Minimizing Odometry Drift by Vanishing Direction References

Tobias Schwarze and Martin Lauer
Institute of Measurement and Control
Karlsruhe Institute of Technology
Karlsruhe, Germany
Email: tobias.schwarze@kit.edu

***Abstract*—Robotic systems that operate in environments without access to global references like GNSS normally estimate their position by odometry. The incremental nature of this measurement is subject to drift that accumulates to large errors in the position estimate. We minimize this drift by exploiting scene knowledge in form of vanishing directions. These provide a scene referenced orientation measurement which we apply in an orientation filter to correct the estimated attitude and in turn improve the position estimation. The experiments show that the resulting accuracy is comparable to a referenced inertial measurement unit and that our method can thus augment or replace such sensors.**

I. Introduction

Robotic systems navigating indoor environments are usually bound to estimate their position incrementally by integrating odometric measurements. Without global reference this estimation is inevitable subject to drift. Only if the infrastructure is augmented with active or passive landmarks in known positions, an absolute position estimate up to a limited error can be obtained by triangulation. To treat accumulating drift, self-contained systems often combine odometry with compass readings as heading reference and augment them with gyroscopes to increase system dynamics and filter magnetic field disturbances. An orientation filter is then applied to fuse the measurements and keep track of the sensor offsets. In return, the accuracy of the position estimation increases strongly, since the largest part of the accumulating error can be avoided by correcting the orientation error. In this work we present an approach to gather such reference readings by a vision sensor. This can be helpful as additional reference measurement to augment the orientation filter, or avoid placing additional expensive and sensitive inertial hardware onto vision only platforms.

Vision only based robotic systems traditionally treat the problem of pose estimation by simultaneously mapping the environment in a structure from motion scheme. The mapped environment is then used to localize the system [1], [2]. Drift is minimized by landmark observations tracked over long periods and can be totally corrected when loop closures happen where the robot returns to a previously mapped part of the scene. If the desired measurement is only the position and orientation of the robot, this comes with the overhead of estimating, storing and maintaining an environmental map.

The effect of measuring fixed landmarks in the environment as orientation references can be achieved in a similar way by measuring scene vanishing directions. Vanishing directions can be understood in the same way as compass or direction of gravity readings, with the difference that vanishing directions are not global references but fixed to the scene or environment. In the scenario of indoor odometry this is no restriction since localization relative to the building is desired, which in this case provides the references. Such direction references have been used to correct heading errors in navigation based on laser scanners [3], or systems building upon cameras augmented with inertial measurement units [4], [5], [6].

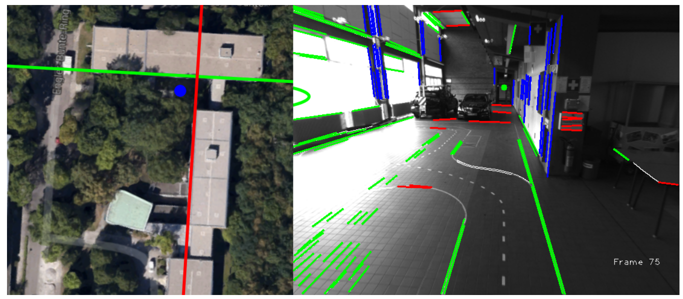


Fig. 1. Three vanishing directions typically found in indoor scenarios. A measurement of these directions provides a scene fixed reference that we apply to correct the drift in odometric position and orientation estimation.

In most indoor scenarios such vanishing directions are known upfront as orthogonal to each other. In cases were this assumption does not apply they can be obtained with little effort. A satellite image might be sufficient since the inner structure with walls and floors is not needed. We apply this kind of a-priori knowledge to robustly track all vanishing directions in the scene. Afterwards we present a minimal orientation filter to correct the estimated orientation from an odometric sensor with all measured directions.

As a challenging application we use binocular visual odometry as base to estimate the pose of a camera in all 6 degrees of freedom. The experiments in a typical building with hallways and stairs show how this leads to an accurate, building referenced position estimate.

II. Methods

A. Estimation of vanishing directions

Given a camera system calibrated using a pinhole model, lines in 3D-space are projected to lines in the image plane. The intersection of projected parallel space lines is known as vanishing point.

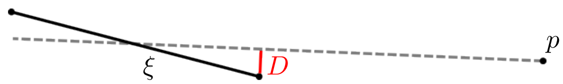


Fig. 2. Error D between edgelet ξ and vanishing point p.

