An Embedded Stereo Vision Module for 6D Pose Estimation and Mapping

G. Spampinato, J. Lidholm, C. Ahlberg, F. Ekstrand, M. Ekström and L. Asplund

***Abstract*—This paper presents an embedded vision system based on reconfigurable hardware (FPGA) and two CMOS cameras to perform stereo image processing and 3D mapping for autonomous navigation. We propose an EKF based visual SLAM and sparse feature detectors to achieve 6D localization of the vehicle in non flat scenarios. The system can operate regardless of the odometry information from the vehicle since visual odometry is used. As a result, the final system is compact and easy to install and configure.**

I. Introduction

AUTOMATIC guided vehicles, commonly known as AGVs, are able to drive autonomously while transporting materials and goods, and are present on the market since the middle of the 20th century. They are both used in indoor and outdoor environments for industrial as well as for service applications for improving the production efficiency and reducing the staff costs.

In field robotics, fully autonomous vehicles are of great interest and still a challenge for researchers and industrial entrepreneurs. Although the large amount of automatic moving platforms already present on the market, almost no one is able to perform automatic navigation in dynamic environments without predefined information. In indoor environments, traditional AGVs typically rely on magnetic wires placed on the ground or other kind of additional infrastructures, like active inductive elements and reflective markers, located in strategic positions of the working area. Such techniques are mostly used in industrial sites, (Danaher Motion, Corecon, Omnitech robotic, Egemin Automation), and in the public service sectors like hospitals (TransCar AGV by Swisslog, ALTIS by FMC and MLR). These systems mainly rely on bi-dimensional views from conventional laser based sensors and sometimes need pre-defined maps. As a consequence they show very low flexibility to environment changes. Furthermore, 2D environmental representations typically provided by laser scanners cannot capture the complexity of unstructured environments, especially in not flat scenarios. To overcome this limitation, often conventional planar scanners are coupled and synchronized with rotational mechanisms but a new-class of 3D ranging sensors has been recently developed. In this class there are high speed 3D laser sensors like the Flash LADAR (LAser Detection And Ranging) SwissRanger SR3000 and the high resolution range finder VelodyneHDL-64E S2 able to deliver a data rate of up to 1.8 million distance measurements per second. The drawback of these sensors is that they use costly and complex equipments.

On the other hand, vision is broadly recognized as the most versatile sensor in non controlled situations with high potential to drastically reduce the costs. At present, modern industrial vision systems are equipped with fast image processing algorithms and highly descriptive feature detectors that provide impressive performances in highly controlled situations. Some solutions available on the market use image processing for recognizing different unique planar patterns placed in strategic positions of the environment (Sky-trax). The obvious drawback is the additional effort required to "dress" the working space with artificial landmarks not related to the production lines. Sometimes it is also impossible to modify the environmental setting due to the highly dynamic conditions in the production operations.

To overcome these drawbacks, a more versatile vision system is required using only pre-existing information from the working setting ("natural landmarks"). This concept requires a new paradigm for the traditional image processing approach that shifts the attention from the two dimensions to the more complete and emerging 3D vision. In this paper we propose a low cost solution for industrial AGVs based on one high speed stereo camera, able to cope with natural landmarks in non flat scenarios and dynamic auto re-calibration of the extrinsic parameters. The system has been designed to be more versatile and cost efficient with respect to existing solutions available on the market.

II. The Stereo Vision Module

The stereo vision system is made of two 5-megapixel CMOS digital image sensors from Micron (MT9P031) and a Spartan-3A-DSP FPGA with 1800K system gates, 84 block RAM (18KB each) and 84 DSP 48A blocks. The board can hold up to seven different configurations for the FPGA stored in Flash, enabling seven different algorithms to be selected at run-time. The FPGA can communicate with the

The work presented has been carried out in the frame of the MALTA project (Multiple Autonomous forklifts for Loading and Transportation Applications), a joint research project between industry and university, funded by the European Regional Development and Robotdalen, in partnership with the Swedish Knowledge Foundation

G. Spampinato, J. Lidholm, C. Ahlberg, F. Ekstrand, M. Ekström and L. Asplund are with the School of Innovation, Design and Engineering, Mälardalen University, Sweden. (e-mail: {giacomo.spampinato, jorgen.lidholm, carl.ahlberg, fredrik.ekstrand, mikael.ekstrom, lars.asplund} @ mdh.se).

external systems over USB, and Ethernet. On the board, there are also a 32 MB SDRAM and 16 MB non volatile Flash. The system architecture is shown in Fig.1. The final configuration of the stereo system includes also a Qseven module featuring an Intel Atom Z530 CPU (not used for the work presented in this paper). All is packed in a compact aluminium box easy to install and configure through the USB connection [2]. The lenses adopted are fisheye lenses with 2.1mm focal length from Mini-Objektiv, with F2.0 aperture and 100 degrees field of view.

