

Real-time Image-based Localization for Hand-held 3D-modeling

Elmar Mair · Klaus H. Strobl · Tim Bodenmüller ·
Michael Suppa · Darius Burschka

Received: 30 April 2009 / Accepted: 3 July 2009
© Springer-Verlag 2010

Abstract We present a self-referencing hand-held scanning device for vision-based close-range 3D-modeling. Our approach replaces external global tracking devices with ego-motion estimation directly from the camera used for reconstruction. The system is capable of online estimation of the 6DoF pose on hand-held devices with high motion dynamics especially in rotational components. Inertial information supports directly the tracking process to allow for robust tracking and feature management in highly dynamic environments. We introduce a weighting function for landmarks that contribute to the pose estimation increasing the accuracy of the localization and filtering outliers in the tracking process. We validate our approach with experimental results showing the robustness and accuracy of the algorithm. We compare the results to external global referencing solutions used in current modeling systems.

Keywords 3D modeling · Visual localization · Hand-held scanning · Inertia aided visual tracking

1 Motivation

The task of hand-held 3D-modeling requires a solution to the following two problems: the distance of the object to the

scan device has to be determined and the scanner has to be located within a global coordinate frame to fuse the acquired measurements. The global registration is commonly done with external tracking devices that calculate the pose from a fixed set of markers which needs to be constantly in view. In opposite to that a localization directly from the perception of a hand-held device needs to cope with continuous changes of the landmarks used for localization and significant motion due to the close distance to the scene. An advantage of such a solution is the correlation of measurement errors with reconstruction inaccuracies. The error appears mostly along the optical axis which is the axis with the smallest spatial resolution, and, therefore, remains mostly unobserved.

There are several applications which prefer the use of optical sensors, like e.g. mobile outdoor robots, due to its passivity, low power consumption, and compactness. In the case of the hand-held scanner, the use of its camera as a sensing device has several disadvantages:

- Since the objects are close to the scanner and the scan process itself demands a motion perpendicular to the optical axis, the area seen by the camera changes rapidly. This impedes the landmark management and causes possible error accumulation due to a continuous change of landmark sets used for navigation.
- The algorithm has to deal with large accelerations and rotations in the projected image, because human wrist is highly dexterous and holding a scan-device worsens the stability of the hand.

These issues make any motion detection at a standard frame-rate of 25 Hz rather complex.

Image-based localization techniques, like visual SLAM, have proven to be accurate and robust for relative slowly moving systems. A robot which has only a bounded acceleration can not change rapidly its velocity. Therefore, its motion and position can be modeled quite reliably. However,

E. Mair (✉) · D. Burschka
Institute of Robotics and Embedded Systems, Technische
Universität München (TUM), Boltzmannstr. 3, 85748 Garching
bei München, Germany
e-mail: Elmar.Mair@cs.tum.edu

K.H. Strobl · T. Bodenmüller · M. Suppa
Institute of Robotic and Mechatronics, German Aerospace Center
(DLR), P.O. Box 1116, 82230 Wessling, Germany

K.H. Strobl
e-mail: Klaus.Strobl@dlr.de



Fig. 1 The upper *left-hand side picture* shows the DLR 3D-Modeler with the IMU module and the PC running the software. The other two images demonstrate each a global registration variant for the scan device: a robot arm (*lower left*), and an external tracking system (*right*)

it is difficult to find correct model descriptions for systems with high camera dynamics. Without restrictive assumptions about the dynamics, the uncertainties are too high and, therefore, a Kalman or particle-based probabilistic filtering would be too weak to provide significant robustification.

In the case of the *DLR 3D-Modeler* [16], in the past the global registration has been performed by external systems. Six reference spheres are mounted on the scanner and an external stereo camera system tracks them in real time (two ARTtrack1 cameras¹). Another approach is to mount the device on the wrist of a robot arm so that the position can be obtained by the inverse kinematics (see Fig. 1). Due to the already mentioned restrictions in the workspace and the high price of such external sensors, the idea arose to develop an image-based visual localization system, which provides a global registration by the two cameras mounted on the 3D-Modeler itself. Such an algorithm has to be online capable and able to deal with the imposed conditions of a human's wrist, which are two conflicting constraints for tracking. Further, no artificial modifications of the environment should be necessary to allow modeling without post-processing and quick changes of the workspace.

