

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

Real-time kinematics for accurate geolocalization of images in telerobotic applications

Sergio Saponara

Sergio Saponara, "Real-time kinematics for accurate geolocalization of images in telerobotic applications," Proc. SPIE 10670, Real-Time Image and Video Processing 2018, 106700J (14 May 2018); doi: 10.1117/12.2303830

SPIE.

Event: SPIE Commercial + Scientific Sensing and Imaging, 2018, Orlando, Florida, United States

Real-time kinematics for accurate geolocalization of images in telerobotic applications

Sergio Saponara

Dipartimento di Ingegneria dell'Informazione (DII), University of Pisa, via G. Caruso 16, Pisa, Italy

ABSTRACT

The paper discusses a real-time kinematic system for accurate geolocalization of images, acquired through stereoscopic cameras mounted on a robot, particularly a teleoperated machinery. A teleoperated vehicle may be used to explore an unsafe environment and to acquire in real-time stereoscopic images through two cameras mounted on top of it. Each camera has a visible image sensor. For night operation, or in case temperature is an important parameter, each camera can be equipped with both visible and infrared image sensors. One of the main issues for telerobotic is the real-time and accurate geolocalization of the images, where an accuracy of few cm is required. Such value is much better than that provided by GPS (Global Positioning System), which is in the order of few meters. To this aim, a real-time kinematic system is proposed which acquires the GPS signal of the vehicle, plus through an RF channel, the GPS signal of a reference base station, geolocalized with a cm-accuracy. To improve the robustness of the differential GPS system, also the data of an Inertial Measurement Unit are used. Another issue addressed in this paper is the real-time implementation of a stereoscopic image-processing algorithm to recover the 3D structure of the scene. The focus is on the 3D reconstruction of the scene to have the reference trajectory for the actuation done by a robotic arm with a proper end-effector.

Keywords: Real-time machine visions, Real-time kinematics, Tele-robotics, Stereoscopic imaging

1. INTRODUCTION

This work addresses the problem of teleoperated machinery, where some operations (e.g. washing, painting, removing garbage, and so on) in harsh environments can be done in a semiautonomous way to reduce operation time and cost and to increase job safety, since the operator is not exposed to risks. The idea is exploiting a RTK (Real Time Kinematic)-based navigation and positioning system for cm-accurate position control of the autonomous machine. Such electronic system may be installed on the machinery, together with a stereoscopic camera system plus an image-processing algorithm that can be used to recover the 3D structure of the scene.

One of the main issues for telerobotic is its real-time and accurate geolocalization, since some operations may require a control of the position with an accuracy of few cm. This is a challenging task when compared to the accuracy provided by GPS (Global Positioning System), which is in the order of few meters [1-5]. To this aim, an RTK system is proposed, which acquires the GPS signal of the vehicle, plus through an RF channel, the GPS signal of a reference base station, geolocalized with a cm-accuracy. To improve the robustness of the differential GPS system, also the data of an Inertial Measurement Unit (IMU) platform, i.e. accelerators and gyro, installed on-board can be used. For the fusion of RTK and IMU data an Extended Kalman Filter (EKF) is used.

With such a robotized system, a single operator is capable of tracking the position of the machine and is capable of controlling its movements and its operations, so that dangerous and time-consuming manual operations can be avoided.

Besides RTK, another key enabling technology for telerobotic operation of industrial machinery is acquiring and geolocalizing stereoscopic images, which can be transmitted to the teleoperator. This way the human remote operator can perceive the same 3D scene that is seen by the machinery.

The feasibility of the developed system was tested with respect to experiments with real images acquired from an online dataset and real industrial scenarios. To be noted that having a machinery that is moving, and operating under safety constraints, there are stringent requirements in terms of low-latency and high throughput of the image processing computation.

Hereafter, Section 2 presents the system architecture, Section 3 the RTK-based positioning/navigation system, and Section 4 presents the 3D reconstruction scene technique. Conclusions are drawn in Section 5.

2. SYSTEM ARCHITECTURE

Fig. 1 shows the architecture of the proposed system, which involves two main parts: the modeling and control of the actuator, e.g. a mobile manipulator, and that for the image vision and 3D reconstruction scene algorithm. To be noted that we are not considering autonomous driving vehicles or robots, so autonomous driving is out of the scope of this paper which is mainly focusing on semiautonomous (since are teleoperated) operation of industrial machinery. Also the problem of wireless link robustness (i.e. Bit Error Rate in the communication between the teleoperator and the machine as a function of distance and of the communication scenario), as discussed for example in [6], is out of the scope of the paper. According to Fig. 1 the vision system on-board the machine provides the teleoperator in real-time with a 3D reconstruction of the scene, as seen by the machinery.

