Vision-based Probabilistic Absolute Position Sensor

Rene Paris, Martin Melik-Merkumians, Georg Schitter

Automation and Control Institute, Vienna University of Technology,
Gußhausstraße 27-29, A-1040 Vienna, Austria
{paris, melik-merkumians, schitter}@acin.tuwien.ac.at

***Abstract*—Many industrial applications require to determine the absolute position of an extended surface without modifying or touching the target object. This contribution presents a concept for an optical absolute position sensor based on an off the shelf camera, operating perpendicular to an extended surface over long strokes, as necessary e.g. for piston actuators. The proposed sensor uses a Particle Filter to measure the absolute position within an once-only learned global feature map with low memory footprint, which is archived by using an adapted feature detector. The probabilistic approach allows for certain robustness against false feature detection and enables fast recovery after power loss without the need for a referencing movement. The absolute position is detected with sub-millimeter accuracy over a stroke of 100 mm.**

I. Introduction

Throughout modern industries piston actuators are widely used for numerous tasks such as pushing objects, folding, or as precise tools using feedback for accurate positioning. In order to control such actuators, as also needed for intermediate positioning or trajectory planning, the accurate position must be measured over the full range. The challenge is to measure the absolute position with moderate precision but over long strokes and in the ideal case without modifying or touching the measurement target. In addition, since measuring from the abutting face is problematic or even impossible by construction, the surface has to be measured perpendicular to its extension as depicted in Fig. 1.

An often used scheme is to measure the position of a magnet which is embedded in the piston. While simple detection of the end position is done by cheap Reed switches, more accurate position sensing is achieved by laterally resolving the magnetic field strength. The latter allows to measure the absolute position, which is a beneficial for increased system performance after sudden power loss. Nevertheless, such sensors are restricted by the length of the sensor head, which is typically on the order of several centimeters only (Fig. 1A). To detect multiple positions or measure over a wider range, embedding more magnets is a possibility [1] (Fig. 1B), but comes at the price of increased costs and likelihood of mechanical failure. In addition, this cannot be applied to existing systems without changing the actuator too.

Here, an optical solution offers an alternative for continuous position measurement even over long strokes. One possible sensing solution are laser Doppler vibrometers [2] which are extremely accurate and low-cost solutions seem to be in reach [3]. Nevertheless, the measurement principle of vibrometers only allows incremental operation and cannot correct for offset errors or drifts after glitches or abrupt target moves. Another

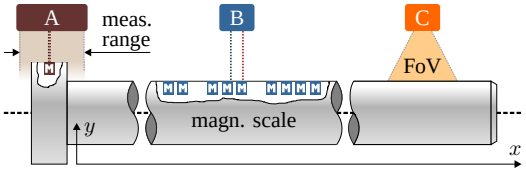

Fig. 1: Several absolute position measurement systems for a piston actuator, showing (A) a magnetic sensor using a single embedded magnet *M* but is limited in its measurement range by the dimension of the sensor head, (B) a magnetic sensor using a specific magnetic scale over the whole stroke and (C) an optical sensor using images within its field of view *FoV*.

class of sensors is based on camera sensors (Fig. 1C), which records images from the surface. Today's camera sensors are compact, cheap, ubiquitous, and also accurate [4].

A simple approach, which allows to transform a camera into a an absolute sensor, is to image the surface with a defined grid and correlate the actual image against this data set [5]. The grid point image with the highest correlation value has the highest probability of being the actual position. Relaxation of the grid density can be archived by interpolation between images and their correlation peak values [6], which leads to a lateral resolution of 1 µm with camera images taken only every 10 µm [6]. Nevertheless, with increasing stroke these methods can get intense in memory consumption.

This contribution presents a concept for an optical absolute position sensor based on a Particle Filter (PF), operating perpendicular to an extended surface. The goal is to archive an absolute position accuracy of less than 100 µm over a total stroke of 100 mm. In addition, memory consumption which grows with the measurable distance is of great concern. The concept is based on principles from vision-based localization [7][8] and extended for the specific sensor geometry, leading to a reduced memory footprint and computational complexity. The sensor estimates the actual absolute position by inferring on a once-only learned global feature map. Implementing a probabilistic approach allows for certain robustness against false feature detection and enables the sensor to recover from an unknown state, such as after power loss.