In this paper, an image-based visual localization technique, which runs in real-time on a small workstation without GPU support or other specialized hardware is presented. In the next section, we provide an overview on related work. In Sect. 3 the modules of the algorithm are described: the image processing and stereo initialization, the IMU-based landmark propagation, the robustification of the motion estimation and the feature set management. Some experimental results illustrate the robustness and the accuracy of the algorithm in Sect. 4 and compare it to the prior global localization methods used for the DLR 3D-Modeler. A discussion in Sect. 5 concludes the paper.

¹<http://www.ar-tracking.com>.

2 Related Work

Even though laser-scanners are still used more often for localization and SLAM than cameras, the computer vision community, working on visual SLAM (VSLAM) grows quickly. The camera projection results in a reduction of the dimensionality of data which needs to be recovered from multiple images.

The VSLAM solution provided by Davison consists of building a probabilistic 3D map with a sparse set of good landmarks to track [3]. The points are used in an Extended Kalman Filter (EKF) for a repeatable localization with limited drift. The amount of landmarks is restricted to permit real-time processing (in [3] a limit of 100 features is set) and another drawback is the necessity of a high feature quality. A single drifting feature can prevent other features to be found again and may also lead to a wrong landmark initialization.

Klein was able to achieve even more accurate results as the EKF based approach of Davison by efficiently separating the tracking and the mapping routines [7]. Further, the keyframe-based SLAM method scales better and even if it is designed for small workspaces, it is also applicable for larger environments. By introducing edgelets, the robustness towards rapid camera motions has been additionally increased [8]. The drawbacks are the high processing and memory requirements for the keyframes and the large number of features.

Eade and Drummond describe a monocular visual Fast-SLAM approach [4], which has no restrictions on the number of landmarks in the map. Due to their particle based method, only those features used for pose estimation have to be processed. The relation to the residual landmarks gets lost, which results in the estimation bias accumulated solely by the local features. Therefore, while it improves processing times, it limits loop closure capabilities.

The approach of Mouragnon et al. [10] provides a good accuracy even at larger distances. This is achieved by a local bundle adjustment, which improves the consistency while not impeding online processing (with a framerate of 7.5 fps). Local features are tracked between the image pairs and keyframes are selected from time to time, according to a minimum necessary feature intersection. These key-images are then used for stereo triangulation and local bundle adjustment. The drawback of this algorithm are curves—features get rapidly lost and many key-frames have to be selected which slows down processing and may decrease the possible framerate.

Mourikis and Roumeliotis developed a dual-layer localization architecture [11]. A combination of a Multi-State-Constraint Kalman Filter (MSC-KF) and a Bundle Adjustment yields highly accurate long-term visual pose estimation. The processing time of their MSC-KF is about 100 ms,

which slows down processing, but in combination with an inertial measurement unit (IMU) they achieved a robust long-range localization.

Thanh et al. present a stereo SLAM method with two EKFs [17]. It is based on a combination of mono- and binocular feature estimation. The stereo-EKF is initialized by odometry and in addition the stereo-matching is done with support from SLAM. The landmarks are then used for monocular pose estimation. By this rather complex stereo algorithm, the accuracy and robustness of conventional MonoSLAM is clearly improved, especially for long range motion. However, stereo processing is time-consuming and hence the maximum possible framerate is about only 9 Hz.

3D-modeling is a wide field in computer vision and with increasing interest due to its low costs and its improving accuracy. There are several possible solutions proposed in literature, covering diverse sensors and techniques. A few of them are mentioned in following. With DAVID² a low-cost laser-stripe profiler is provided, but artificial objects in the environment are needed. In [13] an offline 3D reconstruction method based on uncalibrated images is described. A self-referenced, hand-held 3D sensor is presented in [6], where reference points from a fixed mounted laser are used for pose estimation. However, there appears no online, passively self-referenced and image-based 3D-modeling system without artificial markers in literature—a more comprehensive survey is available at [15].

3 Real-time Image-based Localization

In the online 3D-modeling scenario, we need to compute the pose of a hand-held scanner in real-time. Due to the high dynamics of such a hand-held device, the choice of the correct motion model for a Kalman fusion is problematic. If a too restrictive motion model is used, quick motions of the user's hand are interpreted as wrong measurements, while a too permissive setting does not suppress the measurement noise properly. In such systems, smoothing should be applied only at the data fusion level, because only there smooth values may be assumed. Therefore, our algorithm was developed with the aim for high accuracy and robustness without probabilistic techniques, while satisfying the real-time constraint. In the next sections, we describe the concept and the single modules of the algorithm (see also Fig. 2).

