Stereo Vision Based Navigation for Automated Vehicles in Industry

Giacomo Spampinato, Jörgen Lidholm, Lars Asplund, Fredrik Ekstrand
School of Innovation, Design and Engineering
Mälardalen University, Sweden
{giacomo.spampinato,jorgen.lidholm,lars.asplund,fredrik.ekstrand}@mdh.se

Abstract

This paper proposes a stereo vision based localization and mapping strategy for vehicular navigation within industrial environments using natural landmarks. The work proposed is strictly related to factory automation, since focus is on industrial vehicle autonomous navigation for material handling, in order to increase the operating efficiency with reduced risk for accidents. The stereovision system, proposed as the main sensor, provides the necessary feedback to navigate and simultaneously calibrate the stereocamera parameters (like the camera separation, focal length, camera placement with respect to the robot, etc.). It uses the natural landmarks already present in the environment without additional infrastructures. Some simulation and experimental results are presented in order to explain the proposed method and current status.

1. Introduction

The work presented in this paper has been carried out in the frame of the MALTA project [1] (Multiple Autonomous forklifts for Loading and Transportation Applications, a joint research project between industry and university, funded by the European Regional Development Fund and Robotdalen, in partnership with the Swedish Knowledge Foundation). The project objective is to create fully autonomous forklift trucks for paper reel handling. The result is expected to be of general benefit for industries that use forklift trucks in their material handling through higher operating efficiency and better flexibility with reduced risk for accidents and handling damages than if only manual forklift trucks are used.

In factory automation, automated vehicle navigation is often achieved using specific infrastructure like wires placed on the ground or artificial landmarks placed in different positions of the working environment. The advantage of these approaches is to use simple and efficient systems to localize the robot, but at the same time they require additional effort for preparing the environment. Some examples available on the market are given by laser scanner systems with reflective markers, or video feedback systems using image processing for recognizing different shapes or patterns located in different positions of the working space [2][3]. For overcoming the drawback of "dressing" the environment with artificial landmarks, a different approach is presented in this paper, aiming to use the natural landmarks, directly present in the working space, to localize the mobile robot (the forklift truck).

Typical operative conditions of the forklift trucks are shown in Fig 1. As it is possible to see, the lighting placed in the ceiling as well as the paper reel can be used as natural landmarks, already existing in the working area.

Figure 1. Typical forklift material handling environment into the industrial storage building.

The present work focuses on using Visual SLAM techniques to localize and autonomously navigate in the factory environment. The experimental set-up is made of a stereo camera system with an FPGA, for performing real time feature extraction using Stephen and Harris combined corner and edge detector [4].

2. The navigation strategy

The navigation strategy presented consists of two phases. In the first phase the algorithm works as an optimization algorithm for estimating the extrinsic parameters of the two camera systems. Once the calibration process reaches a stable state, an EKF (Extended Kalman Filter) based probabilistic estimation is performed. In this second phase of the simultaneous localization and mapping of the robot, also the camera

calibration parameters are refined, as indicated by the long arrows shown in the simplified block diagram describing the navigation strategy in Fig.2.

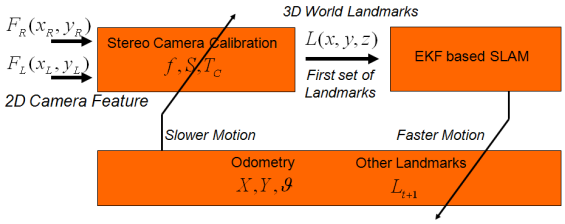

Figure 2. Navigation and auto-calibration strategy block diagram.

The identification of the first landmarks is strictly related to the initial camera view, and will be used by the SLAM algorithm as the main reference frame for the global localization of the robot, and landmarks positioning in the map of the environment.

