

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Endoscopic orientation by multimodal data fusion

Silvio Pulwer, Richard Fiebelkorn, Christoph Zesch, Patrick Steglich, Claus Villringer, et al.

Silvio Pulwer, Richard Fiebelkorn, Christoph Zesch, Patrick Steglich, Claus Villringer, Francesco Villasmunta, Egbert M. Gedat, Jan Handrich, Sigurd Schrader, Ralf Vandenhouten, "Endoscopic orientation by multimodal data fusion," Proc. SPIE 10931, MOEMS and Miniaturized Systems XVIII, 1093114 (4 March 2019); doi: 10.1117/12.2508470

SPIE.

Event: SPIE OPTO, 2019, San Francisco, California, United States

Endoscopic orientation by multimodal data fusion

Silvio Pulwer*^a, Richard Fiebelkorn^a, Christoph Zesch^{a,b}, Patrick Steglich^{a,c}, Claus Villringer^a,
Francesco Villasmunta^{a,b}, Egbert Gedat^a, Jan Handrich^a,
Sigurd Schrader^a, Ralf Vandenhouten^a

^aTechnical University of Applied Sciences Wildau, Hochschulring 1, 15745 Wildau, Germany

^bUniversity of Rome Tor Vergata, Via Cracovia n.50, 00133 Rome, Italy

^cIHP – Leibnitz-Institut für innovative Mikroelektronik, Im Technologiepark 25, 15236 Frankfurt (Oder),
Germany

ABSTRACT

To improve the feasibility of endoscopic inspection processes we developed a system that provides online information about position, orientation and viewing direction of endoscopes, to support the analysis of endoscopic images and to ease the operational handling of the equipment. The setup is based on an industrial endoscope consisting of a camera, various MEMS and multimodal data fusion. The software contains algorithms for feature and geometric structure recognition as well as Kalman filters. To track the distal end of the endoscope and to generate 3D point cloud data in real time the optical and photometrical characteristics of the system are registered and the movement of the endoscope is reconstructed by using image processing techniques.

Keywords: endoscope, orientation, Kalman filter, visual odometry, mems, gyroscope, accelerometer

INTRODUCTION

Endoscopes are used in many fields to help engineers and technicians to diagnose malfunctions, defects or errors. Regularly inspections will lead to an improved long-term performance and reliability. Modern endoscopes allow the user to get structural 3D information of the inspected surface by stereo or structured light approaches. However, endoscopes are difficult to operate, and it takes time for users to learn how to interpret endoscopic images and handle the equipment. In addition to the not visible distal end of the endoscope, one of the difficulties is the discrepancy between the spatial orientation of the endoscopic images and the user's working environment. To overcome these shortages, it is crucial to hone the skills of the inspector with On-the-Job Training (OJT) in real-world situations and virtual training including computer simulation [1]. The addition of haptic feedback also addresses the interaction between the endoscope distal end and surfaces [2]. With these methods engineers are able to plan and optimize the inspection process including the path planning before the actual examination.

To exploit the full potential of modern miniaturized endoscopes in practice, the possibility to estimate the position, orientation and location of the distal end in real time is needed. Therefore, it is necessary to detect the relative position of the distal end of the endoscope by measuring the change in position from a starting point [3].

Previous research used this approach to correct the image orientation and improved the handling in medical applications by using inertial measurement units (IMUs) that are based on gyroscopes and accelerometers [4]. Corresponding to automotive industry approaches to estimate the car position, further data of the already integrated camera, either at the endoscopic distal end (chip-in-tip) or at the proximal end (fiberscopes) can be used to provide real-time vision-aided inertial navigation [5]. The combination of both sensor approaches give the possibility to robustly estimate the trajectory of an equipped system by estimating the position with IMU datasets and re-adjust it with the image registration algorithms [6]. Hereby, Kalman filter techniques enable complementary compensation of sensor limitations, resulting in an improved performance of the sensor system.