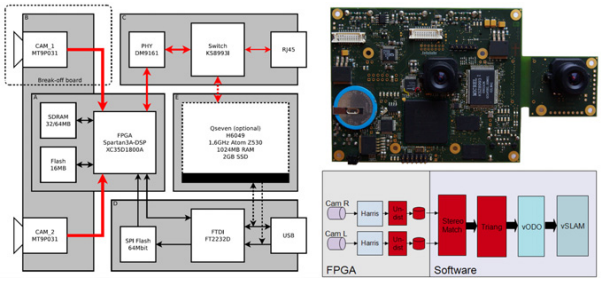

Fig. 1. The stereo vision system and the algorithm blocks diagram.

The calibration procedure has been performed using the camera calibration toolbox for Matlab®. The intrinsic parameters identified include the lens distortion map, the principle point coordinates for the two sensors and the focal lengths in pixel units. For simplicity, the lens distortion has been assumed to be radial, identified by a sixth order polynomial containing only even exponential terms. It is worth to note that the original image resolution 640x480 has been expanded by a factor of 1.6 (1024x768) in order to compensate for the image expansion due to the radial distortion correction, and use all the visual information acquired. The principle point and all pixels have been rescaled according to the new image resolution. The extrinsic parameters are computed through the essential matrix factorization as in [7].

The Harris and Stephens combined corner and edge detection algorithm has been implemented in hardware on the FPGA working in real-time. To gain real-time speed of the system, the algorithm is designed as a pipeline, so each step executes in parallel. (Three different window generators are used for the derivative, factorization, and comparison masks of 5x5 pixels size).

From the autocorrelation mask M and its convolution with the Gaussian kernel G two methods for extracting the "cornerness" value R against a fixed threshold are universally accepted by the research community: the original method from Harris [3] and the variation proposed by Noble [4] in order to avoid the heuristic choice of the k value (commonly fixed to 0.04 as suggested by Harris). The choice of the two methods are rather equivalent and both effective for the case analyzed in our proposed applications. The main difference is the dynamic threshold that has to be three magnitude orders more in the Harris case than for the Noble one. This is due to the division in the Noble case that keeps the "cornerness" lower. In our implementation we decided to implement the original method by Harris since the division implementation in the FPGA would have required a lot more resources. Currently, the frame rate achieved is around 40 fps, even if the actual usable speed is 14 fps due to the bottleneck of the USB communication.

To reduce the computational load, the matching of the interest points in the different cameras has been implemented using binary images. The advantage is to simplify the cross correlation implementation in the FPGA by reducing the amount of information. The binary images are compared with the XOR bitwise operator instead of the binary multiplication, as shown in (1).

$$Corr = \sum_{r=1}^{R}\sum_{c=1}^{C} \frac{not\{XOR[IL(p_{rc}), IR(p_{rc})]\}}{R \cdot C} \qquad (1)$$

As well known from the multiple view projective geometry theory, for each feature extracted from the left image, the corresponding point in the right image lies on the corresponding epipolar line on the right image whose analytical coefficients are easily extracted from the fundamental matrix F. A proper search window has to be defined in order to apply the correlation function (1) to the possible corresponding candidates along the epipolar line. The search window is defined to be large enough to cover the maximum disparity at the investigated depth of view. The search window is heuristically defined and strongly depends on the interested depth and the camera vergence. The proposed system has theoretically zero vergence, so that the corresponding feature in the right camera always lies on the left side along the epipolar line with respect to the left feature coordinates (this is not the case of stereo cameras with non zero vergence). In Fig. 2 an example of the described technique is shown. The feature in the left image defines the epipolar line in the right image, as well as the related search window along the epipolar line.