3. VSLAM theory and simulations

In the first stage, the information coming from the Stephen and Harris feature detection algorithm are combined with the odometry information related to a slower motion in order to use only the relative motion interval ΔT not affected by the incremental error from the odometry. The global robot positioning information is not considered at this stage, but will be estimated in the SLAM phase.

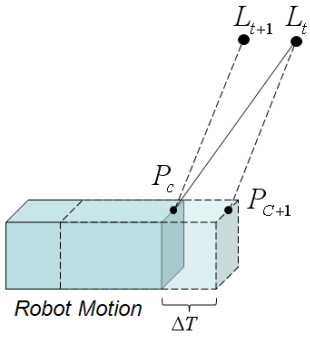

Figure 3. Relative motion of the landmark with respect to the robot.

After a robot movement, fully described by a planar rotation and translation, $\Delta\vartheta$ and ΔT respectively, the corresponding landmarks positions with respect to the robot camera P_C are depicted in Fig 3. The relative positions with respect to two subsequent instants of time are described by the relation (1).

$$L_t = Rot_Z(\Delta\vartheta)\cdot L_{t+1} + \Delta T \qquad (1)$$

Under the hypothesis of a known motion interval $\Delta\vartheta$, ΔT, the actual landmark position L_{t+1} can be predicted from the previous landmark position L_t. The difference from the prediction and the actual landmark measurement coming from the stereo camera, indicated by E_L , should be zero under the hypothesis of a calibrated system. Otherwise, the stereo camera parameters C (like focal length f, cameras separation S, or camera position P_C with respect to the robot reference frame) can be estimated using the difference E_L and the jacobian matrix J, as described in the relations (2).

$$L = L(C, F_L, F_R)\ ,\ J = \frac{\partial L}{\partial C}$$
$$dC = \left(J^T J + kI\right)^{-1} J^T E_L \qquad (2)$$

The jacobian matrix J represents the partial derivative of the landmarks position L with respect to the camera parameters C. Newton's optimization method computes the parameters correction vector in order to minimize the error E_L in the landmarks space. Simulation results are presented in Fig 4, where the focal length f and the camera separation S have been estimated based on visual tracking of the five landmarks shown on bottom-right.

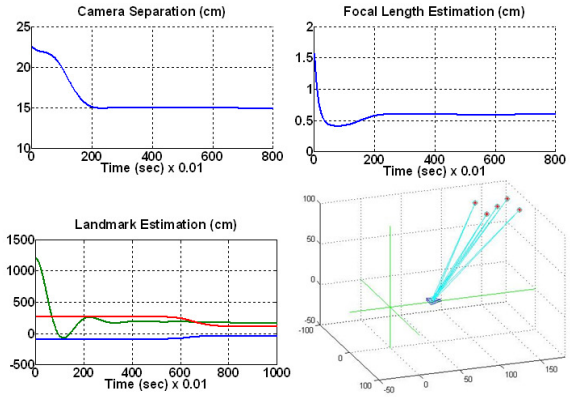

Figure 4. Focal length, camera separation, and central landmark position estimation.

Since the method in general suffers from local mimima problems, the estimation accuracy is strictly dependent on the robot path that should be chosen in order to propose different positions of the robot with respect to the landmarks.

In the second phase of the navigation algorithm, an EKF based probabilistic method has been implemented for simultaneously estimating the camera parameters and the robot landmarks respective positions. The state variables to be estimated are *3+3N+C*, that corresponds to the robot position and orientation (*3* dofs), three dimensional coordinates of *N* landmarks in the environment, and camera parameters *C*, constituting the state vector as shown in (3).

$$x(k) = [X, Y, \vartheta, X_{L1}, Y_{L1}, Z_{L1} \cdots, X_{LN}, Y_{LN}, Z_{LN}, S, \cdots, f]$$
$$u(k) = [V_X, V_Y, V_\vartheta]$$
$$y(k) = [F_{R1}, F_{L1}, \cdots, F_{RN}, F_{LN}]^T \qquad (3)$$