II. Probabilistic Localization

Like in vision-based localization, sensor readings are not taken from the whole image, but from observation of known landmarks stored in a global map [8]. The actual position

Post-print version of the article: René Paris, Martin Melik-Merkumians, and Georg Schitter, "Vision-based probabilistic absolute position sensor," *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015. DOI: 10.1109/I2MTC.2015.7151601

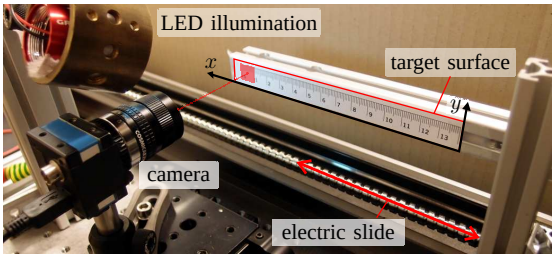

Fig. 2: Laboratory setup with the target surface mounted onto an electric slide with a single degree of freedom, and the camera with its approximate region of interest on the surface, which is illuminated by white LEDs.

within the map is eventually estimated by means an PF. In the following the steps necessary for performing this probabilistic localization is given.

Let the true position of the sensor be described by the state vector $\mathbf{x}_t = (x\ y)^T$, which refers to the global coordinates at time t. As depicted in Fig. 1 the actuation direction is assumed to be along the x axis.

Under the *Markov assumption*, stating that $\mathbf{x}_t$ is complete and the best predictor of the future, it only depends on its previous state $\mathbf{x}_{t-1}$ and the actual control inputs $\mathbf{u}_t$. This dependence is described by the *state transition probability* $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_t)$. Observations $\mathbf{z}_t$ of the sensor are modeled by the *measurement probability* $p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{m})$, taking the actual state and a global map $\mathbf{m}$ into account.

The *belief* is defined by the conditional probability $bel(\mathbf{x}_t) = p(\mathbf{x}_t|\mathbf{z}_t, \mathbf{u}_t)$ and reflects the knowledge of the sensor about the state of the environment. A recursive Bayes filter estimates $bel(\mathbf{x}_t)$ from $bel(\mathbf{x}_{t-1})$ by a *prediction* and an *update* step, incorporating $\mathbf{u}_t$ and $\mathbf{z}_t$, respectively. The most prominent possibility to perform recursive Bayes localization is by means of the well-known Kalman filter [9].

Like in vision-based localization, the sensor readings $\mathbf{z}_t$ emerge from the observation of known landmarks stored in a global map $\mathbf{m}$. The correspondence between observations and individual landmarks is typically established by the maximum likelihood of individual landmark descriptors, meaning that landmarks should be as unique as possible to prevent erroneous assignment. This data association problem can be catastrophic for Kalman filter based localization [10] and is one of the reasons for alternative approaches, such as PFs.

PFs estimate the actual state by a set of random state samples. This random set of samples, or particles, is denoted by $\mathbf{X}_t := \mathbf{x}_t^{[1]}, \mathbf{x}_t^{[2]}, ..., \mathbf{x}_t^{[N]}$, with N being the number of particles used. The key idea is to represent the belief $bel(\mathbf{x}_t)$, also called Bayes filter posterior, by a particle density which has the same distribution as the belief itself. As $N \rightarrow \infty$ the likelihood for the state hypothesis included in $\mathbf{X}_t$ goes asymptotically towards the real posterior. A finite particle count leads to a different distribution but in practice this difference can be neglected as long as N is large enough ($N \geq 100$) [10].

PFs recursively estimate the posterior $bel(\mathbf{x}_t)$ from $bel(\mathbf{x}_{t-1})$, meaning they construct $\mathbf{X}_t$ from the random set $\mathbf{X}_{t-1}$. The algorithm consists of three steps:

1) *Prediction step:* Generate a state hypothesis for each individual particle $\mathbf{x}_t^{[n]}$ based on its former state $\mathbf{x}_{t-1}^{[n]}$ and the control vector $\mathbf{u}_t$ by sampling from the state transition probability $p(\mathbf{x}_t|\mathbf{u}_t, \mathbf{x}_{t-1})$.
2) *Update step:* Weight each particle by incorporating the measurement vector $\mathbf{z}_t$ by means of the measurement probability $\mathbf{w}_t^{[n]} = p(\mathbf{z}_t|\mathbf{x}_t^{[n]}, \mathbf{m})$, where $\mathbf{w}_t$ is called *importance weight*.
3) *Importance sampling:* Approximating $bel(\mathbf{x}_t)$ by drawing N particles with replacement from the actual set based on their weight.