3.1 Approach

In order to estimate the relative motion between the viewpoints of two images, first, relations amongst those images have to be found. The easiest and most popular way for it is

²<http://www.david-laserscanner.com/>.

to find patch correspondences. We chose the Kanade-Lucas-Tomasi (KLT) tracker [14] for landmark tracking due to its speed and robustness. Shifting the camera in a direction perpendicular to the optical axis during the close-range scanning lets the features move rather fast in the images. Therefore, several extensions were necessary to provide real-time tracking for such fast moving landmarks (see Sect. 3.2).

The motion can be calculated in several ways using those feature correspondences in case of calibrated cameras. The most common in the literature are probably the various point algorithms, especially the 8- [5] and 5-point [12] algorithms. We choose an iterative approach, namely vision-based GPS (VGPS) [2], for pose estimation. The iterative nature of the algorithm allows an effective robustification, which is described in detail in Sect. 3.4.

VGPS requires the knowledge of the exact 3D position of the landmarks for the initial image. Therefore, the landmarks in the first image need to be initialized with their true 3D-distances. While the landmarks are tracked from image to image, the pose is always calculated with respect to the first image, for which the 3D location of the features is known. This initial distance estimation can be done by moving the camera and applying a *Structure from Motion* (SFM) approach or by a second camera (*Structure from Stereo* (SFS)) [5]. Any monocular technique has an unknown scale which can not be resolved without further assumptions. Therefore, a second camera is used for the only purpose of landmark initialization. A permanent stereo triangulation would prevent real-time processing. However, the more accurate the initialization is, the more accurate will also the pose estimation be. A novel subpixel-accurate, fast matching, which is based on KLT tracking and an epipolar search, has been implemented [9]. Hence, we can combine the advantage of fast monocular tracking and subpixel-accurate stereo triangulation.

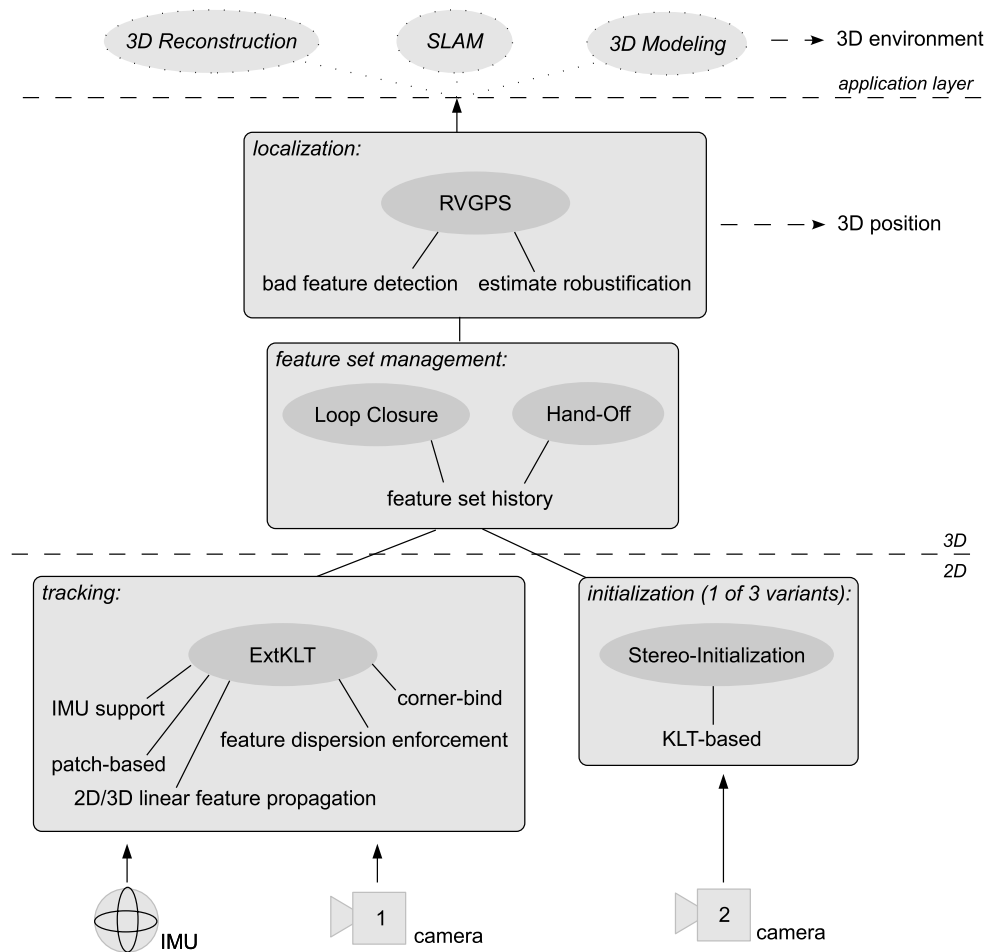
One important issue still needs to be solved: Due to close-range scanning the features are supposed not only to move fast, but also to leave the field of view rather quickly. Thus, new feature sets have to be acquired and initialized to provide permanent tracking. A feature set management has been implemented, which consists not only in an intelligent hand-off, but also in a recovery of already used and initialized feature set. For more details, please refer to Sect. 3.4.

