.

Vision can be used under any of the mentioned approaches. It can serve to detect artificial landmarks [13], [17] or natural ones [3], [18] for absolute positioning, or rather be applied to perform motion estimation, as in the present work. Differing from the other such estimation methods mentioned above, our vision-based technique is not incremental and, therefore, doesn't suffer from the cumulative error problem.

Precision is expensive in terms of both presetting of the environment and sensor resolution. Thus, a cost-effective solution tries to attain it only when needed (in docking for loading and unloading operations), while backing up on a cheaper technique in the other cases. This is the final aim of our project and, to this aim, we are experimenting with a laser-based goniometer using catadioptric marks for accurate positioning, and a monocular vision system that estimates motion on the basis of only the deformation of natural landmarks as seen by the robot as it moves.

In order to force both techniques to work under realistic conditions, experiments were carried out in an industrial plant, namely a warehouse of the brewer company DAMM in El Prat de Llobregat, Barcelona.

After briefly explaining the two techniques in Sections

*This applied research is sponsored by the Catalan Research Directorate, in the framework of its Reference Center for Advanced Production Technologies (CeRTAP). Guillem Alenyà and Carme Torras acknowledge support from the Spanish Science and Technology Council, project DPI2003-05193-C02-01, Guillem Alenyà was partially funded by the Marie Curie FP5 Program of the European Commission, under contract HPMI-CT-2001-00156 .

2 and 3, the experimental set-up is described in Section 4, and the results of the practical experience are discussed in Section 5, before closing the paper with some conclusions and prospects for future work.

II. Laser-based position estimation

Many approaches have been proposed to fuse dead reckoning with an absolute positioning system in order to solve the dynamic positioning problem. Most of these approaches use an Extended Kalman Filter -EKF- to combine all measurement data [11]. However, several authors agree that in real operation the signals used have nongaussian noise density and, consequently, other recursive algorithms have been put forth [10], [16].

In this work robot positioning is achieved by dynamic triangulation relying on a geometric algorithm based on circle intersections [6]. The proposed technique [2] takes into account the kinematics of the vehicle to estimate in real time the evolution of the length (ρ_i) and bearing (θ_i) between the laser-based goniometer and the set of landmarks used (see Fig.1). The evolution of these variables can be expressed as:

$$\dot{\rho}_i = -\left(v_{1p}\cos\theta_i + v_{2p}\sin\theta_i\right)$$
$$\dot{\theta}_i = \frac{\left(v_{1p}\sin\theta_i - v_{2p}\cos\theta_i\right)}{\rho_i} - \dot{\Psi}$$

where v_{1p} and v_{2p} are the components of the speed of the driving wheel expressed in the laser reference system (P), and Ψ is the orientation of the vehicle in world coordinates.

Time integration of these equations between actual measurements leads to the real-time estimation of ρ_i and θ_i:

$$\rho_i(t) = \rho_i(t_k) + \int_{t_k}^{t} \dot{\rho}_i dt$$
$$\theta_i(t) = \theta_i(t_{i,k}) + \int_{t_{i,k}}^{t} \dot{\theta}_i dt$$

where t_k is the time of the last positioning measurement, and $t_{i,k}$ is the time of the last reflection from landmark i. In this way, the static triangulation algorithm can be accurately applied at any time under dynamic conditions. The computational cost of the presented technique has been shown to be lower than the one required by other methods based on predictive algorithms [2].

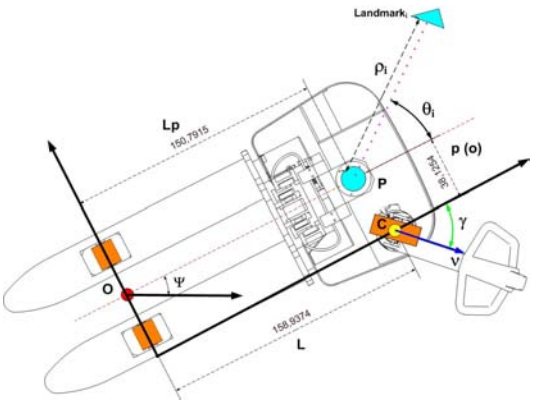


Fig. 1. Different reference frames defined in the forklift mobile vehicle.