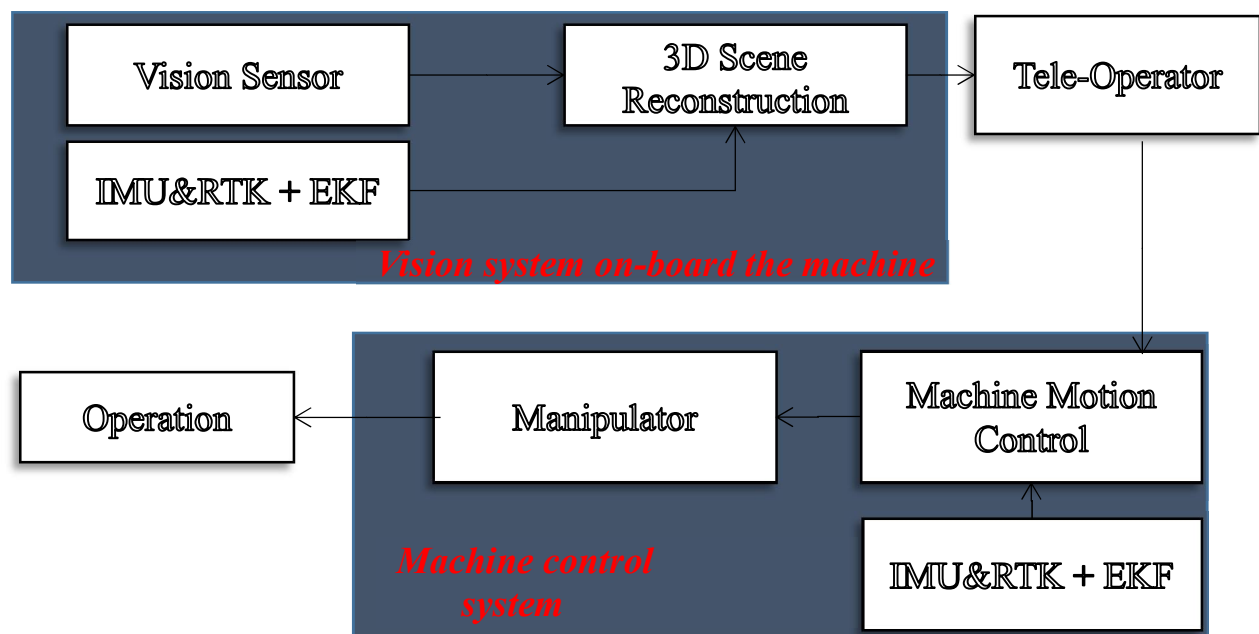


Figure 1: Architecture of the proposed system

Thanks to this information of the environment surrounding the machinery, the teleoperator can decide when and how to force the machinery making an operation. Once this decision is taken by the human teleoperator, then a motion controller on-board the machinery implements the actuation. This is de facto a human-in-the-loop approach. The control on the robotized actuator is divided in two steps:

- The first step is to finding the configuration that minimizes the error in the operative space through a closed loop inverse kinematic algorithm, a technique that is quite know at the state of art [7].
- The second step is to minimize the error in the joint space through the inverse dynamics of the manipulator. The principle of this family of controllers is that the non-linear system is cascaded with its inverse so that the overall system has a constant unity gain. Since in real applications the inverse is not perfect, then a feedback loop is required to deal with errors.

In other words, the manipulator (End-Effector) has to follow a trajectory defined from a planning algorithm: on the base of the position, velocity and attitude of the vehicle, it is defined by means of an inverse kinematic control the desired position of the end-effector minimizing the relative error. The same principle can be applied to different robotic arms.

3. POSITIONING AND NAVIGATION

When making an actuation, the robotic system must know the current position with an accuracy on the scale of few cm. To this aim, it is necessary that the system is equipped with a localization device that exploits both RTK technique, see Fig. 2, and on-board inertial sensors. The localization can be executed by means of a commercial off-the-shelf RTK unit, such as Piksi Multi by Swift Nav [8] which is able to receive GNSS (Global Navigation Satellite Systems) signals with a single antenna on both L1 and L2 GPS bands. However, real-world tests proved that if we rely only on satellite signals, the accuracy below 10 cm can be achieved only if a minimum number of 5 satellites of the constellation are contemporary seen by the base station and by the rover.