3.2 Patch-based Real-time Tracking

KLT is a fast local tracker, but some improvements were still necessary to achieve a 25 Hz image processing on a standard workstation without a GPU. In the following the main extensions to the standard implementation of the KLT algorithm are described:

- The preprocessing and the image convolution has been restricted to small patches around the tracked points to

Fig. 2 The modules used for image-based localization



- speed up tracking. This makes also tracking on images with high resolution feasible.
- A linear motion model allows not only a larger feature displacement between the images but reduces also the search area. The results are fewer tracking iterations and smaller image-patches.
 - An improved pose estimation is achieved through a good landmark distribution in the image. We enforce such a good distribution by splitting the image into commensurate subimages for feature selection.

3.3 IMU Supported Tracking

Initial experiments have shown, that even with 2D feature propagation it is rather difficult to keep the wrist calm enough to not lose all landmarks. The need for a simple and passive sensor with a high test frequency led to the integration of an inertial measurement unit (IMU). The resulting rotation R^{IMU} can be easily applied to landmark propagation by using the measured rotational speed—indeed, a calibrated camera is basically an angular sensor. Thus, the ro-

tation propagation is computed for each unit vector r representing the direction to the tracked point.

$$\hat{r}_{t+1} = R_{t+1}^{IMU} r_t \tag{1}$$

The linear translational propagation in 2D, d_{T_t} , is calculated by the difference from the current measured feature position m_t to the prior one, but where the rotational component, measured by the IMU, is already removed:

$$d_{T_t} = m_t - d_{R_t} \tag{2}$$

where d_{R_t} is the feature displacement on the image plane resulting from R_t^{IMU} . Finally, the landmark propagation consists of

$$\hat{m}_{t+1} = m_t + \hat{d}_{R_{t+1}} + d_{T_t} \tag{3}$$

where $\hat{d}_{R_{t+1}}$ is the pixel displacement as difference of the projections of \hat{r}_{t+1} and r_t . Therefore, the rotation becomes propagated based on the IMU measurement while the translation is predicted in 2D. The measurement and motion propagation are consequently completely separated which prevents an accumulation of the estimation errors.

However, using this prediction step, the bottleneck of successful tracking during fast motions (up to 75°/s and

0.5 m/s) are not anymore the large feature displacements between two images (up to 50 pixels), but mainly the motion blur.³ The shutter time of the cameras can not be arbitrarily reduced, because short exposure times force the camera gain to be increased, which amplifies the image noise and impedes proper tracking.

3.4 Feature Set Management

Visual GPS is a sequential algorithm for estimation of transformation [2]. In several iterations the rotational and the translational component of the motion are estimated independent from each other. This iterative approach allows an easy weighting of the different rays according to their re-projection error. For the weighting, M-Estimators with the biweight Tukey-function are used, because of their continuous derivatives and their convenient weights. The modifications to the original VGPS algorithm are briefly sketched in the following—for a more detailed description of the innovations, please refer to [9].

Fast movements and close objects cause the landmarks to leave quickly the camera's field of view. However, well before the number of features becomes insufficient for pose estimation, the calculation becomes inaccurate. Whenever all points lie within a small region of the image, and especially if that region is on the image border, any image based pose estimation becomes inaccurate. The translational and rotational components cannot be separated properly and the influence of noise increases. Thus, we identified two cases, where a new set of up to 25 features should be initialized by the stereo triangulation:

- Each time, when the number of trackable features falls below a certain threshold.
- Whenever the features lie at the image border and prevent accurate pose estimation and increase the risk to get all lost soon.