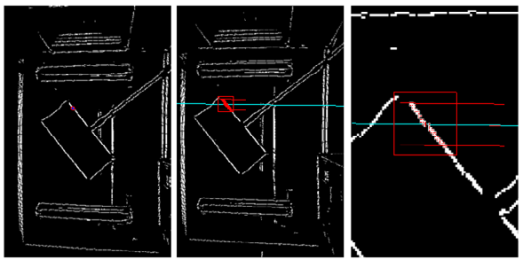

Fig. 2. Stereo matching using the epipolar constraint and the correlation on the right candidate matches. The search window is represented by the two red lines under and above the epipolar line blue, and the contained candidates within the window are marked in red. Also the correlation window is shown around the correspondent feature indicated by the red square.

III. Stereo Triangulation and Depth Error Modeling

The stereo triangulation projects the interest points collected from the two images in the 3D space. Unfortunately the triangulation is affected by heteroscedastic error (non homogeneous and non isotropic) as described in [5]. An error analysis has been performed to provide an uncertainty modeling of the stereo system. The

uncertainties in the 3D reconstruction directly affect the covariance matrices used in the localization and mapping algorithm that is based on probabilistic estimation. A 3D error modelling is derived assuming that the feature points are affected by a uniform distributed Gaussian noise, represented by two uncertainty circles in the left and in the right images. The corresponding 3D uncertainty is obtained by the solid intersection of the two projective cones as shown in Fig.3.

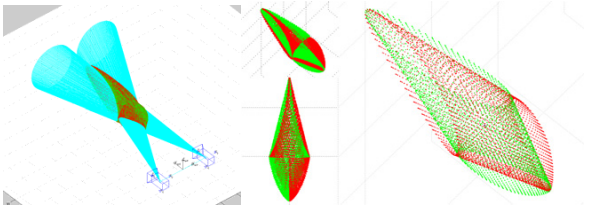

Fig. 3. The diamond shaped 3D uncertainty produced by a circular uncertainty in both the left and right images.

Since the projective lines could be skew lines in absence of epipolar constraints a general model for the stereo triangulation has been adopted. The triangulation makes use of a least square solution to minimize the re-projection error in both images. The initial hypothesis comes from the extrinsic parameters R and T that relates the two cameras $P_R = R \cdot P_L + T$, that can be rewritten as $P_{ZR} \cdot F_R = R \cdot P_{ZL} \cdot F_L + T$, using the projective transformations for each camera.

$$F = \begin{bmatrix} F_x & F_y & 1 \end{bmatrix}^T = \begin{bmatrix} \frac{x}{f} & \frac{y}{f} & 1 \end{bmatrix}^T = \begin{bmatrix} \frac{P_x}{P_z} & \frac{P_y}{P_z} & 1 \end{bmatrix}^T \quad (2)$$

Using the matrix formulation the problem can be rewritten.

$$\begin{bmatrix} F_R & -R \cdot F_L \end{bmatrix} \cdot \begin{bmatrix} P_{ZR} \\ P_{ZL} \end{bmatrix} = T \quad (3)$$

Posing $A = \begin{bmatrix} F_R & -R \cdot F_L \end{bmatrix}$ and solving using the LSM, the 3D point P can be computed both in the left and right reference frames.

$$\begin{bmatrix} P_{ZR} \\ P_{ZL} \end{bmatrix} = \left(A^T \cdot A\right)^{-1} \cdot A^T \cdot T \qquad \begin{aligned} P_R &= F_R \cdot P_{ZR} \\ P_L &= F_L \cdot P_{ZL} \end{aligned} \quad (4)$$

To make the analysis of the triangulation accuracy, the analytical relation between the uncertainty in the image space and the related uncertainty in the 3D space has been defined. The partial derivatives of the reconstructed landmarks with respect to the feature points in the two images have been computed through the jacobian matrix J_{PS} shown in (5). The error propagation in the 3D space ΔP is related to a given uncertainty ΔR and ΔL in X and Y in both images.