In the field of medical applications, endoscopic tracking systems are state of the art. They are based on the analysis of external electro-magnetic field distributions relative to the endoscope and hence not usable for industry purposes including machines made of metal that disturb the external field. Nader et al. [7] described a vision based tracking and 3D reconstruction method, which delivers real time information of a monocular endoscope for medical purposes. It takes

advantage of the feature tracking method ORBSLAM introduced by Mur-Artal and Tardós [8]. The requirements for industrial endoscopes, increase towards more precise object detection, defect recognition and measurement. This will lead to a further degree of automatization, which supports the inspection process and the digitalization of the peripheral data. With these newly attained datasets valid long-term measurements are possible. In this work, we propose a method for estimating the position, orientation and viewing direction of the distal end of an endoscope in industrial environments based on two sensing modules.

* pulwer@th-wildau.de; phone: +49 3375508771; www.th-wildau.de/photonik

METHODS

The propose endoscope employs two sensing modules as described in the following.

1. IMU System based on accelerometers and gyroscopes for three axes, integrated into the distal end of the endoscope. By that, the trajectory of the device can be acquired.
2. A camera which is attached to a lens, with entrance pupil located at the distal end of the endoscope. The images taken by the camera are filtered if necessary and afterwards loaded to a direct sparse odometry algorithm (DSO) introduced by Engel et al. [9]. As a result the measurement of the intensities of pixel and gradients will acquire position and orientation of the endoscope.

The multiple output data from each module are fused together into a Kalman filter framework which will include additional constraints coming from CAD datasets. In Fig. 1 a block diagram of the used sensor system for endoscopic orientation is shown.

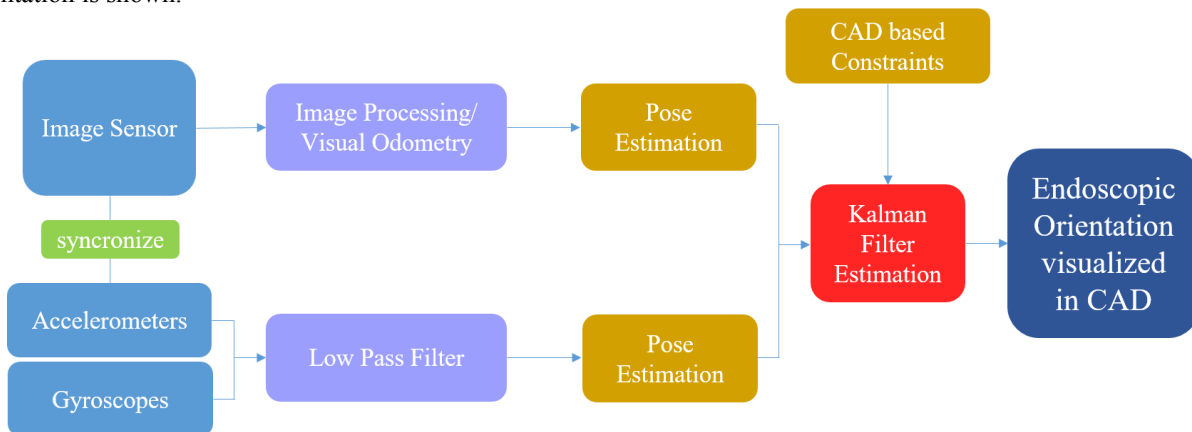


Figure 1. Conceptual sketch of the proposed multiple sensor system inputs fused together and computed with Kalman filter techniques for a visualized orientation of an endoscope.

For the IMU, the definition of three orthogonal vectors of acceleration and rotation, respectively for the recognition of translational movement and of angular rate for the estimation of rotational movement are used. By acquiring the acceleration data relative to the moving endoscope inertial system, the travelled distance is proportional to the double integration of the acceleration value. This operation is prone to drift, due to errors that accumulate over time. The error function for the position is proportional to t^2 and is orientation dependent.