The first case can be solved by a simple minimum threshold on the number of features. This is chosen based on the minimum amount of landmarks necessary for pose estimation, which is 3.

The second issue is addressed by calculating the feature set's centroid and projecting it onto the image plane. It represents the mean feature location and if it leaves the inner area of the image, it means that most features moved also to the border. However, this set-centroid plays an other important role, too.

Typically, several feature sets are used during a scan process. A possible bias is accumulated each time a new set is

acquired. This error is removed if an old set is reused. Additionally, bad and drifting features are removed from the sets as soon as they are detected by our robustified VPGS (RVGPS). These two facts favor the reuse of feature sets, which moved outside the field of view but become trackable again. Without knowing the quality of features, the most robust and best conditioned pose estimation can be achieved if the landmarks lie around the image center. We use again the projection of the feature set's centroid to determine which set to use for the next tracking step. Every time, the projection of the centroid of an other feature set is more central than the current one, we switch to those landmarks.

4 Experiments

Before we compare the presented algorithm to the existing global registration methods, some results depicting the accuracy improvements through robustification and feature management are presented.

4.1 IMU Fusion

Significantly faster rotations are feasible if the tracking is supported by an IMU. Figure 3 depicts a fast sweep of the camera with an average rotation of 11° and a maximum of 14° per frame. A maximum feature displacement of about 45 pixels has been measured. Nevertheless, the landmark location can be estimated accurately: the yellow squares (feature propagations) match the red circles (tracking results). The exact feature propagation makes also a change of the active feature set during such motions feasible. Good matches of the red circles and the yellow circles (projections of the landmarks according to their 3D location and the camera pose) confirm an accurate stereo-based initialization.

4.2 RVGPS Accuracy

The DLR 3D-Modeler is mounted on the wrist of a robotic manipulator KUKA KR16⁴ for this accuracy examination. The robot's kinematics is used as a ground truth with a relative error of 0.1 mm in position and 0.1° in rotation. Based on that data, the accuracy of the visual navigation algorithm can be computed. In the experiment, the sensor is moved half around the object and back again. 710 images were processed on a total path of 125 cm. The orientation changed up to 55° . The images have been processed twice: with the robustification enabled and disabled. Figures 4 and 5 illustrate the errors in rotation and translation. The blue dots represent the RVGPS output and the red dot VGPS without robustification.

³A short video of such fast motion is available at http://www6.in.tum.de/~maire/videos/3dMo_imuTracking.avi.

⁴<http://www.kuka.com/>.

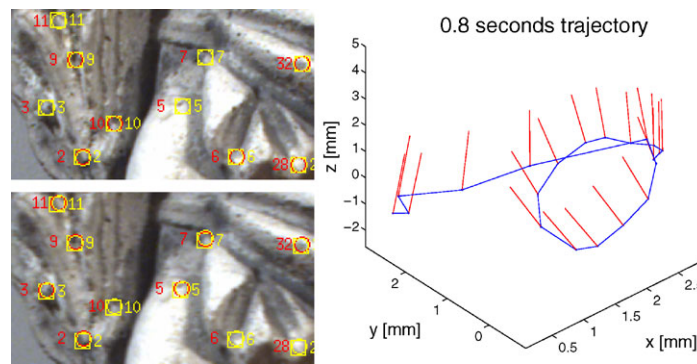


Fig. 3 (Color online) The left two images show two screenshots acquired at the beginning and at the end of a fast camera motion (up to 14° per frame). The yellow squares mark the IMU-supported propagation of the landmarks, while the red circles are the tracked features.