$$J_{PS} = \frac{\partial P}{\partial S} = \begin{bmatrix} \frac{\partial P_X}{\partial L_X} & \frac{\partial P_X}{\partial L_Y} & \frac{\partial P_X}{\partial R_X} & \frac{\partial P_X}{\partial R_Y} \\ \frac{\partial P_Y}{\partial L_X} & \frac{\partial P_Y}{\partial L_Y} & \frac{\partial P_Y}{\partial R_X} & \frac{\partial P_Y}{\partial R_Y} \\ \frac{\partial P_Z}{\partial L_X} & \frac{\partial P_Z}{\partial L_Y} & \frac{\partial P_Z}{\partial R_X} & \frac{\partial P_Z}{\partial R_Y} \end{bmatrix} \quad \Delta P = \begin{bmatrix} \Delta P_X \\ \Delta P_Y \\ \Delta P_Z \end{bmatrix} = J_{PS} \cdot \begin{bmatrix} \Delta L_X \\ \Delta L_Y \\ \Delta R_X \\ \Delta R_Y \end{bmatrix} \quad (5)$$

In Fig. 4 the 3D distribution of the uncertainty along the principle axis of the 3D diamond shape is presented, showing the heteroscedastic behaviour of the stereo model adopted. Indeed they have different behaviours in the three dimensions (non isotropic) and non linear along each axis (non homogeneous).

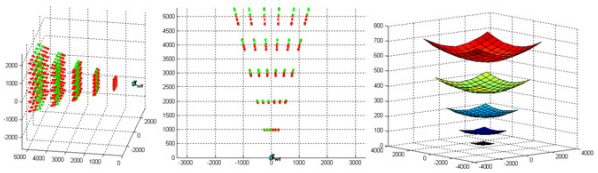

Fig. 4. The 3D uncertainty of the major axis of the ellipsoid related to a grid pattern analyzed at different depths from the cameras.

The known grid pattern, shown in the upper part of Fig. 4, has been used to measure the triangulation error under the hypothesis of three pixels uncertainty in the image space re-projection. For the stereo system adopted, the 3D reconstruction mostly suffers of uncertainty along the principle axis of the 3D diamond, that is, along the projective line connecting the observed landmark with the centre of the stereo rig. Extending the reference plane to arbitrary heights, so that the image projections remain unchanged, the average uncertainty in the three dimensions has been reported in Fig. 5 for distances to the stereo rig from 1 to 30 meters, showing the non linear behaviour as expected. The distribution of the error in the three directions is also presented in the left-most pictures for the specific depth of 3 meters.

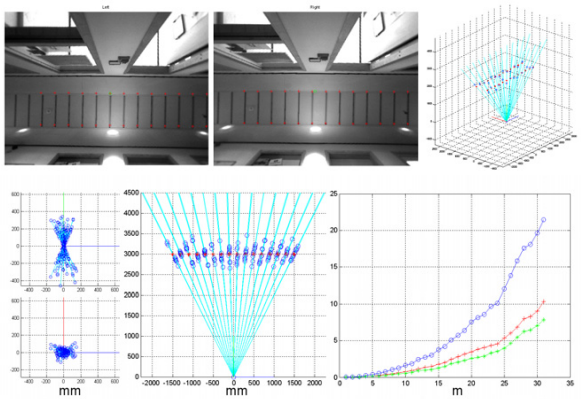

Fig. 5. Top : The reference pattern used to analyse the triangulation error at 3 m distance from the ceiling. Bottom: The distribution of the error along the three dimensions for a fixed depth of view of 3 m is also shown together to the non linear behaviour of the average errors increasing the depth from 1 to 30 m.

The covariance matrix w associated with the 3D uncertainty of each reconstructed landmark can be computed using the error propagation model as in [6] through the jacobian J, where w_L, w_R are the image covariance matrices for the two images.

$$w = J \cdot \begin{bmatrix} w_L & \\ & w_R \end{bmatrix} \cdot J^T \quad (6)$$

IV. Visual SLAM

The SLAM algorithm has been implemented using an

Extended Kalman Filter (EKF) able to estimate the complete 6D pose of the vehicle. It is based on the visual information coming from the stereo-camera, and the quaternion algebra has been adopted to efficiently deal with the 3D orientations. During an initial calibration phase, the odometry coming from the vehicle is used to simultaneously estimate the camera parameters together with some specific landmark positions, as detailed in [2]. The state variables to be estimated are *7+3N+C*, corresponding to the robot position and orientation (3 DoFs for the position and 4 for the normalized quaternion *Q*), three dimensional coordinates of *N* landmarks in the environment, and camera parameters *C*, constituting the state vector. The inputs to the system are the robot velocities for both position and orientation, whereas the outputs are *4N* feature coordinates for the right and left cameras.