For computing the camera based data, an extended pinhole camera model is assumed. Therefore, the geometric intrinsic camera calibration detects the field of view (FOV), image center and distortion. During the calibration process, a checkerboard can be used as an external 2D pattern. One further requirement is a separate photometric calibration of the system. The independence between geometric and photometric calibration is crucial for the DSO method and leads to much better robustness, for example in cases of auto-exposure controlled cameras. For the photometric image correction, we use the vignetting response function (V) in addition to the non-linear response function (G) described by Engel et al. in Ref. 9. It is given by

$$I'(x) := tB(x) = \frac{G^{-1}(I(x))}{V(x)}, \quad (1)$$

where B is the irradiance and I is the observed pixel intensity, while t is the exposure time. Usually for chip-in-tip endoscopes, the lens is fix attached to the image sensor, but some systems allow additional supplementary lenses or prisms to shift the focus distance or change the direction of view. In this case a multiple correction is necessary. For

sensor fusion, a standard Kalman filter was implemented to estimate position and orientation from IMU and camera data. Whenever the visual based DSO algorithm produces a position estimation, the difference between the position estimates of the two systems is used as the input for a filter that tries to estimate the errors of the orientation solution, as well as the errors for the IMU sensors. As the visual pose estimation operates at typically 20 frames per second and the IMU can be accelerated up to 1 kHz sample rate, the latter has to work suitably stable in these time windows until the next supporting visual pose estimation is delivered. Whenever this is the case the Kalman filter is used with the latest IMU pose estimation as a priori prediction and the visual pose estimation as a posteriori measurement, both with adjustable and recursively refined noise terms. The estimated errors are then used to correct the orientation solution and to approximate the reading of the sensor. Furthermore, it involves CAD based data, which represents a suitable ensemble of motion constraints. To implement these in a Kalman filter, it can be realized by projecting the unconstrained estimates in the direction of the constraints [10].

In this work, we use as basic setup the fiberscope IF6C5X1 from Olympus with 6 mm outer diameter and an attached camera via relay lens optical adapter with a far focus lens with 60° full viewing angle. The endoscope is equipped with integrated illumination fibers which are connected to a white light LED source. In Fig. 2 the endoscope and the used axes conventions that are based on the standardized ISO 1151-2:1985 norm are shown.

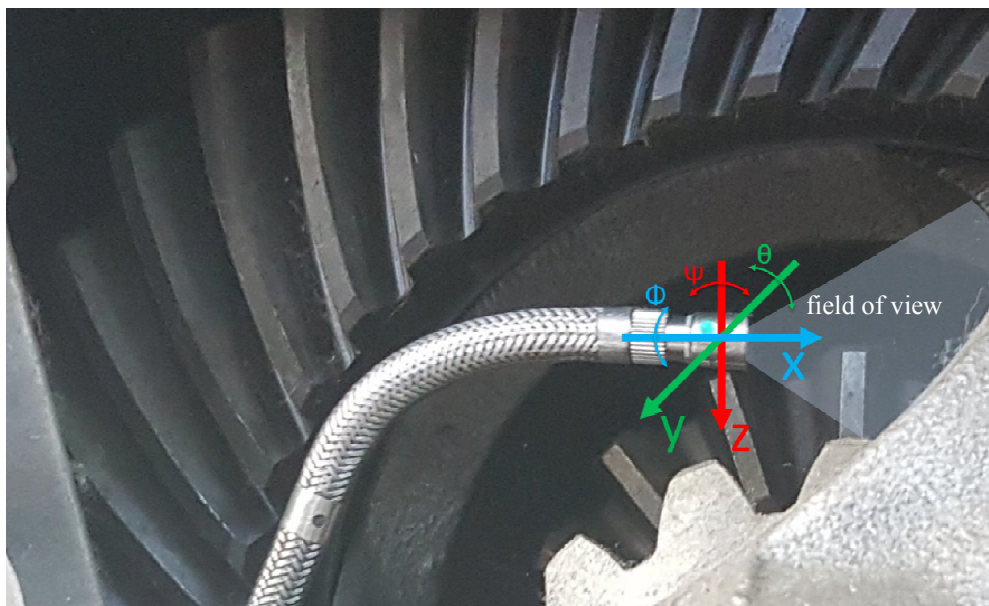


Figure 2. Image of a fiberscope placed inside a half cut gearing box with axes conventions and field of view.