The experiments have been performed with a laser-based goniometer (from a Guidance Control Systems Ltd.) rotating at 8 *Hz*, with an accuracy of 0.095 *mrad* and with a maximum reflection distance of 30 m. Catadioptric rectangles - retro-reflecting - that polarize the laser signal are used as landmarks. The positions of these landmarks have been topographically measured with submillimeter accuracy in order to reduce the uncertainties of the method.

The accuracy of the triangulation algorithm depends on the landmark arrangement, the position of the vehicle in the workspace, and the laser-based goniometer resolution. Figure 2 shows the accuracies of positioning measurements obtained under static condition in the laboratory environment, represented over a $4m \times 6m$ grid. Under static condition, the maximum error -worst accuracy- is located at the bottom-left-hand corner of the laboratory and its value is 1.72 *mm*. The best accuracy (less than 0.25 *mm*) is obtained near the center of the laboratory. The accuracy variability is due to the discretization of the laser-based goniometer.

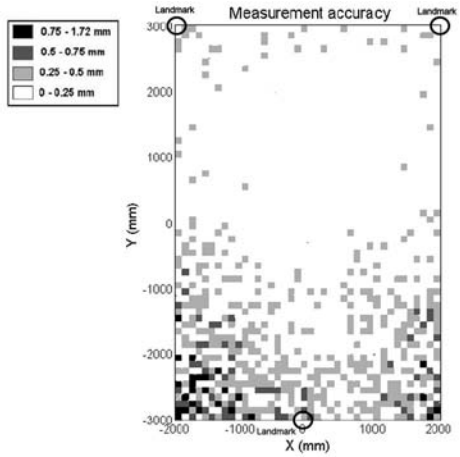


Fig. 2. Measurement accuracy of the triangulation algorithm used under static condition in the laboratory.

III. Vision-based motion estimation

By observing the deformation of an active contour along an image sequence, egomotion can be computed. Equations have been developed [14], [15] to obtain the complete motion parameters: translation and rotation, up to a depth scale factor. Note that the image sequence is acquired with a monocular camera system, so depth can only be recovered with some additional information [7].

A weak-perspective camera model is assumed. This simplified model is a linear approximation to the pinhole camera when some conditions are satisfied [9] as follows. The first one is a small field of view: the farther the image points are from the image center, the worse is the projection approximation. So interest points (in the current work, the points defining a contour) must be near the center of the image. The second condition is that the contour should be

planar, i.e., all points used to model the contour must have small depth variations compared to the distance from the camera to the object.

Assuming only rigid motions of a planar contour, it can be proved that transformations in 3D space project in the camera plane as affine deformations of the contour [14]. A contour can be defined in b-spline form just by providing the set of control points Q [8]. Consider Q_0 the initial state of the contour, then the deformed contour Q can be represented as

$$Q = WS + Q_0, \tag{1}$$

where W is the *shape matrix*, determining the family of all possible affine deformations, and S is the *shape vector*, coding a particular one.

In the general 6 d.o.f. case of planar affine transformations, the matrix W has the form

$$W = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} Q_0^x \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ Q_0^y \end{bmatrix}, \begin{bmatrix} 0 \\ Q_0^x \end{bmatrix}, \begin{bmatrix} Q_0^y \\ 0 \end{bmatrix} \right), \tag{2}$$

and the shape vector is expressed as [15]:

$$S = (t_x, t_y, M_{11} - 1, M_{22} - 1, M_{21}, M_{12})^T. \tag{3}$$

where t_x and t_y correspond to the translations in the plane, and $[M_{ij}]$ is the matrix encoding rotation and scalings.

Different deformation spaces can be defined corresponding to constrained robot motion. In the case of a planar robot, with 3 degrees of freedom, the working space is parametrized with two translations (T_x, T_z) and one rotation (θ_y). Obviously, the rest of the possible motions are not possible with this kind of robot. Forcing these constraints in the equations of the affine deformation of the contour, a new shape space can be deduced. This corresponds to a shape matrix with also three dimensions.