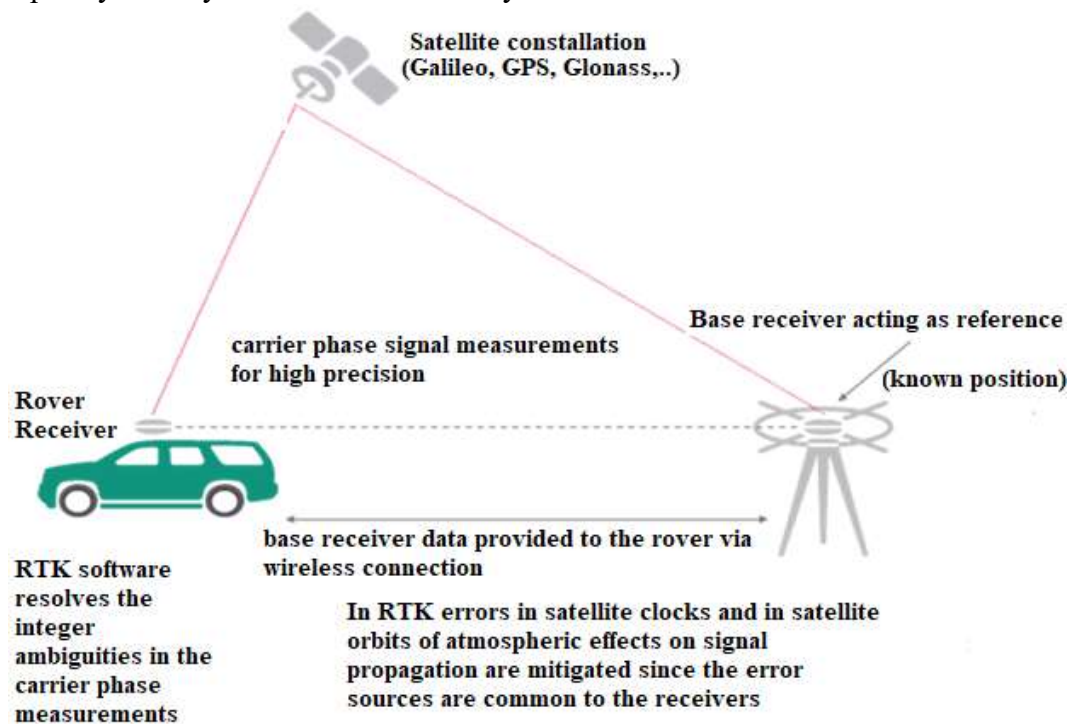


Figure 2: RTK system for cm-accuracy positioning/navigation

In case the machinery is operating in an industrial scenario where buildings may create occlusions, or some operation are done in an indoor environment, the support of an IMU is required. As we discussed in [9], the accuracy of IMU is limited by bias issues of accelerometers and gyroscopes. Due to the double integration process of an IMU-based navigation system, it is difficult to mitigate the effect of the bias errors. Indeed, in an Inertial Navigation Systems (INS) position and orientation are determined from the previous solution by integrating multiple times the acceleration and angle rate data from the IMU. Because of multiple integrations, the solution is extremely susceptible to many sources of error intrinsic to the sensors. Each sensor shows an offset (or bias), with a temperature-dependent component. Bias is given in *mg* for accelerometer, and *degrees/h* for gyroscopes. Due to bias, the solution, even at rest, drifts with time. The error on the estimated position caused by accelerometer and gyroscope bias increases as the second and third power of time, respectively, and grows to unacceptable levels for inertial navigation purposes. For example, Table 1 shows an error estimate after ten seconds for the different IMU grades (i.e. different values of the angular rate and accelerometer bias). The last two rows are those typical of consumer IMU, often realized with MEMS technology. The error is unacceptable. This is why we proposed the system in Fig. 3 where the data from the IMU are used to increase the robustness of RTK in case of occlusions, indoor mitigation and so on, while the accuracy of the RTK unit is used to correct the effect of accelerometers and gyro bias. Error filtering is done through an EKF.

Table 1: Ten-second position errors due to sensor bias

Acceleration bias, mg	Angular rate bias, deg/hr	10-s position error due to acceleration bias	10-s position error due to angular rate bias
$10^{-3} - 10^{-2}$	$10^{-4} - 10^{-3}$	$< 0.5 \times 10^{-3} \text{ m}$	$< 8 \times 10^{-6} \text{ m}$
$10^{-2} - 1$	$10^{-3} - 0.1$	$0.5 \times 10^{-3} - 0.5 \text{ m}$	$8 \times 10^{-6} - 0.8 \times 10^{-3} \text{ m}$
$1 - 30$	$0.1 - 30$	$0.5 - 15 \text{ m}$	$0.8 \times 10^{-3} - 0.25 \text{ m}$
> 30	> 30	$> 15 \text{ m}$	$> 0.25 \text{ m}$