An advantage of PFs over single hypothesis recursive Bayes filter is their ability to cope with erroneous data associations. Each particle evaluates observations only with respect to itself. If an observation is now assigned to the wrong landmark, only the state of the actual particle will be slightly altered. This will lead to a decreased importance factor for the altered particle and as a consequence this particle is less likely to be resampled during the importance sampling step and will "die out".

III. Proposed Approach

Similar to a landmark based approach in mobile robotics [10], a system is proposed which extracts features of interest from the actual sensor image and uses the above described PF to localize the sensor.

The *state hypothesis* within the *prediction step* is modeled by $\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{u} + \epsilon$, where the control input u is generated from the actual frame-to-frame displacement, and the random term ϵ is a random displacement based on the variance over all displacements during map generation (see Sec. IV).

To enable absolute position measurement, an once-only initial scan of the surface is executed, the coordinates of selected features are extracted, and a global map is generated. During operation, the coordinates of all visible features $\mathbf{z}$ within the map $\mathbf{m}$ are taken into account within the *update step*. Feature coordinates extracted from the live image are set into a 1:1 relation with respect to the global map by a nearest neighbor search. For calculation of the normalized *importance weight* $w_t^{[n]}$ the measurement noise is assumed to be Normal distributed. This might not reflect the exact noise generated by the feature detector, but is robust enough to cope with small modeling errors [10]. It also allows to use the Mahalanobis distance [11] as weighting base, which enables a direct incorporation of the covariances extracted during map generation. Since no rotational movement is expected, x and y variances are assumed to be independent. This finally leads to a diagonal covariance matrix which further reduces the calculation to the normalized Euclidean distance

$$d(\mathbf{m}, \mathbf{z}_t^{[N]}) = \sqrt{\frac{(m_x - z_{t,x}^{[N]})^2}{s_x^2} + \frac{(m_y - z_{t,y}^{[N]})^2}{s_y^2}}, \quad (1)$$

where $z_{t,x}^{[N]}$ and $z_{t,y}^{[N]}$ are the coordinates of the current observations with respect to the actual particle, m_x and m_y are

Post-print version of the article: René Paris, Martin Melik-Merkumians, and Georg Schitter, "Vision-based probabilistic absolute position sensor," *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015. DOI: 10.1109/I2MTC.2015.7151601

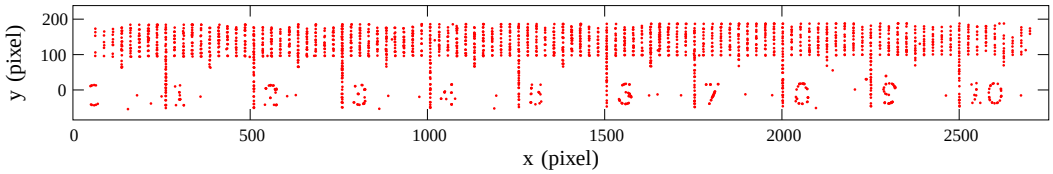

Fig. 3: Global map generated through stitching feature coordinates from adjacent images

the matched map coordinates, and s_x and s_y are the standard deviations of these coordinates. This procedure is executed for all particles and the resulting weights are normalized.

Finally the sensor position is estimated by taking the weighted average over all particles before resampling by

$$\mathbf{x}_{est} = \sum_{n=1}^{N} w_t^{[n]} \mathbf{x}_t^{[n]}. \tag{2}$$

IV. Implementation

A. Laboratory Setup

For the laboratory setup (Fig. 2), a camera (sensor) is mounted in parallel to a moveable target and focused at its surface.

As target acts a ruler printed onto a regular white paper. The pattern has been selected to guarantee reliable feature detection. The target surface is mounted on a rigid frame and actuated along its x-axis by an electric slide (EGSP-33-500-6P-P, Festo AG, Esslingen, Germany) with a repeatability of $\pm 3\,\mu m$. As the desired resolution is $100\,\mu m$, the sensor reading of the electric slide is taken as ground truth.

The camera (DMK22BUC03, The Imaging Source Europe GmbH, Bremen, Germany) has a gray scale CMOS sensor (MT9V024, Micron Technology Inc., Boise, Idaho, USA) with a maximum resolution of 744×480 pixel at 76 frames/s, a square pixel pitch of $6\,\mu m$, and a dynamic range of 8 bit. The objective used, limits the field of view to approximately 30 by 19 mm. Illumination is done by diffuse white light using white LEDs and a diffuser. Exposure and gain values of the camera are fixed and adjusted once-only manually. In addition, the setup is shielded from stray light by a cardboard box.