However, for this to be so, both reference frames, the camera one and the robot one, should be perfectly aligned. Any misalignment between both frames will result in errors in tracking and egomotion computation, since deformations produced in the camera plane will not be well modelled by the reduced shape matrix.

To avoid misalignment problems and ensure a correct contour tracking, it suffices to allow the shape space to include also the parameters that model T_y translations. The deduced shape matrix can then be expressed as

$$W = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} Q_0^x \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ Q_0^y \end{bmatrix} \right), \tag{4}$$

and the shape vector as

$$S = (t_x, t_y, M_{11} - 1, M_{22} - 1)^T. \tag{5}$$

Note that each shape vector parameterizes the deformation of the active contour in relation to the initial set of points. The main advantage is that, since egomotion is expressed relatively, this method does not suffer from typical bias problems.

Once a contour deformation has been obtained as a shape vector, it is possible to derive the 3D camera motion up to a scale factor. Centering the reference frame on the calibrated camera, the equations to obtain the scaled translations are

$$\frac{T_x}{Z_0} = \frac{t_x}{f\sqrt{\lambda_1}} - R_{13}, \tag{6}$$

$$\frac{T_y}{Z_0} = \frac{t_y}{f\sqrt{\lambda_1}} - R_{23}, \tag{7}$$

$$\frac{T_z}{Z_0} = \frac{1}{\sqrt{\lambda_1}} - R_{33}., \tag{8}$$

where f is the focal length of the camera, Z_0 is the estimated depth of the plane containing the contour points, λ_1 is the largest eigenvalue of the $\mathbf{M}\mathbf{M}^t$ matrix with $\mathbf{M} = [M_{ij}]$ as before, and R_{ij} are components of the Euler rotation matrix

$$\mathbf{R} = \mathbf{R_z}(\phi)\mathbf{R_x}(\theta)\mathbf{R_z}(\psi). \tag{9}$$

This matrix gives the rotation components of the 3D motion, and can be deduced with equations

$$cos\theta = \sqrt{\frac{\lambda_2}{\lambda_1}}, \tag{10}$$

$$\mathbf{v_1} = \begin{bmatrix} cos\phi \\ sin\phi \end{bmatrix}. \tag{11}$$

where $\mathbf{v_1}$ is the eigenvector of the largest eigenvalue, and

$$\mathbf{R_z}|_{\mathbf{2}}(\psi) = (1 + \tfrac{T_z}{Z_0}) \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{cos\theta} \end{bmatrix} \mathbf{R_z}|_{\mathbf{2}}(-\phi)\mathbf{M}. \tag{12}$$

IV. Set-up description

The mobile robot used (see Figs. 1 and 3) is a Still EGV-10 modified forklift. This is a manually-guided vehicle with aids in the traction. To robotize it, a motor was added in the steering axis with all needed electronics. This electronics was specifically designed. The kinematics of this platform is that of a tricycle-like vehicle, and it is provided with a driving encoder and a steering encoder for dead reckoning. These odometric sensors are used to determine the variables v and γ shown in Fig. 1.

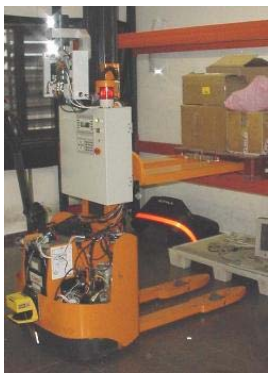


Fig. 3. The mobile robot used in the experience.

The hardware used to compute the vehicle positioning is an industrial PC (PC104 based) Pentium III Celeron clocked at 400 *Mhz* smartcore. This PC runs with a real-time OS RT-Linux 3.2. For the odometric and laser signals capture, specific firmware implemented in FPGA is applied.

By means of this equipment, positioning can be calculated at a rate as fast as one measurement every 126 *ms*.