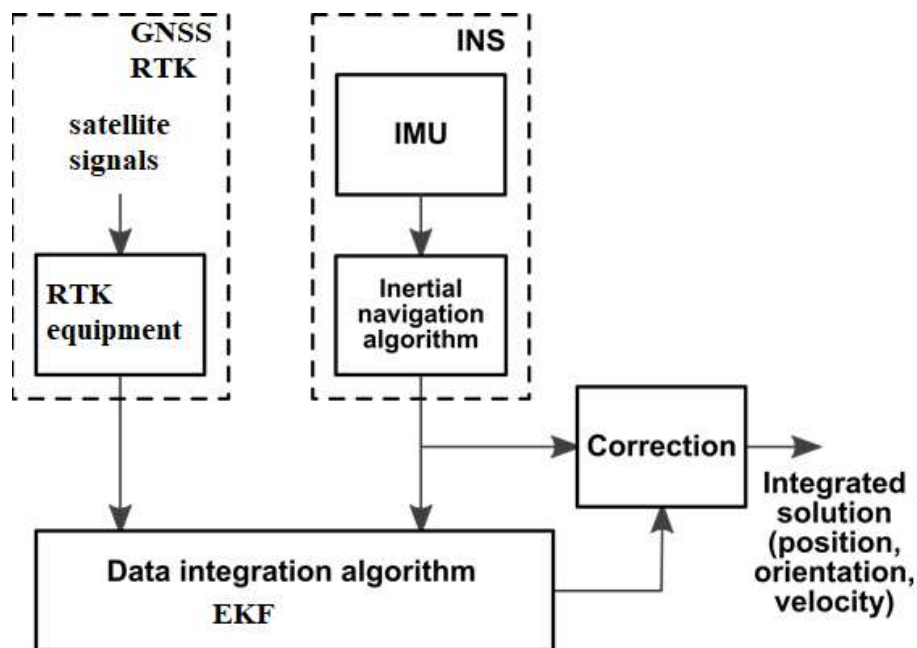


Figure 3: RTK and IMU data fusion for robust and accurate navigation

4. 3D RECONSTRUCTION SCENE

This Section focuses on the vision system. The Section analyzes the stereo correspondence problem and proposes a novel 3D reconstruction scene algorithm that outperforms the existing ones. In Fig. 4 it is showed a system vision process to reconstruct the depth map of a scene from two stereo images acquired with two cameras. Indeed, with two or more cameras we can infer depth, by means of triangulation, if we are able to find corresponding points in the images. In Fig. 4 the offline calibration procedure aimed at finding intrinsic parameters of the two cameras (e.g. focal length, image center, lenses distortion) and extrinsic parameters (e.g. alignment of the two cameras). The offline calibration procedure in Fig. 4 is carried out acquiring and processing tens of stereo pairs of a known pattern (typically checkerboard). The rectification step in Fig. 4 re-projects image planes onto common planes parallel to line between optical centers. Moreover, using the information from the calibration process the rectification step removes from the images the lens distortion and turns the stereo pair in standard form. The correspondence problem in Fig. 4 refers to the problem of ascertaining which parts of one image correspond to which parts of another image, where differences are due to movement of the camera, the elapse of time, and/or movement of objects in the photos. According to the taxonomy proposed in literature most stereo algorithms perform subset of the following steps:

1. Matching Cost Computation
2. Cost Aggregation
3. Disparity Computation
4. Disparity Refinement

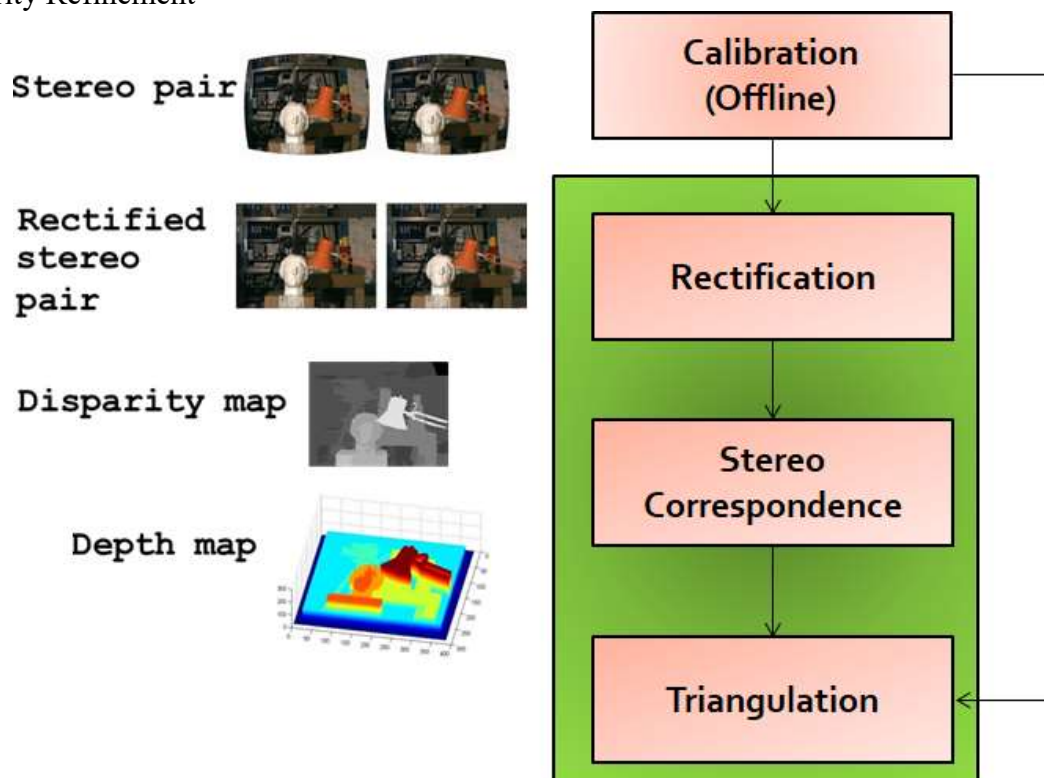


Figure 4: Depth map reconstruction from stereo images acquired with two cameras

Two different strategies are typically implemented in state of art for stereo correspondence:

- Local algorithms, which use the simple WTA (willingness-to-accept) disparity selection strategy but reduce ambiguity by aggregating matching costs over a support window.
- Global algorithms, which search for disparity assignments that minimize an energy function over the whole stereo pair using a pixel-based matching cost.

Given the disparity map, the baseline and the focal length, the triangulation step in Fig. 4 computes the position of the correspondences in the 3D space.

The 3D scene flow estimation aims at jointly recovering dense geometry and 3D motion from stereoscopic image sequences. To realize this objective and overcome limitations of many existing methods, it is proposed to represent the scene as a collection of rigid planes, into which the input images are segmented. Geometry and 3D motion are then jointly recovered alongside an over-segmentation of the scene, and the egomotion of the stereo rig is measured with a RTK positioning system. This enables to define suitable scene priors, perform occlusion reasoning, and leverage discrete optimization schemes toward stable and accurate results. Assuming the rigid motion to persist approximately over time additionally enables to incorporate multiple frames into the inference.

3D scene flow is commonly defined as a flow field describing the 3D motion at every point in the scene. The 3D scene flow problem is traditionally formulated in a variational setting, where optimization proceeds in a coarse-to-fine manner and local regularizers are leveraged to encourage smoothness in depth and motion. To this aim, [10] proposed a slanted plane model which assigns each pixel to an image segment and each segment to one of several rigid moving 3D plane proposals, thus casting the task as a discrete optimization problem which can be solved using α -expansion and Quadratic Pseudo. Boolean Optimization (QPBO) [11]. The goal of the proposed scene flow algorithm is to determine depth and 3D motion vector for every pixel in all the view considered. We keep track of a superpixel segmentation in every view denoted as S_v^t , the set of segments in view v at time t . Given the segmentation, an energy term is defined as a function of two sets of mappings. The first set of mappings is $\Gamma = \{\Gamma_v^t: t, v\}$ with $\Gamma_v^t: I_v^t \rightarrow S_v^t$ assigns each pixel of frame I_v^t to a segment of S_v^t . The second set is $P = \{P_v^t: t, v\}$, a rigidly moving plane is selected for each segment in each view $P_v^t: S_v^t \rightarrow \Pi$, with Π candidate set of possible 3D moving planes.

The energy term includes (i) a Data Term that takes into consideration that corresponding points in different views have similar appearance, and check the geometric configuration for plausibility, test for occlusions and verify the consistency of the segmentation; (ii) a Regularization Term that encourages smoothness of geometry and motion at segment boundaries; (iii) a Boundary Term, which evaluates the quality of the segmentation, encouraging a compact and edge preserving over-segmentation of the reference image.

To perform inference over the 3D geometry and motion of the segments, it is required a set of proposal planes together with their rigid motion. These inputs are created from a combination of stereo and optical flow methods. To convert the pixelwise correspondence information to the representation in terms of planes and motion, it is separately fitted the parameters of a 3D moving plane and its rigid motion to each superpixel of the initial segmentation.

The strategy of selecting parts of the solution from a set of proposals allows to include additional information in an unbiased way, without the need for altering the energy formulation. We exploit this property by including the estimated ego-motion of the stereo system as an additional proposal. Ego motion is found by the measures of an RTK system. The process starts with a fixed segmentation, and optimize the energy with respect to a set P , selecting a suitable moving plane for each segment from the proposal set. After it is found a solution for the set P , the mapping is kept fixed and the energy is optimized with respect to S , reassigning the pixels to segments.