In the calibrated camera case, each vanishing point defines a direction vector $\mathbf{n}(\theta, \varphi)$ originating from the focal point of the camera. We estimate this vector following the approach of Tardif [7]. Based on a Canny edge detector, edge crossings are suppressed and connected points extracted by flood fill region growing. The edge hypotheses are split into straight segments. The remaining segments longer than a minimal length (e.g. 30px) are fitted with a line model and constitute the edge list ξ.

From ξ we seek the subset of edges ξ_{VP}, which support the expected direction of the vertical vanishing point. To evaluate the support of an edge ξ for a given vanishing direction $\mathbf{n}_{VP}$ we define its error as the orthogonal distance $D(\mathbf{n}_{VP}, \xi_j)$ of one of the line endpoints to the line connecting the vanishing point with the edge centroid (Figure 2). All edges with D smaller than a threshold ϵ are used to update the vanishing direction by minimizing

$$\mathbf{n}_{VP}^{+} = \min_{\mathbf{n}_{VP}} \sum_{\xi \in \xi_{\mathrm{VP}}} D(\mathbf{n}_{VP}, \xi) \tag{1}$$

Internally, we represent the vanishing direction $\mathbf{n}_{VP}$ with spherical coordinates (θ, φ) and also perform minimization in this domain.

Multiple vanishing directions Typical indoor environments with non-curved hallways consist of three vanishing directions, of which the forward directed is the most prominent and depicted with its vanishing point in the visible image plane. Besides, the vertical direction which corresponds to the vector of gravity is depicted in many vertical edge segments and usually always measurable (compare e.g. Figure 3(e)). The third direction is usually orthogonal pointing sideways. The relation between these directions is independent of the position of the observer, hence, the angles between the directions do not change within the building.

The estimation of multiple vanishing directions requires a preceding classification step to assign edge observations to vanishing directions. To this end we evaluate the distance function $D(\cdot)$ for all directions and do nearest-neighbour assignment. In our experiments this appeared to be sufficient in comparison to more complex soft-classification schemes like expectation maximization [8].

Instead of minimizing the cost functions (1) for each vanishing direction VP_j with the classified support edgelets ξ_{VP_j} and enforcing the known angles between vanishing directions as optimization constraint, there is an alternative and more intuitive solution. Since the spatial angles between the vanishing directions are fixed, we do not optimize the vanishing directions directly, but instead seek the spatial rotation R that needs to be applied to all vanishing directions in order to minimize their cost functions. This is expressed as

$$\min_{\theta,\phi,\psi} \sum_{VP_j} \sum_{\xi \in \xi_{VP_j}} D(\ R(\theta, \phi, \psi)\ \mathbf{n}_{VP_j}, \xi) \tag{2}$$

The updated vanishing directions are then obtained from

$$\mathbf{n}_{VP_j}^{+} = R(\theta, \phi, \psi)\ \mathbf{n}_{VP_j} \tag{3}$$

This way we have to optimize 3 rotation parameters for arbitrary many vanishing directions instead of 2 parameters for each direction in an independent optimization. Furthermore, we benefit from the fact that each edgelet contributes to each direction. This leads to more observations for less parameters and consequently to a much more robust estimation process.

B. Visual Odometry

In general, any odometric sensor with known extrinsic calibration w.r.t. the camera can be applied to gather an incremental position update. In this work we demonstrate that the deployment of such additional sensors can be avoided in camera setups and visual measurements alone suffice to estimate the position with minor drift. An incremental motion update solely from image data is commonly referred to as visual odometry (VO). The general goal is to find the transformation in all six degrees of freedom that relates the camera poses of frame $k-1$ and k. Most methods build upon salient image feature points tracked over consecutive frames and minimize their reprojection error. Various standalone implementations exist, an overview can be found in [9].

In this work we employ the stereo variant of libViso2 [10]; the frame to frame transformation is provided as a translation vector $\mathbf{t}$ and a 3x3 rotation matrix R which can be written as an affine transformation $T = \begin{bmatrix} R_{vo} & \mathbf{t}_{vo} \\ 0^{\mathrm{T}} & 1 \end{bmatrix}$. The total transformation T_G applied to the camera increments with each frame by $T_{G,k} = T_{G,k-1} T^{-1}$, respectively