The practical experience was carried out in a warehouse of the brewer company DAMM in El Prat del Llobregat, Barcelona. Figure 4 shows the environment of the mobile robot, a portion map of the factory. One can see the catadioptric landmarks (Rx).

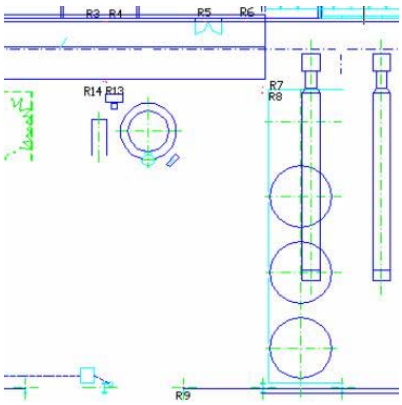


Fig. 4. Map of the plant where the experiment was carried out.

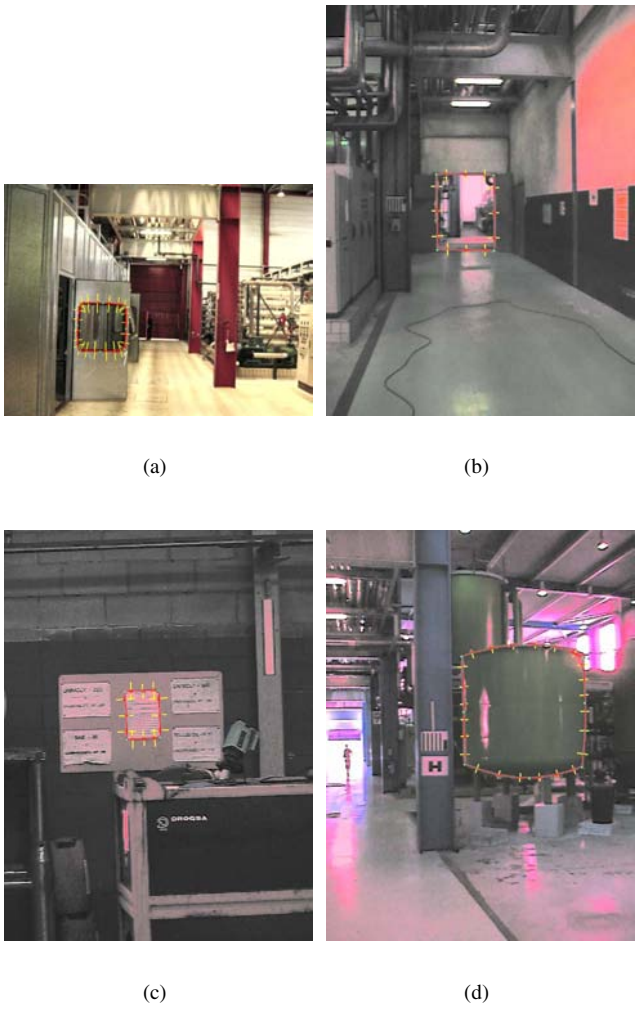
During the experience, the robot was guided manually. A logger software recorded the following simultaneous signals: the position obtained by dynamic triangulation using a laser-based goniometer, the captured reflexes, and the odometry signals provided by the encoders. At the same frequency, a synchronism signal was sent to the camera and a frame was captured.

A log file was created with the obtained information. This file permitted multiple processing to extract the results for the performance assessment and comparison of the presented estimation techniques. Although this experiment was designed in two steps: data collection and data analysis, the current implementations of both algorithms run in real time, that is, 20 *fps* for the camera subsystem and 8 *Hz* for the laser subsystem.

V. Experimental results

Several experiments were performed involving trajectories of different types (straight, oscillating, turning), varied environments and landmarks, and also contours falling in different positions in the image (Fig. 5). Among these experiments, a representative one has been chosen to be presented in detail below.

For the vision subsystem, the set of data to analyze were 200 frames. An active contour was initialized manually on an information board appearing in the first frame of the chosen sequence (Fig. 6). This board was used as target object to track, since it appears entirely in all frames. Once the active contour is initialized in one frame, the tracking algorithm finds the most suitable affine deformation of this contour that fits the target in the next frame, yielding an estimated affine deformation [4]. This is expressed in terms of a shape vector (5), from which the corresponding Euclidean 3D transformation is derived: a translation vector (equations 6-8) and a rotation matrix (equations 9-12).