The proposed experiments to assess the validity of the design mainly employs the KITTI training dataset [12, 13], which is suitable for a detailed performance due to images with public ground truth. The results in terms of disparity map [14,15] and flow map are showed in the next figures (Fig. 5 and Fig. 6). The standard KITTI metric is to compute the number of outliers pixels that deviate more than a certain threshold from the ground truth. Fig. 7 shows results for error thresholds of 2, 3, 4 and 5 pixels for the entire image, in terms of Disparity Flow and Flow Estimation, with the comparison of the algorithm developed in [10].

Once all the disparity and flow maps for the frames are obtained, the measures for all points of interest are obtained, see as example in Fig. 8 the case of a vehicle measuring through the remote stereoscopic cameras the width of a road.

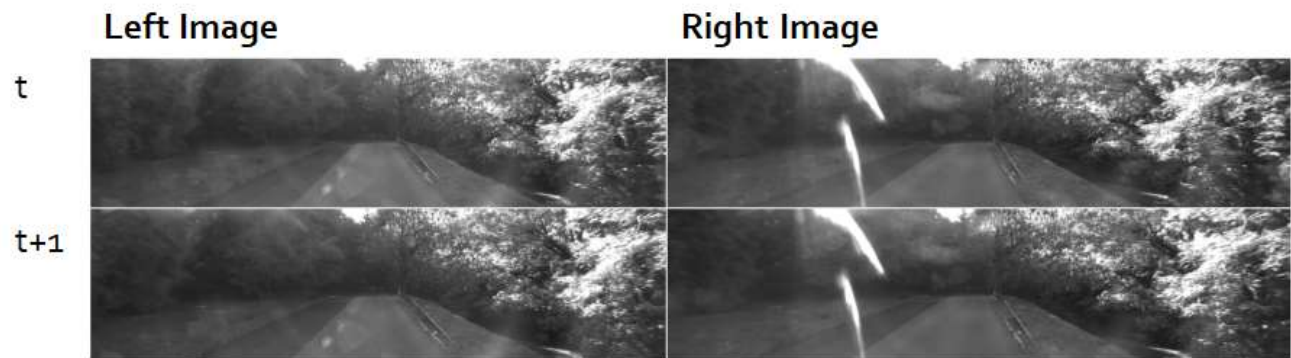


Figure 5: Example image selected from the KITTI algorithm

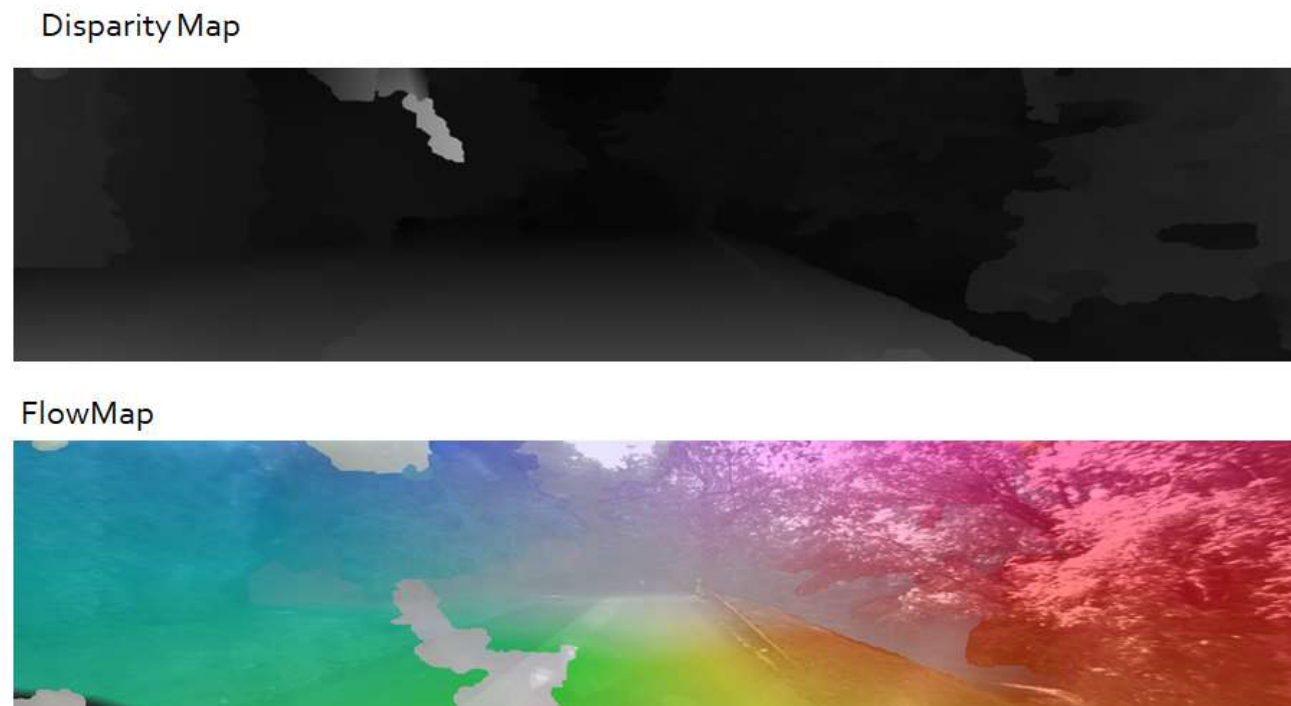


Figure 6: Disparity Map and Flow Map obtained from the images in Fig. 5

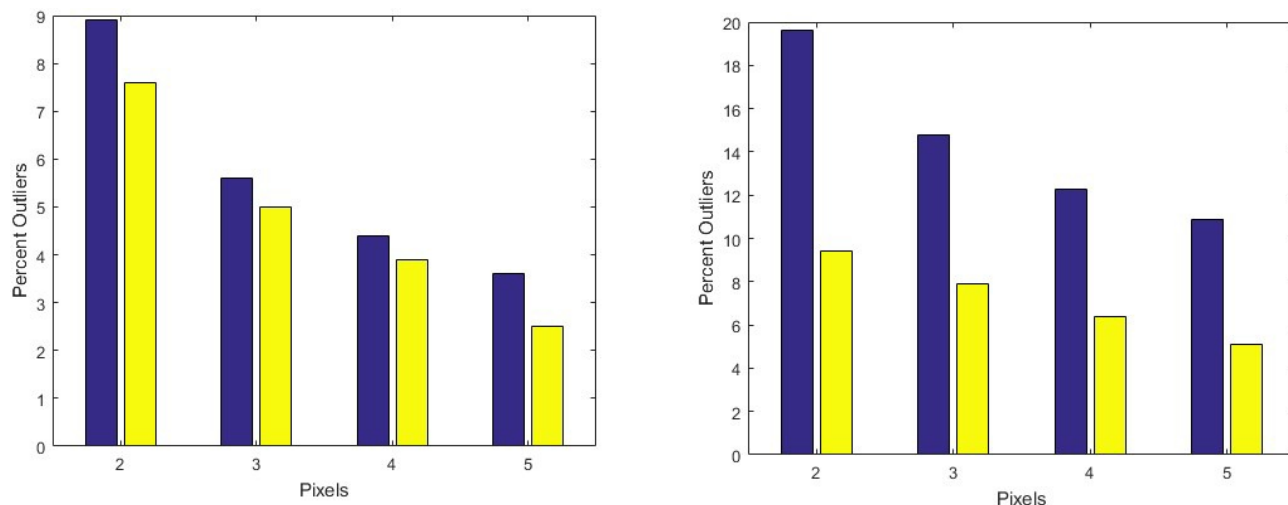


Figure 7: Comparison for Outliers in KITTI metric between the algorithm developed (yellow bar) and PRSF (periodic real Schur form) algorithm (Blue bar), either in Disparity map (Left) and Flow Map (Right)