$$x(k) = [X, Y, Z, Q, X_{L1}, Y_{L1}, Z_{L1} \cdots, X_{LN}, Y_{LN}, Z_{LN}, C]$$
$$u(k) = [V_X, V_Y, V_\vartheta] \quad (7)$$
$$y(k) = [F_{R1}, F_{L1}, \cdots, F_{RN}, F_{LN}]^T$$

The non linear function *f,* representing the system non linear model $x(k+1) = f(x(k), u(k), k) + v(k)$ according to the EKF theory imposes a zero dynamics for the landmarks, and a state propagation in the *predict phase* as presented in (8) where *Rq* indicates the Rodrigues tensor associated to *q*.

$$x(k+1) = \begin{bmatrix} fp_{K+1} \\ fq_{K+1} \end{bmatrix} = \begin{bmatrix} fp_K + R_{q_K} \cdot \Delta_P \\ fq_K \cdot \Delta_Q \end{bmatrix}$$
$$P(k+1) = G_v(k) \cdot P(k) \cdot G_v(k)^T + G_u(k) \cdot v(k) \cdot G_u(k)^T \quad (8)$$
$$G_v(k) = \left.\frac{\partial f}{dx}\right|_{x=\hat{x}(k)} = \begin{bmatrix} I & \frac{\partial f_p}{\partial q_K} \\ 0 & \frac{\partial f_q}{\partial q_K} \end{bmatrix} \quad G_u(k) = \left.\frac{\partial f}{du}\right|_{x=\hat{x}(k)} = \begin{bmatrix} R_{q_K} & 0 \\ 0 & \frac{\partial fq}{\partial \Delta q} \end{bmatrix}$$

The system model and the measurement uncertainties are respectively indicated by a 7x7 and 4x4 diagonal matrices *v* and *w* containing the variances terms. The model uncertainty are computed based on the specific kinematics involved, whereas the uncertainties in the measurements are based each time on the depth involved as reported in the previous section.

To keep the whole system simple to use and easy to maintain, more effort has been devoted to avoid to read the odometry data from the vehicle. At the same time, the localization algorithm results more robust to uncertainties that easily arise in the vehicle kinematic model. After the calibration phase, the calibrated stereo rig is used to estimate the vehicle motion using visual odometry. Back-projecting the features coordinates in the image space to the 3D space using triangulation, the problem is formalized as an *orthogonal Procrustes* problem [1] in estimating the rotation and translation terms that minimize the functional (9) between two subsequent camera views, using the SVD.

$$\sum_{i=1}^{n} \left\| P_{t,i} - T - R \cdot P_{t,i+1} \right\|^2 \quad (9)$$

The complete computation is reported in [2]. The robustness of this method is strongly related to the accuracy in the reconstruction and on the cameras resolution related to the investigated landmarks depths. In the current implementation, for object farther than 10 meters from the cameras, a low pass filter has been applied to smooth the reconstruction noise. In some cases, a zero dynamics on the *predict phase* of the Kalman filter is preferable to a noisy visual odometry. In these cases, higher values of the Gaussian variances have been used to model the uncertainties reflecting the 3D reconstruction error modeling.

V. Experimental Results

A. Small scale 6D vSLAM

A 6D pose estimation of the stereo rig with respect to a generic 3D landmarks configuration is shown in the first experiment, performed holding the camera by hand and moving it freely in all 6 DoFs in front of a test pattern shown in Fig. 6. The test pattern is composed by a five LED deformable "non plane" structure, that allows the complete 6D vSLAM without the aid of any homographic information.

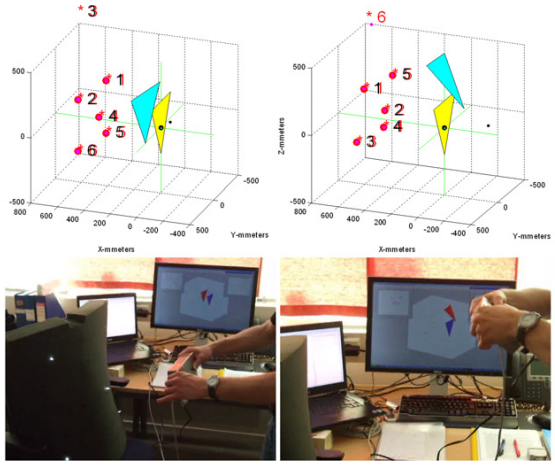

Fig. 6. The reference pattern used to test the vSLAM algorithm moving the stereo camera by hand in different positions and orientations. The cyan marker indicates the camera pose, whereas the yellow one shows the starting pose.