The yellow circles are the reprojection of the features according to the 3D structure. The right picture illustrates the trajectory (blue) and the camera viewing directions (red)

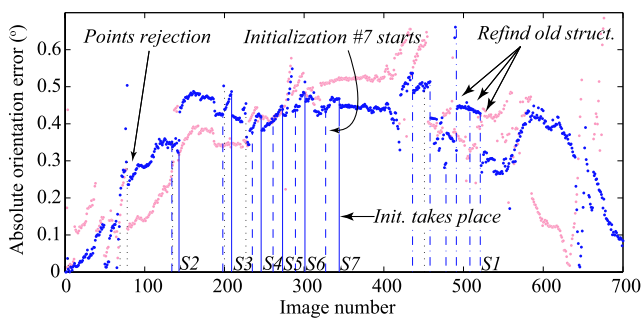


Fig. 4 (Color online) Residual orientation error with respect to the KUKA manipulator. Blue: using RVGPS. Pink: using VGPS

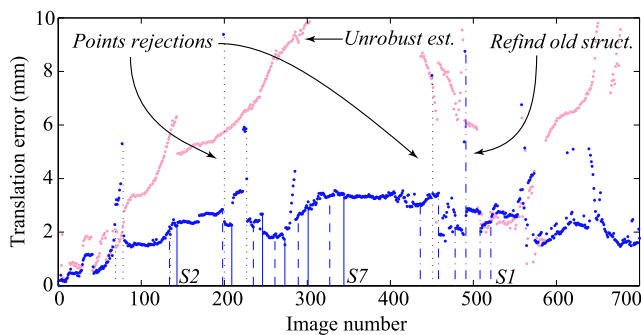


Fig. 5 (Color online) Residual translation error with respect to the KUKA manipulator. Blue: using RVGPS. Pink: using VGPS—the error rises up to more than 20 mm, which is not visible to preserve an adequate scale

Disregarding the outliers in the performance charts, the translational error increases up to 4 mm and the maximum accumulated rotational error was 0.5°. The outliers are the result of bad, drifting features which are detected by RVGPS and removed from further processing.

In the run with standard VGPS, the orientational error remains almost the same, while the translation estimation accumulates an error up to 20 mm. This can be explained

by the nature of such an iterative gradient descent method as VPGS: A difference in the measured landmarks and the reprojection can always be lessened by zooming out the camera. The algorithm simply follows the gradient on the optical axis of the camera to diminish the error, while the gradient in the rotational space may be too flat to be concerned.

Further, the effect of the feature set management becomes apparent in both runs. The error which is accumulated at each switch of set on the way to the end of the castle becomes again removed moving back to the initial position. The old feature sets are refound and used for the pose estimation on the way back. The error accumulated between the changes is removed again and the final residual error can be explained by the inaccuracies in the robot kinematics and tracking drifts.

4.3 Accuracy of the Registration Alternatives

In the following experiment, the ART-system, a robot arm (KUKA KR16) and our visual navigation approach are compared (see Fig. 1). Therefore, the 3D-Modeler has been mounted again on the wrist of the KR16 and while measuring the poses with the ART-system and the robot's kinematics, the images for the visual localization algorithm have been acquired. To simulate an ordinary scan sweep, the robot arm was commanded to move down 40 cm and up again in front of a putto shown in Fig. 7. Figure 6 illustrates the estimated trajectories of the scanner and the absolute rotational and translational difference of the ART-system and the visual navigation to the KUKA robot. While with equally scaled axis the difference in the trajectories is not really recognizable, the zoom in the top figure makes it visible. Due to the inaccuracies of the robot-to-camera (hand-eye) and ART-to-camera calibration, a small error between the results of the robot and the ART exists. This error is also illustrated in the blue