The inputs to the system are the robot velocity for both the position and orientation, whereas the outputs are *4N*

feature coordinates on the right and left camera sensors. The model of the system is computed as shown in the relations (4), constituting the *predict phase* of the algorithm.

$$x(k+1) = F(k)x(k) + G(k)u(k) + v(k)$$
$$y(k) = h(x(k),k) + w(k) \quad (4)$$

The state equations are still linear with respect to the robot velocities, since the robot kinematics has been solved a part. Besides, the output of the model is non linear, and represents the core of the estimator. The state matrix *F(k)* provides the robot position and orientation, computing the corresponding state variables from the input velocities. On the other hand, the landmarks' positions and the camera parameters have a zero dynamic behavior.

$$G(k) = \begin{bmatrix} \underset{3\times1}{I} \\ \underset{N\times1}{I} \\ \underset{C\times1}{I} \end{bmatrix} \quad F(k) = \begin{bmatrix} \underset{3\times3}{I} & \underset{3N\times3N}{0} & 0 \\ 0 & I & \underset{C\times C}{0} \\ 0 & 0 & I \end{bmatrix} \quad (5)$$

During the *update phase* of the EKF, the state variables, and the related covariance matrix *P*, are updated by the correction from the Kalman gain *R* and the innovation vector *e*, as reported by the relations (6).

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + R \cdot e$$
$$P(k+1|k+1) = P(k+1|k) - RH(k+1)P(k+1|k) \quad (6)$$

The innovation vector represents the difference between the estimated model output *h* and the real measurements from the stereo camera sensors.

$$e = y(k+1) - h(x(k+1|k), k+1)$$
$$R = P(k+1|k)H(k+1)^T S^{-1} \quad (7)$$
$$S = H(k+1)P(k+1|k)H(k+1)^T + W(k+1)$$

The computation of the Kalman gain *R*, comes from the linearization of the output model around the current state estimation, through the corresponding jacobian matrix *H*, as presented in (8).

$$P = \begin{bmatrix} \underset{3\times3}{\sigma^2_R} & 0 & 0 \\ 0 & \underset{3N\times3N}{\sigma^2_F} & 0 \\ 0 & 0 & \underset{C\times C}{\sigma^2_C} \end{bmatrix} \quad H(k+1) = \left.\frac{\partial h}{dx}\right|_{x=\hat{x}(k+1|k)} \quad (8)$$

$$\begin{matrix} & & & & & & & & & & & dx \\ & & & & & & & & & & & (3+3N+C)x1 \\ & & & H:\ 4N\ x\ (3+3N+C) & & & & & & & & \\ 4 & \frac{\partial F_1}{\partial X_R} & \frac{\partial F_1}{\partial Y_R} & \frac{\partial F_1}{\partial \vartheta_R} & \frac{\partial F_1}{\partial L_1} & 0 & 0 & 0 & \frac{\partial F_1}{\partial f} & \cdots & \frac{\partial F_1}{\partial S} & dX \\ & & & & & & & & & & & dY \\ & & & & & & & & & & & d\vartheta \\ 4 & \frac{\partial F_2}{\partial X_R} & \frac{\partial F_2}{\partial Y_R} & \frac{\partial F_2}{\partial \vartheta_R} & 0 & \frac{\partial F_2}{\partial L_2} & 0 & 0 & \frac{\partial F_2}{\partial f} & \cdots & \frac{\partial F_2}{\partial S} & dL_1 \\ & & & & & & & & & & & dL_2 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 & \ddots & 0 & \vdots & \cdots & \vdots & \vdots \\ & & & & & & & & & & & dL_N \\ 4 & \frac{\partial F_N}{\partial X_R} & \frac{\partial F_N}{\partial Y_R} & \frac{\partial F_N}{\partial \vartheta_R} & 0 & 0 & 0 & \frac{\partial F_N}{\partial L_N} & \frac{\partial F_N}{\partial f} & \cdots & \frac{\partial F_N}{\partial S} & df \\ & & & & & & & & & & & \vdots \\ & 1 & 1 & 1 & 3 & 3 & \cdots & 3 & 1 & \cdots & 1 & dS \end{matrix} \cdot = \begin{matrix} dy \\ 4Nx1 \\ dF_1 \\ dF_2 \\ \vdots \\ dF_N \end{matrix}$$