(a) (b) (c) (d)

Fig. 5. Some of the experiments performed to log data. First frames, with contours fitted to salient objects, are shown for four sample robot trajectories.

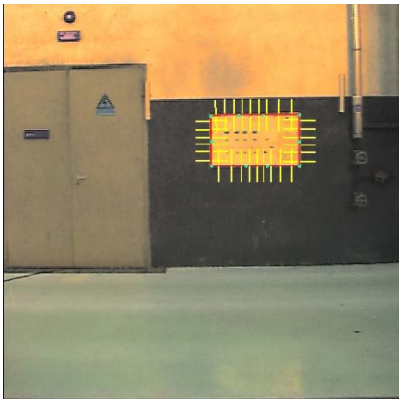


Fig. 6. First image, with an active contour fitted to an information board, of the sequence for which results are presented in detail.

The tracking process produces a new deformation for each new frame, from which 3D motion parameters are obtained. If the initial distance Z_0 to the target object can be estimated, a metric reconstruction of motion can be

accomplished. In the present experiment, the value of the initial depth was estimated with the laser sensor, as the target (the information board) was placed in the same wall as some catadioptric marks, yielding a value of 7.7 *m*. The performed motion was a translation of approximately 3.5 *m* along the heading direction of the robot perturbed by small turnings. The computed T_x, T_y and T_z values can be seen in Fig. 7. The reason why the T_y translation is also computed has been explained in Section III.

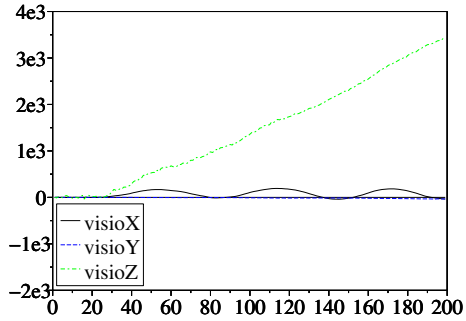


Fig. 7. Translati on im millimeters computed by the vision subsystem for each frame.

Placing the computed values for the X and Z translations in correspondence in the actual motion plane, the robot trajectory can be reconstructed (Fig. 8).

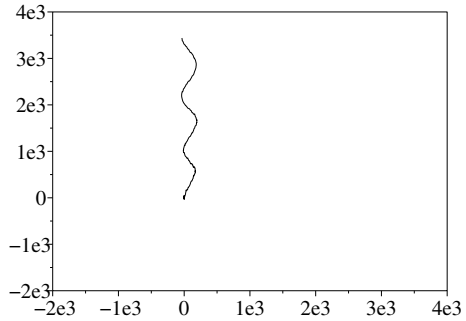


Fig. 8. Trajectory in millimeters in the *XZ* plane computed by the vision subsystem.

Once the laser data was processed off-line, the computed motion estimation was obtained. As explained in Section II, this estimation can be globally located in the working space if landmark positions are known. As mentioned above, catadioptric positions where precisely placed in the digital factory map (refer to Fig. 5); for this particular run, the positioning method used landmarks R5, R6 and R9, which were automatically selected by the algorithm as the most informative ones. Thus, the computed motion in the plane can be represented within the map (Fig. 9).

Extrinsic parameters from the laser subsystem and the vision subsystem are needed to be able to compare the obtained results. They provide the relative position between both acquisition reference frames, which is used to put in correspondence both position estimations. In the moment of the data collection no specific calibration method was

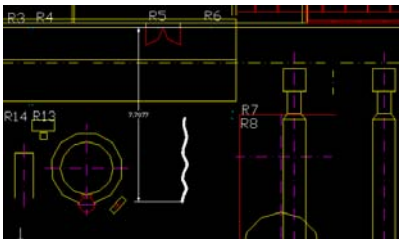


Fig. 9. Trajectory computed by the laser subsystem and plotted in the factory map.