The camera is calibrated using the MATLAB *Computer Vision Toolkit* (The Mathworks Inc., Natick, Massachusetts, United States) and a standard checkerboard pattern with a square size of $1\,mm^2$. To further reduce nonlinear distortions caused by lens aberrations especially at the boundaries of the image, a region of interest is chosen in the center of the image with a size of about a third of the whole camera image.

B. Feature detection

Feature detection for PF based systems is a tradeoff between feature detection complexity and particles necessary. The less distinguishable features are, the more particles are needed to cope with additional erroneous data association. In order to guarantee a certain uniqueness of features, more complex detectors are used, which comes again at the price of increased computational effort. For good performance, feature detection needs to be adapted for the task given.

One of the most prominent detectors due to its robustness is Scale Invariant Feature Transform (SIFT) [12], which can be seen as *gold standard* for many fields, such as visual object tracking [13] or object classification [14]. The major asset of SIFT is definitely its ability to detect features which are changing in size and rotation, as often necessary in camera based applications. Dissecting SIFT allows to develop a robust feature detection.

For the task given with a fixed sensor-to-surface orientation the required calculation steps in order to achieve rotation invariance can be ignored completely. In addition, the distance between sensor and surface can be regarded constant, meaning that individual features must not be detected at different scales. Nevertheless, surfaces naturally exhibit features which differ in size and shape. Under the assumption that the expected size of the features is known, as possible for targets with similar surface structure, the necessary image scaling steps can be adjusted and reduced to a minimum.

In a final step, SIFT calculates the sub-pixel position of the detected features. This step can be omitted under the assumption, that quantization errors occur statistically independent. As long as enough statistically independent features are visible at the same time, errors are averaged out by the PF itself. This expected quantization noise can be covered by increasing the standard deviation used in the Gaussian measurement model.

Summarizing, the object detector stage based on the Laplace of Gaussian (LoG) operator applied to different scales of the image, should be sufficient for robust detection of features for the task given. For the actual implementation only the coordinates extracted using the LoG detector are used. For more complex surfaces also intensity or histogram information at the detected coordinates could be matched.

C. Global feature map

The state is estimated within a global feature map, which is generated once-only for the setup. As in [6] images are taken only every 10^{th} increment of the desired resolution. This means that the target is scanned with a step size of 1 mm. Second, the scanned images are processed by the feature detector to extract the local *x* and *y* coordinates of each feature of interest.

To generate a global coordinate system, the local coordinates are corrected by a displacement with respect to the first image. The displacement is measured with a sub-pixel accuracy of $1/10$ pixel by cross-correlation peak detection between consecutive images, as described in [15]. Third,

Post-print version of the article: René Paris, Martin Melik-Merkumians, and Georg Schitter, "Vision-based probabilistic absolute position sensor," *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015. DOI: 10.1109/I2MTC.2015.7151601

the corrected coordinates are stitched by averaging nearest neighbors within close proximity. The standard deviation for each merged coordinate is stored individually. Outliers are filtered by removing coordinates with high standard deviation. The coordinates of the resulting map is shown in Fig. 3, clearly showing the image of the ruler.

The generated map consists of 2254 coordinates plus their according standard deviation stored as single precision floating point numbers and consumes only about 70 kB. If the same images are used for a correlation based detection of the position, about 11 MB of memory are necessary, which is 165 times more. This significant reduction in required memory is one of the specific strengths of the proposed approach.

V. Results

For the following experiments, a step size of 100 µm over a stroke of 100 mm is applied to the electric slide. As mentioned above, the ground truth is given by the internal sensor of the electric slide. To prevent dynamic effects, the algorithms are tested in a quasi-static scenario. Note that the images for map generation are recorded in an independent scan one week before the other experiments to ensure sample set independence.

To check whether the assumptions made for the LoG descriptor hold, two PF implementations using 100 particles are compared to each other in a first experiment. The first implementation computes observations generated by SIFT feature coordinates, while the second uses the LoG detector. Fig. 4 shows a comparison of the displacement error of the detected position using SIFT and LoG with respect to the ground truth. As expected, both detectors show similar performance, even though LoG is using integer pixel coordinates only and less computational steps. Therefore, further experiments are executed using LoG only. The reasons for the positive offset will be explained later in detail.