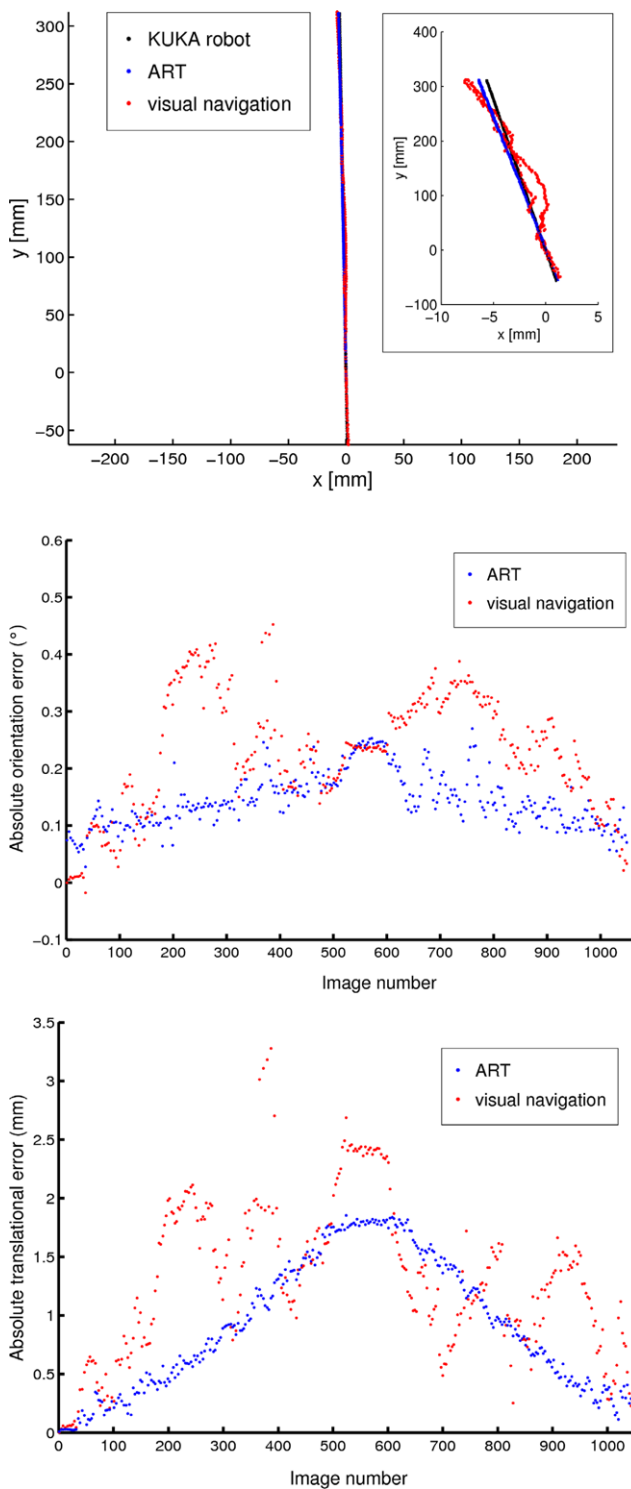


Fig. 6 (Color online) *Top*: The scan trajectories computed by three different references. Only by zooming in the *x*-axis (*small plot*) the error becomes evident. The absolute rotational error (*middle*) and the absolute translational error (*bottom*) are referenced to the KUKA poses

dots of the bottom plots, where the KUKA robot estimates are used as ground truth due to its high accuracy. The visual localization result is noisier, but the average er-



Fig. 7 *Left*: the scanned putto-statue. *Middle*: triangle-meshes based on the ART-system. *Right*: triangle-meshes based on the visual localization. All measurements were done in real-time and also the meshes where computed online

ror is comparable to the ART-system. Five landmark sets were acquired during the sweep and refound on the way back.

During longer sweeps without loop closure, the visual navigation results are drifting like in every SLAM-based method, while the ART-system provides drift-less results within its workspace. This is the most problematic drawback regarding external registration devices and the price for compactness, low costs and especially arbitrary workspace location and size. However, with this method, the error is not accumulated from image to image, but only with every new landmark set.

Figure 7 shows online computed triangle-meshes for the statue on the left, once acquired with the ART-based global registration (middle) and once with the presented visual navigation approach (right). The 3D-Modeler was moved by hand down and up in front of the putto, in the same way as in the experiment with the KUKA.⁵ The distance to the object is measured by a laser-stripe profiler (LSP) and the acquired data is fused, together with the global pose, to 3D-meshes [1]. Due to the smoothing by the data fusing process, the noise in the visual navigation estimates is removed and the results are comparable. The differences in the meshes result from the fact that two hand-held runs were necessary to evaluate the *online* processing performance, which yields unequal views for the LSP.

Regarding the quality of the pose estimation, following points are worth mentioning: The visual navigation is well-conditioned if the features are spread over the whole image. Further, accurate results are provided if the landmarks are accurately initialized, which depends on the features' dis-

⁵An example of the operation is available at http://www6.in.tum.de/~maire/videos/3dMo_liveModeling.avi.

tance and the base-line of the cameras. The ART-system has the systematical drawback of its inaccurate calibration capabilities. It is difficult to calibrate the setup only by the six small and tight spheres, which are quite far away from the cameras. If the markers are occluded by the sensor device or by the object to scan, the accuracy of the pose estimation is massively reduced. Further, the error of the LSP and the ART do not rely on the same images and thus the error correlation is worse compared to the visual navigation. The robot arm is the most precise of the three variants, provided an accurate hand-eye calibration. However, it lacks the same drawbacks as the ART-system and furthermore the degrees of freedom of the scan device are reduced at the workspace border.