available. Two catadrioptic landmarks (R5 and R6) used by the laser were placed in the same plane as a natural landmark used by the vision tracker. A rough estimation of the needed calibration parameters (d_x and d_y) was obtained with measures taken from controlled motion of the robot towards this plane, yielding the values of 30 *mm* and 235 *mm*, respectively. To perform the reference frame transformation the following equations were used:

$$x_{cam} = x - (sin(\psi) * d_x) + (cos(\psi) * d_y),$$
$$y_{cam} = y - (sin(\psi) * d_x) + (cos(\psi) * d_y).$$

While laser measurements are *global*, the vision system ones are relative to the initial position taken as reference [15]. To compare both estimations, laser measurements have been transformed to express measurement increments.

The compared position estimations are shown in Fig. 10, where the vision estimation is subtracted from the laser estimation to obtain the difference for each time step.

The computed difference in the Z direction is more noisy, as estimations from vision for translations in such direction are more ill conditioned than for the X or Y directions. Physically, translation of the camera along the latter directions induces a larger change in the angle of incidence of the projection rays from the contour to the image plane. Rather, translations along the Z direction induce a smaller change, which is harder to sense due to the pixelization effect. In all, it is remarkable that the computed difference is only about 3%.

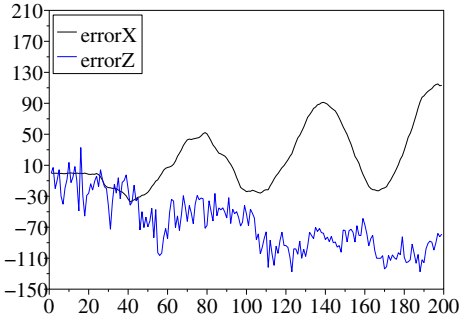


Fig. 10. Difference in millimeters between translation estimates provided by the laser and the vision subsystems for each frame.

The computed differences in X are less noisy, but follow the robot motion. Observe that for larger heading motions, the difference between both estimations is also larger. This

is probably due to misalignments in the reference frame transformation from the laser to the camera.

Finally, to compare graphically both methods, the obtained translations are represented in the *XZ* plane (Fig. 11).

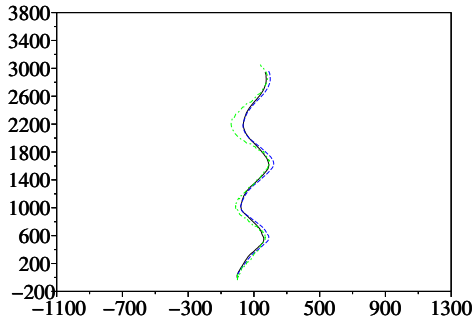


Fig. 11. Trajectories in millimeters in the *XZ* plane. The black line corresponds to the laser trajectory, the blue dashed line to the laser-estimated camera trajectory, and the green dotted one to the vision-computed camera trajectory.

VI. Conclusions and Future Work

A practical experience of mobile robot localization is presented. A heavy industrial robot was placed in a real production plant, and some experiments including data collection were performed. The robot was equipped with two main sensor subsystems: laser goniometer and image acquisition.

For the laser subsystem, it was indispensable to place some catadioptric landmarks. These landmarks were very precisely located in the environment, and this permitted the computation of a global referenced motion.

For the vision subsystem, no intervention in the environment was needed. The obtained motion was always given relative to the initial robot position, although not in an incremental way, thus errors were not cumulative. It has been shown that, knowing the initial target depth, the scale indetermination can be removed. In the same manner as in the laser case, if the target is located in the environment map, the global position can also be computed.

The relatively small deviation (about 3%) of the vision-based estimation with respect to the laser-based one qualifies the former technique as a promising alternative in situations with low-precision demands, and encourages us to continue our investigation of motion estimation based on active contours.