The three groups of parameters to be estimated are quite evident by the structure of the *H* matrix, where the central part is block diagonal indicating the feature-landmark correspondences. Also the predicted state covariance matrix *P* is block diagonal, symmetric and positive definite, containing the predicted variances of the state elements.

Simulation results are reported in Fig 5, where the camera separation *S* and the focal length *f* have been estimated together with landmark positions with respect to the robot along a square path.

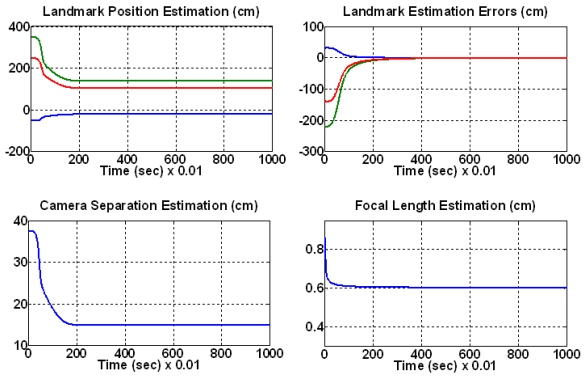

Figure 5. Camera separation estimation together with the landmarks positions with respect to the robot.

4. Experimental Set Up and results

The experimental platform is made up of the "two camera-system" mounted on an FPGA support-board for real-time feature extraction as shown in Fig.6. Five LED landmarks have been used to simulate the ceiling lamps in the storage building.

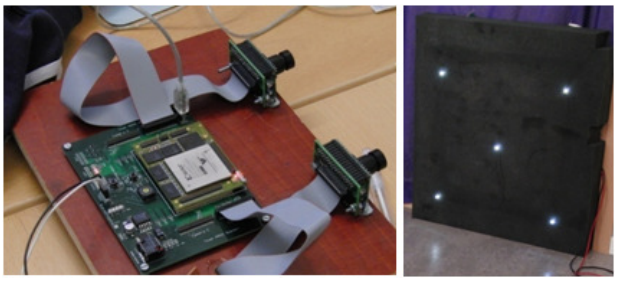

Figure 6. The two camera system used during the experiments, directly connected to the FPGA board.

The Stephen and Harris combined corner and edge detector, implemented on the FPGA, provides the feature extraction for both the right and left camera views, as shown in Fig.7. The nearest neighbor algorithm has been used to track the features between two subsequent frames, solving the data association problem with the landmarks.

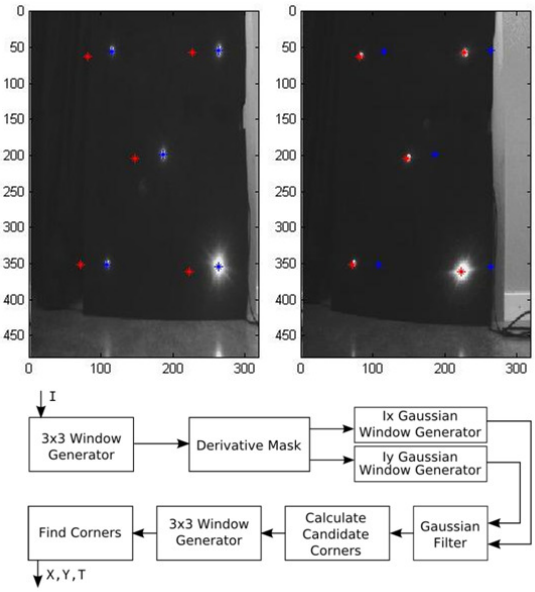

Figure 7. Block diagram of the VHDL implementation of the Stephen and Harris corner detector, and the feature extraction for the left (blue) and the right (red) cameras.