In a second experiment, the recovery behavior of the LoG implementation is tested, as would be necessary after a sudden power loss. The state is assumed to be unknown and therefore initialized by a uniform distribution along the x coordinate. The physical extension of the map is a-priori knowledge and is exploited to filter particles based on their location. This is done by a plausibility check after each control update, which assigns particles a zero importance weight if they leave the known boundaries. Since such particles are not selected for resampling, condensation of the state estimate is improved. The target surface is again moved in 100 µm steps starting from the coordinate origin. As shown in Fig. 5, the system is able to recover from the unknown state. Within 20 iterations, the position error reduces drastically and finally reaches a stable and correct measure. The same behavior can be observed in the state variance which can be used as direct measure for measurement uncertainty (not shown). Even though the position is a weighted average over the particle positions and their importance weight, the initial error is half of the maximum stroke. This is a result of the plurality of possible solutions given by the periodic ticks of the target. As the additional cyphers of the ruler break this periodicity, false particles still can be identified. From that it can be concluded, that a truly random surface would decrease the cycles necessary for a state recovery.

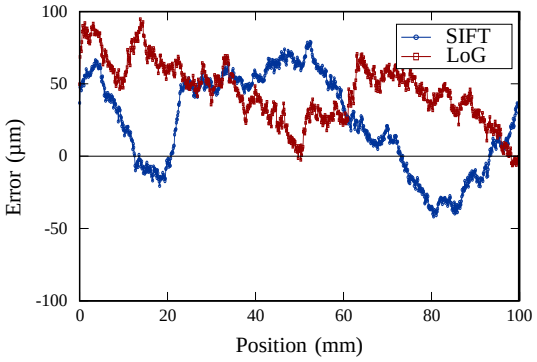

Fig. 4: Performance comparison of two PFs using coordinates extracted by SIFT and LoG.

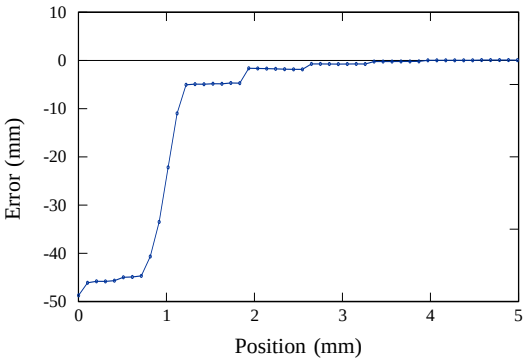

Fig. 5: Position error for the first 50 cycles for an unknown initial state.

In a third experiment, repeatability under real world conditions is tested. Therefore the target is moved from 0 mm to 100 mm (trace) and back to 0 mm (retrace). The position error shown in Fig. 6 has an RMS value of 34.9 µm and a standard deviation of 14.4 µm. It shows a positive position offset and a strong correlation between trace and retrace, which is also observed in the former experiments. This effect can be explained by uncertainties of the displacement estimation within the map generation, used to model the measurement uncertainty for each coordinate (see Sec. IV). This displacement estimate is afflicted with errors causing a mismatch between ground truth and position estimate. Knowledge of this mismatch allows to correct each measured position, leading to a more accurate measurement. To acquire correction values an independent measurement cycle is executed. For low variance, the posterior is sampled with 500 particles instead of 100 (see Sec. II). The position is matched against the ground truth by a polynomial fit, which acts as look-up table for the measured values of the last experiment. Fig. 7 shows the corrected position error which stays below 51 µm with a significantly reduced RMS value of 12.3 µm, and a standard deviation of 12.4 µm.

Summarizing, the experiments show that PFs allow to

Post-print version of the article: René Paris, Martin Melik-Merkumians, and Georg Schitter, "Vision-based probabilistic absolute position sensor," *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015. DOI: 10.1109/I2MTC.2015.7151601

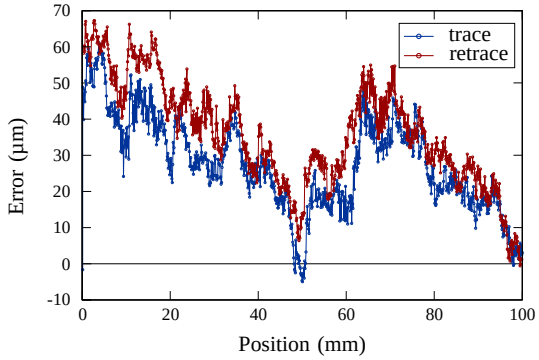

Fig. 6: Position error for a run with 100 particles.