The comparison between both estimations has shown the importance of an accurate laser-camera extrinsic parameter calibration. It may be the case that part of the error attributed to the vision-based estimation comes in fact from a slight miscalibration and, thus, the precision of the proposed contour-based technique may turn out to be even better. To elucidate this, an algorithm to obtain the extrinsic laser-camera parameters with high precision needs to be developed. Moreover, a physical device to record robot trajectories is being developed in order to get a "ground truth" against which to compare both estimations.

The presented experiments have involved comparisons between two techniques for position estimation with the aim of assessing their respective performances in a variety of situations within an industrial environment. Future work will combine both techniques in a single procedure in order to exploit the accuracy of laser when needed (in docking, narrow crossings, etc.) and to fall back on vision for transfer between these high-precision demanding places.

Acknowledgments

The authors thank the brewer company DAMM, and specially the workers of its plant in El Prat de Llobregat for their help in performing the experiments. The work of Xevi Castell in acquiring the video sequences is particularly acknowledged, and also the collaboration of Toni Benedico.

References

[1] G. Alenyà, E. Martínez, C. Torras, Fusing visual and inertial sensing to recover robot egomotion, Journal of Robotic Systems, Vol. 21, pp.23-32, 2004.

[2] J. A. Batlle, J. M. Font, J. Escoda, Dynamic positioning of a mobile robot using a laser-based goniometer, IEEE IFAC Symposium on Intelligent Autonomous Vehicles, Lisboa 2004.

[3] M. Betke and L. Gurvits, Mobile robot localization using landmarks. IEEE Trans. on Robotics and Automation, 13(2): 251-263, 1997.

[4] A. Blake and M. Isard, Active contours, Springer, 1998.

[5] Borenstein J., H. R. Everett, L. Feng and D. Wehe, Mobile Robot Positioning - Sensors and Techniques. Journal of Robotic Systems, Vol.14, No. 4, pp. 231 - 249, 1997.

[6] C. Cohen and F.V. Koss A Comprehensive Study of Three Object Triangulation SPIE Vol. 183 Mobile Robots VII (1992) / 95

[7] A. Criminisi, Accurate visual metrology from single and multiple uncalibrated images, Springer, 2001.

[8] J. Foley, A. van Dam, S. Feiner and F. Hughes, Computer Graphics. Principles and practice. Addison-Wesley Publishing, 1996.

[9] D. Forsyth and J. Ponce, Computer vision, a modern approach, Prentice Hall, 2003.

[10] Hanebeck, U. D. and G. Schmidt, Set Theoretic Localization of Fast Mobile Robots Using an Angle Measurement Technique. Proc. of the IEEE Int. Conference on Robotics and Automation, pp.1387-1394, 1996.

[11] Hu, H. and D. Gu, Landmark-based Navigation of Autonomous Robots in Industry. International Journal of Industrial Robot, 27(6): 458-467, 2000.

[12] Leondes, C. T., Mechatronic Systems Techniques and Applications. Vol. 2 Transportation and vehicular Systems, pp. 338 - 359. Gordon and Breach Science Publishers, Amsterdam, 2000.

[13] G. Jang, S. Kim, W. Lee and I. Kweon, Color landmark-based self-localization for indoor mobile robots. IEEE Intl. Conf. on Robotics and Automation (ICRA'02), Washington D.C., pp. 1037-1042, 2002.

[14] E. Martínez, "Recovery of 3D structure and motion from the deformation of an active contour in a sequence of monocular images", PhD Thesis, 2000.

[15] E. Martínez and C. Torras, Qualitative vision for the guidance of legged robots in unstructured environments, Pattern Recognition, 34: 1585-1599, 2001.

[16] Sasiadek, J. Z. and P. Hartana, Sensor Fusion for Dead-reckoning Mobile Robot Navigation. A Proceedings volume from the IFAC Workshop on Mobile Robot Technology (J. Sasiadek. (Ed)), pp. 251 - 256. Pergamon, Oxford, 2001.

[17] D. Scharstein and A. Briggs, Real-time recognition of self-similar landmarks. Image and Vision Computing, 19(11): 763-772, September 2001.

[18] R. Sim and G. Dudek, Learning environmental features for pose estimation. Image and Vision Computing, 17: 445-460, 2001.