The landmarks covariance ellipses converge to an uncertainty within 1 cm after about 30 cycles, and so also the 6D pose estimation of the camera. The landmarks 3D reconstruction and the camera trajectory are shown in Fig 7.

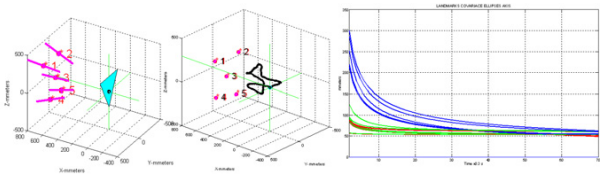

Fig. 7 From left to right: the principal axis of the covariance ellipses at the beginning of the test, and the final map after 150 cycles. The SLAM algorithm converges in about 30 cycles to 1 cm accuracy.

The full 6D pose estimation has been tested on a mobile

platform to reconstruct the executed 6D path in not flat indoor environment using the S-shaped slope shown in Fig 8. The system has been mounted on a wheeled table and was driven manually form the station A to the station B up to the slope.

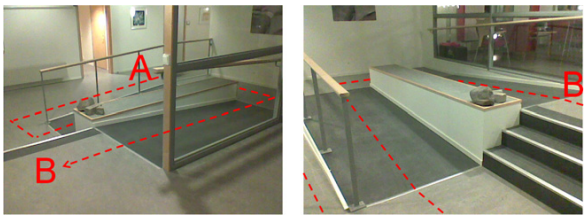

Fig. 8. S-shaped non planar path used to test the vSLAM algorithm. The path total length is 12 m, three 4m straight lines and three 1m diameter curves, covering 1 m difference in height.

The system was pointing upwards while driven along the assigned path, extracting the interest features form the lamps in the ceiling. The vSLAM algorithm has been executed minimizing the re-projection errors of the estimated landmarks positions in the 3D space and estimating the 6D pose of the platform along the path. The platform was driven with a speed of 0.5 m/s and the corresponding precision of the pose estimation was within 5 cm in the planar coordinates X and Y, and 10 cm in height coordinate Z. The same accuracy has been registered in the landmarks estimation. The accuracy of the orientation estimation was within 10 degrees depending on the number of landmarks seen at the same. The accuracy of the vSLAM reconstruction gets worst if three or less lamps are seen at the same time. Fig 9 reports the reconstructed path and the convergence of the estimated landmarks positions with the related covariance.

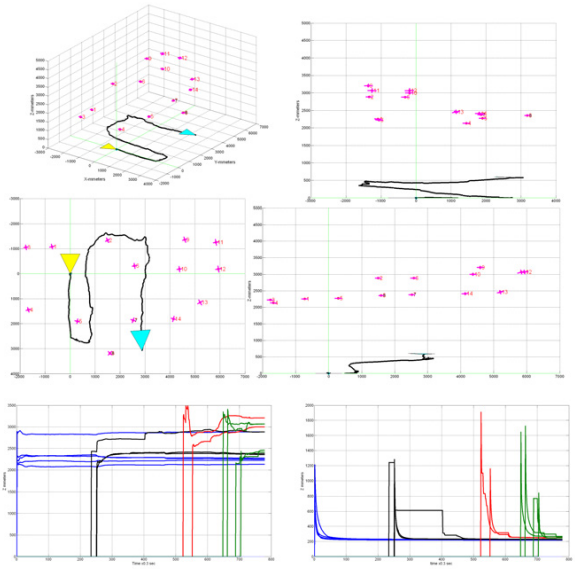

Fig. 9. Vehicle path reconstruction and landmarks estimation. From top to bottom: 3D, lateral, top and front views of the estimated path. Convergence of the Z-position and covariance ellipses principal axis associated to the reconstructed landmarks.

B. *Large scale vSLAM*

To test the vSLAM in a bigger scale, some experiments have been performed inside the university hall. In the working platform at Mälardalen University, the stereo system has been placed on a passive wheeled table. The vision system looks upwards, extracts information from the lamps in the ceiling, builds a map of the environment and localizes itself inside the map with a precision within the range of 1-5 cm depending on the height of the ceiling. Two different tests are reported here. The first case, shown in Fig 10, presents a localization and mapping experiment in which the table was moved at about 1m/s producing the map with 40 landmarks on a surface of about 600 m^2. The landmarks are mainly grouped in two layers (represents in red and blue in Fig 10), respectively 4 and 7 m far from the cameras, reflecting the architectural structure of the entrance hall. The accuracy of the reconstruction is within 5 cm.