5 Conclusions and Future Work

In this paper, a hand-held 3-D modeling device for close-range applications has been presented that references itself passively from its own images in real-time. The high flexibility of a human wrist makes the application of conventional VSLAM techniques problematic. Therefore, a non-probabilistic approach for vision-based egomotion estimation has been implemented. An accurate pose estimation in real-time could be achieved by speeding up and supporting the image processing with IMU-measurements and by robustifying the motion estimation. Experimental results validated the algorithm and the method was compared to other common external registration alternatives. Low costs, compactness, easy handling and no workspace restrictions have been achieved for the price of a small drift over longer sequences in cases where no loop closure was possible.

In future, an absolute reference for the scanner should be provided by using global trackable landmarks for initialization. Further, we plan to use the IMU information not only for landmark propagation but also for a Kalman-based fusion to provide a more accurate, robust and smooth pose estimation.

Acknowledgements This work is supported in part by the DFG excellence initiative research cluster *Cognition for Technical Systems* (CoTeSys) and by the *German Aerospace Center* (DLR).

References

1. Bodenmueller T, Hirzinger G (2004) Online surface reconstruction from unorganized 3d-points for the DLR hand-guided scanner system. In: Proc. of 2nd symposium on 3D data processing, visualization, transmission
2. Burschka D, Hager GD (2004) V-GPS(SLAM): vision-based inertial system for mobile robots. In: IEEE international conference on robotics and automation
3. Davison AJ, Reid ID, Molton ND, Stasse O (2007) Monoslam: real-time single camera slam. IEEE Trans Pattern Anal Mach Intell. doi:[10.1109/TPAMI.2007.1049](https://doi.org/10.1109/TPAMI.2007.1049)
4. Eade E, Drummond T (2006) Scalable monocular slam. IEEE conference on computer vision and pattern recognition
5. Hartley R, Zisserman A (2003) Multiple view geometry in computer vision. Cambridge University Press, Cambridge
6. Hebert P (2001) A self-referenced hand-held range sensor. In: Proceedings of the third international conference on 3-D digital imaging and modeling, 2001
7. Klein G, Murray D (2007) Parallel tracking and mapping for small AR workspaces. In: Sixth IEEE and ACM international symposium on mixed and augmented reality (ISMAR'07)
8. Klein G, Murray D (2008) Improving the agility of keyframe-based SLAM. In: 10th European conference on computer vision (ECCV'08), pp 802–815
9. Mair E, Strobl KH, Suppa M, Burschka D (2009) Efficient camera-based pose estimation for real-time applications. In: IEEE/RSJ international conference on intelligent robots and systems
10. Mouragnon E, Dekeyser F, Sayd P, Lhuillier M, Dhome M (2006) Real time localization and 3d reconstruction. In: IEEE conference on computer vision and pattern recognition
11. Mourikis AI, Roumeliotis SI (2008) A dual-layer estimator architecture for long-term localization. In: Workshop on visual localization for mobile platforms. IEEE computer society conference on computer vision and pattern recognition
12. Nister D (2004) An efficient solution to the five-point relative pose problem. IEEE Trans Pattern Anal Mach Intell. doi:[10.1109/TPAMI.2004.17](https://doi.org/10.1109/TPAMI.2004.17)
13. Pollefeys M, Koch R, Vergauwen M, Gool LV (1999) Hand-held acquisition of 3D models with a video camera. In: IEEE 2nd international conference on 3D digital imaging and modeling
14. Shi J, Tomasi C (1994) Good features to track. In: IEEE conference on computer vision and pattern recognition
15. Strobl KH, Mair E, Bodenmüller T, Kielhöfer S, Sepp W, Suppa M, Burschka D, Hirzinger G (2009) The self-referenced DLR 3D-modeller. In: IEEE/RSJ international conference on intelligent robots and systems
16. Suppa M, Kielhöfer S, Langwald J, Hacker F, Strobl K, Hirzinger G (2007) The 3d-modeller: a multi-purpose vision platform. In: IEEE international conference on robotics and automation
17. Thanh TN, Sakaguchi Y, Nagahara H, Yachida M (2006) Stereo slam using two estimators. In: IEEE international conference on robotics and biomimetics