The camera used for recording the sequences is the Chameleon 3 CM3-U-13S2C global shutter CMOS camera from Point Grey. It is capable of recording 1288x964 videos at up to 30 fps. For the IMU consisting of the accelerometer and gyroscope, which are attached to the distal end of the endoscope, we use the MPU-6050 from IDK InvaSense that integrates these sensors into a small package. It is connected to an Arduino microcontroller board which enables us to acquire sensor data via MATLAB® with 16-bit resolution for accelerometer and gyroscope data individually.

RESULTS

First of all we checked the performance of the IMU for tracking abilities. A modified version of I Skog's GPSaidedINS MATLAB® code [11, 12] was used for data fusion based on the IMU data and global navigation satellite system (GNSS) data in a loosely-coupled closed-loop structure by a Kalman filter. The algorithm was tested with car motion data from the KITTI database [13]. In Fig. 3 the results of this inertial navigation system (INS) supported with global positioning system (GPS) data are displayed for the dataset 2011_09_26_drive_009 from the KITTI database. The sample rate for the IMU and GPS in this dataset are approximately 10 Hz. Here for GPS data only each tenth point was used, resulting in a GPS sample rate of approximately 1 Hz. For applications in endoscopic examination, GNSS data will be replaced by 3D positions, extracted from a combination of image processing algorithms, visual odometry and optical flow analysis.

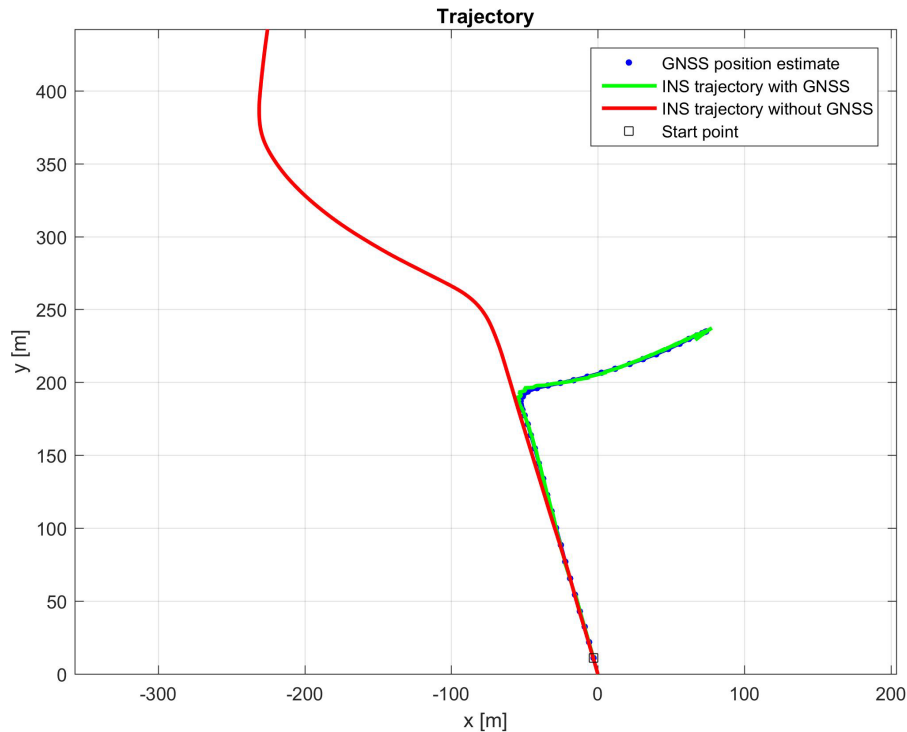


Figure 3. Example of the simulated performance of the chosen Matlab algorithm [11, 12] for guided IMU navigation. The data are based on the KITTI raw database.