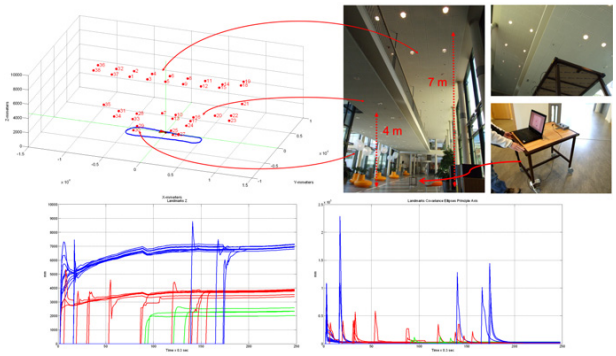

Fig. 10. The 40 landmarks reconstructed from the faculty entrance hall are estimated in 250 cycles moving the platform in a closed path. The convergence of the covariance ellipses (right) and the estimated heights (left) are also reported. The green landmarks are other interest points coming from lower windows.

The second test presented in Fig. 11, shows a smaller scale platform motion area (about 10 m^2), but a higher distance of the farthest landmarks, 16 m. The accuracy of the reconstruction decreases from 5 to 10 cm, revealing the high sensitivity of the system to higher depths of view.

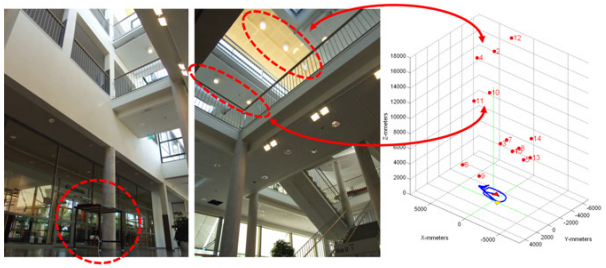

Fig. 11.Second test performed inside the university building, in which four main different layers of 15 landmarks have been identified.

C. *Preliminary tests in industrial site*

During the frame of the MALTA project, some tests have been organized inside the Stora Enso paper mill in Karlstad, Sweden. The vehicle used during the experiments was the H50 forklift provided by Linde Material Handling, and properly modified by Danaher Motion to provide the low level trajectory control. The tests site, as well as the industrial vehicle used, are shown in Fig. 12. The stereovision navigation system has been placed on top of the vehicle, like shown in the picture, making the system integration extremely easy.

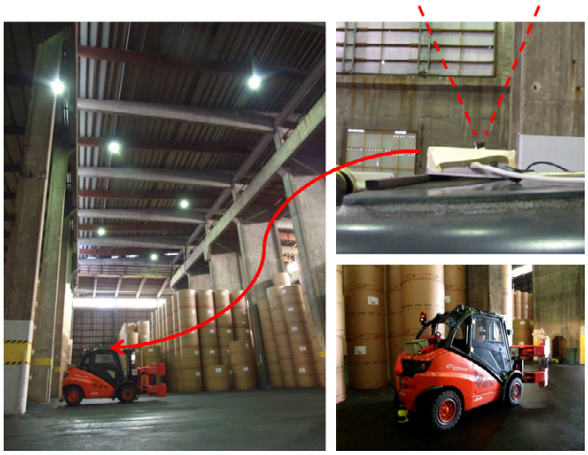

Fig. 12. The industrial set-up inside the Stora Enso paper mill.

The environmental conditions were very different from the university both bigger and darker. The test area was about 2800 m^2 and the height of the ceiling, and so the distance of the lamps from the vehicle (used as natural landmarks), was about 20 m. The tests have been performed, estimating the position of the robot and building the map of the environment. The estimated position and orientation of the vehicle were provided to the Danaher Motion navigation system as "epm" (external positioning measurement). In Fig. 13 the path estimation is reported while the vehicle was performing a cubic b-spline driving with a speed of 0.5 m/s. Also a longer path was performed with the purpose of collecting as many landmarks as possible and build a more complete map. In this case the map employed a total of 14 landmarks on 800 m^2 surface with a precision of about 15 cm at low speed.