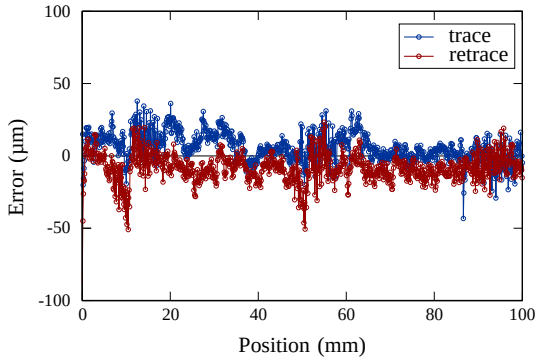

Fig. 7: Position error after compensating for map mismatch.

implement a vision-based absolute position sensor with a significantly reduced memory footprint. They further demonstrate that the proposed approach is able to track the in-plane motion of planar surfaces with a resolution better than 100 µm over a measurement range of 100 mm.

VI. Conclusion

This contribution presents a concept of an optical absolute position sensor that is based on camera in combination with Particle Filtering, and is able to track the in-plane motion of planar surfaces with a resolution better than 100 µm over a measurement range of 100 mm. Given the fact, that the initial scan has been performed with 1 mm steps, the resolution matches the expected values and is high enough for the targeted application. In addition, the measurement range is only limited by the target and can be easily extended without modification of the algorithms.

Instead of correlating against an image database, the proposed system uses observations of individual features, which are matched against a global feature map by means of probabilistic localization. It is shown that the careful adaption of the feature detection steps to the application leads to comparable results as with more complex detectors, such as SIFT, even when using coordinates only. Uncertainties of the used lightweight LoG detector has been successfully countered by incorporating the measurement uncertainty into each coordinate stored within the once-only generated global feature map. The system performed well using only 100 particles and was also able to recover from an unknown initial state, as necessary for absolute position sensors. A specific strength of the presented approach is the low memory footprint of the generated map, which is about 165 times smaller than a purely correlation based approach using the same data set.

Future work will aim at improved map stitching, operation on different surfaces without preparation, as well as robustness against surface changes.

VII. Acknowledgments

Financial support by Festo AG (Esslingen, Germany) as part of the Festo System Laboratory at the ACIN is gratefully acknowledged.

References

[1] S. Yang, M. Lee, and M. Lee, "Absolute position detection method for stroke sensing cylinder," Jun. 29 1999, uS Patent 5,918,199.

[2] P. Castellini, M. Martarelli, and E. Tomasini, "Laser Doppler Vibrometry: Development of advanced solutions answering to technology's needs," *Mechanical Systems and Signal Processing*, vol. 20, no. 6, pp. 1265–1285, 2006.

[3] L. Scalise and N. Paone, "Laser Doppler vibrometry based on self-mixing effect," *Optics and Lasers in Engineering*, vol. 38, no. 3, pp. 173–184, 2002.

[4] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 1052–1067, 2007.

[5] S. Patzelt, K. Pils, A. Tausendfreund, and G. Goch, "Optical absolute position measurement on rough and unprepared technical surfaces," in *Proceedings of the 12th euspen International Conference*, vol. 1. Stockholm: euspen, June 2012, pp. 84–87.

[6] M. Farsad, G. Goch, and C. Evans, "Displacement measurement using speckle correlation with reduced number of database patterns," in *Proceedings of the 28th Annual Meeting of the American Society for Precision Engineering (ASPE)*, 2013.

[7] J. Wolf, W. Burgard, and H. Burkhardt, "Robust vision-based localization by combining an image-retrieval system with monte carlo localization," *IEEE Transactions on Robotics*, vol. 21, no. 2, pp. 208–216, April 2005.

[8] S. Se, D. Lowe, and J. Little, "Vision-based global localization and mapping for mobile robots," *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 364–375, June 2005.

[9] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[10] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.

[11] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences (Calcutta)*, vol. 2, pp. 49–55, 1936.

[12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[13] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," *Computer Vision and Image Understanding*, vol. 113, no. 3, pp. 345–352, 2009.

[14] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 490–503.

[15] R. Paris, T. Thurner, and G. Schitter, "Compensation based displacement measurement using objective laser speckles," in *Proceedings of 6th IFAC Symposium on Mechatronic Systems*, no. 1, 2013, pp. 264–270.

Post-print version of the article: René Paris, Martin Melik-Merkumians, and Georg Schitter, "Vision-based probabilistic absolute position sensor," *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015. DOI: 10.1109/I2MTC.2015.7151601