Next, the camera data were used. To avoid instable photometric corrections, it was crucial to fix the camera adapter to the endoscopic imaging fiber bundle properly. Additional low pass Gaussian filtering techniques were added to each image. The operation frequency was approximately 10 Hz. For the experiments, we added external white light sources to the scenery.

In Fig. 4 the results of visual based tracking are shown including details of the inspected object. Therefore we used the first compressor stage of the axial compressor of a Klimow TW2-117 shaft turbine. The object contains a complex geometrical structure with prominent angles and areas. The surfaces have a specular and diffuse reflection behavior which leads to high dynamic amplitudes in the images. The endoscope was hand controlled and the live view image was monitored via notebook. Most trackable features were found on the edges of the rotor blades. The trajectory together with 3D point cloud data could be extracted and linear as well as fast accelerated shaking movements could be detected.

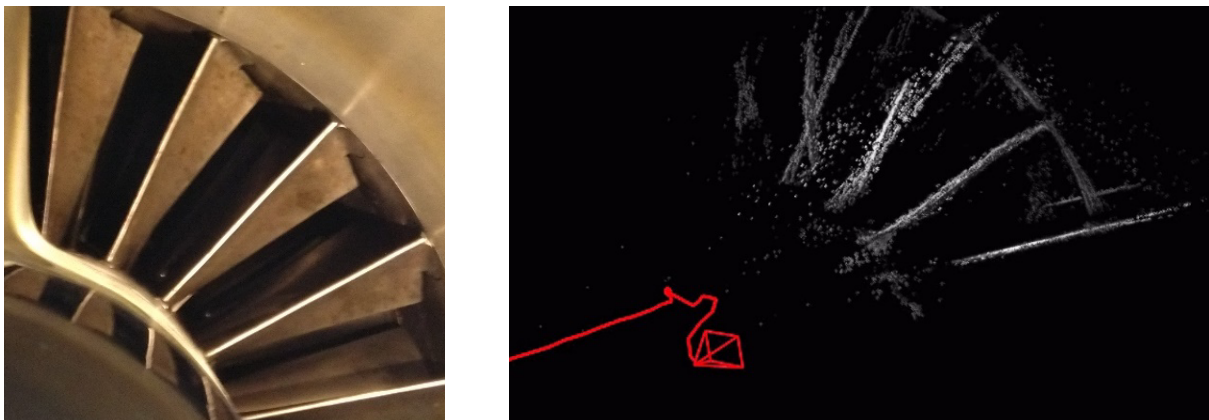


Figure 4. On the left, image of the first compressor stage of the axial compressor of a Klimow TW2-117 shaft turbine. On the right, example for trajectory extraction with camera pose (red colored) based on DSO algorithm including 3D point cloud data (white colored).

DISCUSSION

From our INS calculations we come to the conclusion that an INS that includes inertial sensors can only be used for short time periods to analyze the position, orientation and viewing direction of an endoscope due to the permanent drift of acceleration data. To overcome this problem, the information should be frequently fused with additional information provided by 3D positions, extracted from a combination of feature/geometric structure recognition algorithms, visual odometry or optical flow analysis. It can be seen in Fig. 3 that the tracking of the vehicle only with INS data leads to severe mislocation and the repeated support with additional data – GPS in the case – stabilizes the location.