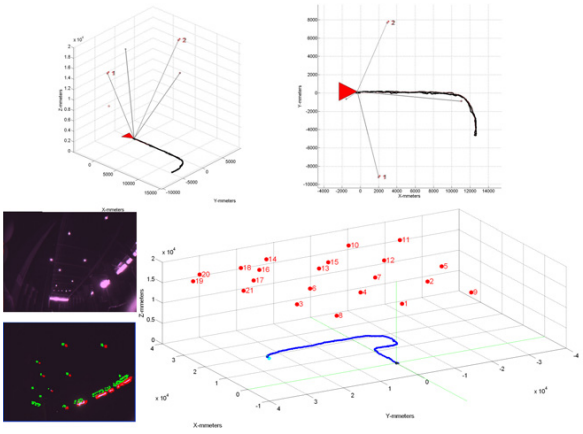

Fig. 13. Top: Vehicle path estimation while performing a b-spline trajectory inside a 12x5 m^2 area. Bottom: Three dimensional representation of the vehicle path and map built inside a 20x40 m^2 area. On the left side, a screenshot of the feature extraction process is shown.

Table 1 reports a summary of the experimental results relating the accuracy of the stereo rig to the landmarks distance and the speed of the moving platform. Both increasing the distance and the vehicle speed worsen the reconstruction accuracy as expected. Increaseing the speed of the vehicle to 1-1.5 m/s, the accuracy decays to 30-40 cm from the desired path in the worst cases.

TABLE 1
SUMMARY OF THE ACCURACY OF THE STEREO RIG

Landmarks Distance	Small scale ≤ 1m	Medium Scale ≤ 10m	Large Scale ≤ 20m
Accuracy at 0.5 m / s	± 1 cm	± 5 cm	± 10-15cm
Accuracy at 1.5 m / s	± 5 cm	± 10 cm	± 30cm

In order to address the target of autonomous navigation at full speed (30 Km/h required in many industrial sites to be comparable with human drives), the core of the vSLAM system (running in software at 3 Hz in the current version presented in this paper) is currently being updated to work at 10 Hz, and speed up the low level communication with the trajectory control loop. The main bottlenecks of the system are represented by the 1Mbps USB communication that is currently being updated to 100Mbps Ethernet, and the vSLAM algorithm implementation that has not jet being parallelized into the FPGA.

VI. CONCLUSION

In this paper we propose a stereovision based vSLAM approach with high potential to be used in large environments and in not flat indoor terrains, due to the full 6 DoFs vehicle pose estimation. The main contribution is related to the real time image processing on the FPGA and the use of vision as the only source of information. To make the system more robust, the accuracy has been analytically formulated and modeled. The choice of the EKF as vSLAM algorithm is motivated by a tradeoff between simplicity of implementation and reliability of the method. Even if the computation of the filter grows exponentially with the number of landmarks, in the specific applications targeted, sparse features clustered identifying lamps in the ceilings rarely produce more than 30-40 landmarks in the final map.

REFERENCES

[1] Eggert, DW; Lorusso, A; Fisher, RB (1997), "Estimating 3-D rigid body transformations: a comparison of four major algorithms", Machine Vision and Applications 9 (5): 272–290, 1997.

[2] G. Spampinato, J. Lidholm, F. Ekstrand, C. Ahlberg, L. Asplund, M. Ekström, "Navigation in a box. Stereovision for industry automation", book chapter within "Advances in Theory and Applications of Stereo Vision" ISBN 978-953-307-516-7, INTECH publisher, January 2011.

[3] C. Harris, M. Stephens, (1988). A combined corner and edge detection. In Proceedings of The Fourth Alvey Vision Conference, pp 147-151, Manchester, UK, 1988.

[4] A. Noble, (1989), Descriptions of Image Surfaces, PhD thesis, Department of Engineering Science, Oxford University.

[5] G. Dubbelman, F. Groen, Bias reduction for stereo based motion estimation with applications to large scale visual odometry. Proc. Of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, Florida, June, 2009.

[6] L. Matthies and S. A. Shafer, "Error modeling in stereo navigation," IEEE Journal of Robotics and Automation, vol. 3, no. 3, pp. 239 - 248, June 1987.

[7] Nister, D. "An efficient solution to the five-point relative pose problem", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 26 Issue 6, page(s) 756 - 770, June 2004.