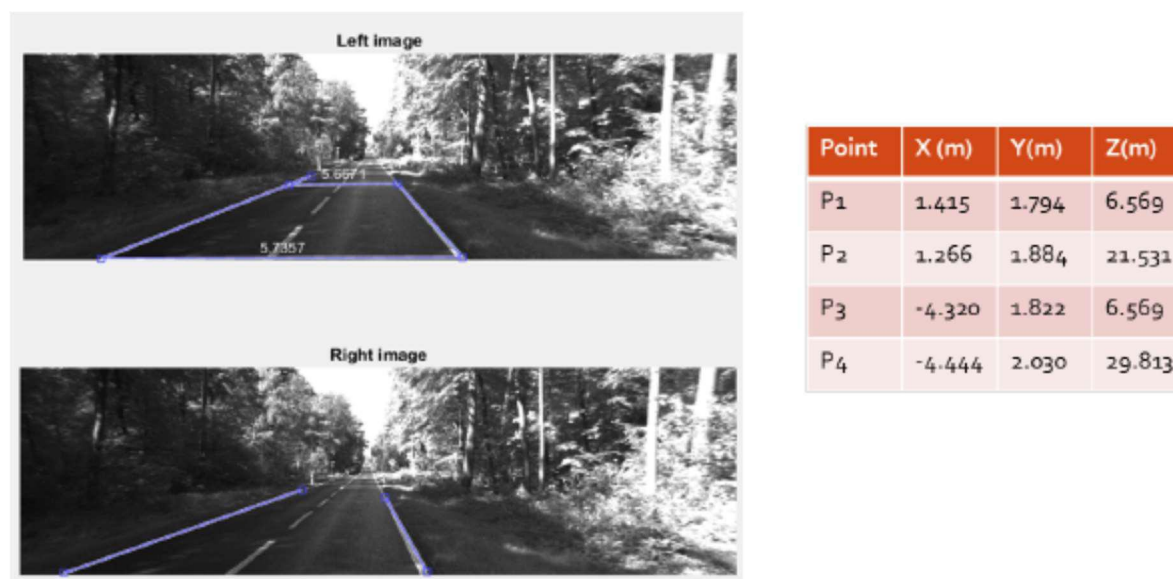


Figure 8: Triangulation of the selected points on the lines

To validate the theoretical contributions, a stereo vision prototype has been also developed with two industrial cameras and an RTK sensor. Repeating the measure in an industrial scenario we have seen that the accuracy of the measurements extracted from the stereoscopic images are in the range of few cm (below 10 cm).

5. CONCLUSIONS

This work has proposed the design of a novel electronic system that can be easily installed on a machinery to ensure cm-accurate positioning, thanks to fusion of RTK and INS through a Kalman extended estimator, plus the real-time implementation of a stereoscopic image processing algorithm

to recover the 3D structure of the scene. The focus is on the 3D reconstruction of the scene to have the reference trajectory for the actuation done by a robotic arm with a proper end-effector.

To this aim, a 3D reconstruction scene algorithm has been developed which can be used in many applications. The image processing technique has been validated through the KITTI benchmark that is ideal for a detailed performance due to images with public ground truth. Furthermore, experimental tests have been done and the results in simulation and in the experiments showed a good matching.

ACKNOWLEDGMENT

This work was partially supported by the RST funded project of Tuscany Region and b the PRA E-TEAM 2017 project; discussions with Pure Power Control srl, Navacchio (Cascina), Italy as well as the contribution of Dr. Donato Cammarano are gratefully acknowledged

REFERENCES

- [1] L. Huisman, P.J.G. Teunissen, D. Odijk, "On the Robustness of Next Generation GNSS Phase-only Real-Time Kinematic Positioning", FIG Congress 2010, Sydney, Australia, 11-16 April 2010
- [2] C. Cristodaro et al., "GNSS Receiver Performance in Urban Environment: Challenges and Test Approaches for Automotive Applications", IEEE Int. Conf. on Electrical and Electronic Tech. for Automotive, 2017
- [3] J. Meguro et al., "Development of an Autonomous Mobile Surveillance System Using a Network-based RTK-GPS", IEEE International Conference on Robotics and Automation 2005
- [4] Achieving Centimeter Level Performance with Low Cost Antennas, White Paper, UBX-16010559 - R01
- [5] M. Skoglund et al., "Static and dynamic performance evaluation of low-cost RTK GPS receivers", IEEE Intelligent Vehicle Symposium, 2006
- [6] S. Saponara et al., "Exploiting mm-Wave Communications to Boost the Performance of Industrial Wireless Networks", IEEE Transactions on Industrial Informatics, 2017, vol. 13, n.3, 1460-1470
- [7] A. Colomè, "Closed-Loop Inverse Kinematics for Redundant Robots: Comparative Assessment and Two Enhancements", IEEE Transactions on Mechatronics, vol. 20, n. 2, 2015
- [8] https://www.swiftnav.com/sites/default/files/piksi_multi_product_summary.pdf
- [9] F. Pieri, C. Zambelli, A. Nannini, P. Olivo, S. Saponara, "Limits of sensing and storage electronic components for high-reliable and safety-critical automotive applications", IEEE International Conference of Electrical and Electronic Technologies for Automotive, 2017
- [10] S. Roth, C. Vogel, K. Schindler, "Piecewise rigid scene flow", International Conference on Computer Vision, 2013
- [11] M. Paget, J.-P. Tarel, L. Caraffa, "Extending alpha-expansion to a larger set of regularization functions", IEEE International Conference on Image Processing, Sept 2015
- [12] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, "Vision meets robotics: The kitti dataset". International Journal of Robotics Research (IJRR), 2013.
- [13] http://www.cvlibs.net/datasets/kitti/raw_data.php
- [14] S. Mattoccia, "Stereo Vision Algorithms for FPGAs", IEEE Conference on Computer Vision and Pattern Recognition Workshops (CPVR) 2013, pp. 636-641
- [15] Rostam Affendi Hamzah et al., "Literature Survey on Stereo Vision Disparity Map Algorithms", Journal of Sensors, vol. 2016 (2016), pp. 1-23,