The camera motion with respect to the landmarks has been performed in a straight path along the *X* axis. The landmarks' relative motions with respect to the robot have been computed through a proper calibration and triangulation method [5]. On the other hand, their relative positions with respect to the camera have been computed according to the relation (1), and shown in Fig.8.

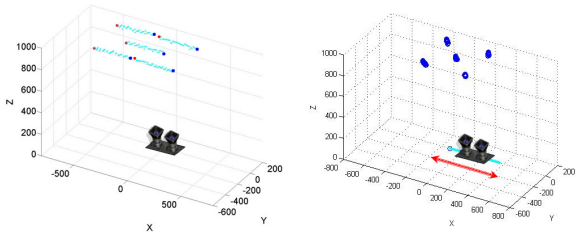

Figure 8. Landmarks 3D reconstruction with respect to the robot (left) and to the world reference frame (right).

The localization and mapping algorithm has been implemented using the odometry data for the *predict* phase, and the stereo vision feedback for the *update* phase. The implementation follows the description explained in the previous section, but the video feedback has also been considered directly in the three dimensional space. The state vector is made out of 19 elements, (having one camera parameter *C=1*, and five landmarks *N=5*), representing the three robot DoFs, the 5 three dimensional coordinates of the landmarks, and the camera separation *S*. Some experimental results are shown in Fig.9 in which the five landmarks' locations are estimated simultaneously with the robot motion back and forth along the *X* axis, and the camera separation. The position estimation of the central landmark is presented in the upper part of Fig.9 together with the error with respect to the real three-dimensional coordinates. The algorithm errors, with respect to the sensor feedback (representing the innovation vector *e* as described in (7)), are also reported in both the three dimensional space and in the pixels space.

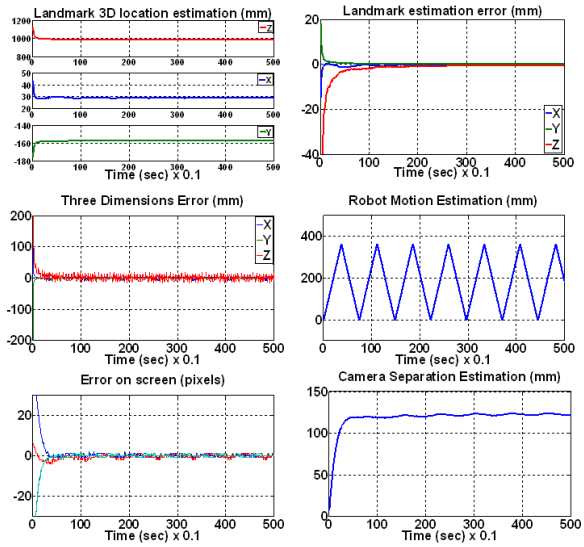

Figure 9. Experimental results related to the landmarks, robot motion, and camera separation estimation.

5. Conclusions and future works

The presented work describes an EKF-based method for performing SLAM using stereovision. The presented method has been validated in simulation and through a simplified robot motion along a straight path, obtaining an error within 5 mm in estimating the landmarks' position, and 2 mm in estimating the camera separation. More experiments will be performed using more complex paths, and also estimating other camera parameters like focal length and camera position and orientation with respect to the robot.

References

[1] http://aass.oru.se/Research/Learning/malta/index.html

[2] www.danahermotion.com : Automated Guided Vehicles

[3] www.sky-trax.com

[4] C. G. Harris and M. Stephens, "A combined corner and edge detector," In *Proc. 4th Alvey Vision Conf.*, Manchester, pages 147-151, 1988.

[5] J. Lidholm, F. Ekstrand, and L. Asplund, "*Two camera system for robot applications; navigation*", In IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2008, pages 345–352.