Hereby, the visual based navigation can provide sufficient data about the position, orientation and viewing direction. But this depends strongly on the existence of unambiguous features, geometric structures or intensity gradients. In addition to that, the illumination system is very close to the optical axis. Therefore it is prone to create direct reflexes on the image sensor when the object surface shows specular reflection or strong Gaussian scattering behavior due to polished metal surfaces. For this reason, prominent geometrical features like edges or holes enable a much more stable trajectory extraction. With the obtained 3D point cloud coordinates and additional mesh processing a direct feedback loop with the CAD dataset of the object can be applied to improve trajectory extraction. This requires a well-controlled geometric noise function. The global shutter of the camera, the geometric calibration and the direct sparse model of Engel et al. [8] lead to sufficient results, even at fast accelerations of the tip of the endoscope. In practice, most endoscope manufacturer integrate rolling shutter based image sensor with well corrected lenses that will tend to much more complex geometric noise. To correct that, Li et al. proposed a method that will include the readout time of the sensor into the Kalman filter [14]. Furthermore, additional Gabor filter techniques and structured light can improve the recognition of geometrical features. Future work is focusing to fuse the methods on one platform together and extend the Kalman filter with CAD constraints.

CONCLUSION

In this paper, we proposed a method that provides online information about the position, orientation and viewing direction of an endoscope. We particularly implemented the method on an industrial endoscope and tested the individual sensors for their tracking abilities. We found that visual based analysis in combination with IMUs can work well when fused together into a Kalman filter framework. This work serves as a first attempt to use these estimates for the development of an inexpensive endoscopic orientation system.

ACKNOWLEDGEMENTS

This project is funded by the Ministry for Science, Research and Culture of the Federal State of Brandenburg in the frame of the ERDF program (European Regional Development Fund) under contract no.85009217.



EUROPEAN UNION

European Regional
Development Fund

REFERENCES

- [1] Vora, J. et al., "Using virtual reality technology for aircraft visual inspection training: presence and comparison studies", *Applied Ergonomics* 33, 559–570 (2002)
- [2] Vembar, D. et al., "A Haptic Virtual Borescope for Visual Engine Inspection Training", *IEEE 3DUI-2008* (2008)
- [3] Harada, T. et al., "Portable orientation estimation device based on accelerometers, magnetometers and gyroscope sensors for sensor network", *IEEE MFI-2003* (2003)
- [4] Höller, K. et al., "Orientation of endoscopic images: rectification by gravity", *Biomedical Engineering* Vol. 55, No. 4, 211–217 (2010)
- [5] Mourikis, A. I. and Roumeliotis, S. I., "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation", *2007 IEEE ICRA-2007* (2007)
- [6] Kouroggi, M. and Kurata, T., "A method of personal positioning based on sensor data fusion of wearable camera and self-contained sensors", *IEEE MFI-2003* (2003)
- [7] Nader, M. et al., "ORB-SLAM-based Endoscope Tracking and 3D Reconstruction", *MICCIA CARE-2016* (2016)
- [8] Mur-Artal R. and Tardós, J. D., "ORB-SLAM: Tracking and Mapping Recognizable Features", *Science and Systems MVIGRO* (2014)
- [9] Engel, J. et al., "Direct Sparse Odometry", *IEEE TPAMI-2017* (2017)
- [10] Massicotte, D. et al., "Incorporation of positivity constraint into a Kalman-filter-based algorithm for correction of spectrometric data", *IEEE Trans. Instrum. Measure* (1995)
- [11] Isaksson, M., <https://gitlab.com/martisak/GPSaidedIMU>, visited 12/19/2018
- [12] Skog, I and Bernhardsson, B., https://kth.instructure.com/files/677996/download?download_frd=1, visited 12/19/2018
- [13] Geiger, A. et al., "Vision meets Robotics: The KITTI Dataset", *IJRR* (2013)
- [14] Li, M. et al., "Real-time Motion Tracking on a Cellphone using Inertial Sensing and a Rolling-Shutter Camera", *IEEE ICRA-2013* (2013)