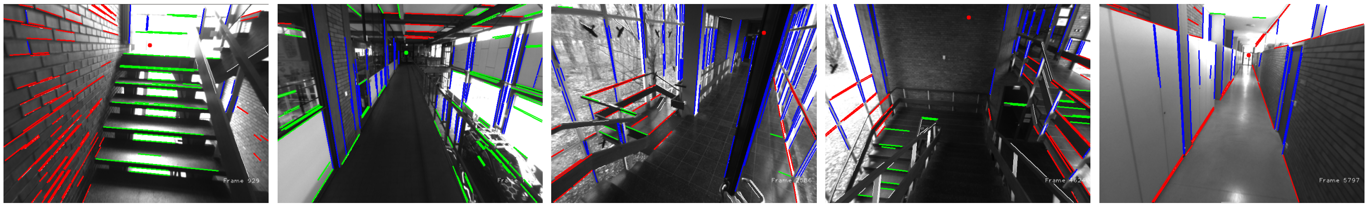
$$R_{G,k} = R_{G,k-1} R_{vo}^{\mathrm{T}} \tag{4}$$

$$\begin{aligned} \mathbf{t}_{G,k} &= \mathbf{t}_{G,k-1} + (-R_{G,k-1} R_{vo}^{\mathrm{T}} \mathbf{t}_{vo}) \\ &= \mathbf{t}_{G,k-1} - R_{G,k} \mathbf{t}_{vo}. \end{aligned} \tag{5}$$

Since this estimation is done incrementally it is inevitably subject to drift like any other odometric estimate.

C. Information fusion

The aim of fusing visual odometry with the vanishing directions is to include a referenced measurement of camera attitude. The vanishing directions do not give any information about the camera position but offer an environment referenced attitude measurement. If two or more known vanishing directions are measured (by optimizing (2)), we can determine the orientation that is equivalent to the incrementally accumulated orientation R_{G_k} but free of drift. This could directly replace R_{G_k}, however, in case of an erroneous measurement of the vanishing directions the egopose estimation would be set to an erroneous attitude. The translation component t_{vo} would in consequence be incremented along an incorrect direction and lead to large errors in the position estimate. To prevent such errors we do not discard the incremental measurement R_{vo} completely but use it as a (non-referenced) attitude prediction. Then, we determine the predicted vanishing directions, compare them with the actual measurements and correct the



(a) (b) (c) (d) (e)

Fig. 3. Example shots from the image sequence. Positions are marked in Figure 4.

attitude to minimize their differences. The corrected attitude is afterwards basis for the translation update following (5). We model this in an extended Kalman filter and use a quaternion $\mathbf{q}$ to represent the attitude. The scene vanishing directions are fixed as (θ_G, φ_G) for each direction.

Prediction We use the visual odometry transformation T between the last frame $k-1$ and the current frame k to obtain the prediction of the current camera attitude in the global reference frame. R_{vo} is transformed into a rotation quaternion $\mathbf{q}_{vo}$. Equivalent to (4) we calculate

$$\mathbf{q}_k^- = \mathbf{q}_{vo}^* \mathbf{q}_{k-1} \tag{6}$$

using the Hamilton product.

Correction The measurement prediction for each global vanishing direction (θ_G, φ_G) from current attitude $\mathbf{q}_k^-$ is found by

$$h(\mathbf{q}_\mathbf{k}^-, \theta_G, \varphi_G) = \begin{bmatrix} \theta \\ \varphi \end{bmatrix} = g_{sph}\left(R(\mathbf{q}_k^-) \begin{bmatrix} \sin(\theta_G)\cos(\varphi_G) \\ \sin(\theta_G)\sin(\varphi_G) \\ \cos(\theta_G) \end{bmatrix} \right) \tag{7}$$

where $R(\mathbf{q})$ is the left-handed rotation matrix equivalent to the rotation quaternion $\mathbf{q}$ and

$$\begin{bmatrix} \theta \\ \varphi \end{bmatrix} = g_{sph}(\mathbf{n}) = \begin{bmatrix} \arccos(n_z) \\ \text{atan2}(n_y, n_x) \end{bmatrix} \tag{8}$$

the transformation between euclidean and spherical coordinates.

For each vanishing direction we apply one correction step with the local measurements (θ_j, φ_j) and obtain the updated orientation estimate $\mathbf{q}_k^+$.

To complete the data fusion we apply the translation update using the filtered orientation $\mathbf{q}_k^+$ according to (5) as

$$\mathbf{t}_k = \mathbf{t}_{k-1} + \mathbf{d} \tag{9}$$

$$(0, \mathbf{d}) = \mathbf{q}_k \ (0, -\mathbf{t}_{vo}) \ \mathbf{q}_k^* \tag{10}$$

III. Experiments

Our experimental setup consists of a calibrated stereo rig with a baseline of around 18 cm and wide-angle lenses of 3.5 mm focal length mounted on a helmet. We recorded a sequence of 7500 images with 30fps while we were walking a 240 meter loop through a building including two staircases connecting the two floors as well as glass doors in the corridors that had to be opened during the passage. Opening doors is a challenging situation for visual odometry due to the fact that a large part of the observed scene is moving and violates the static scene assumption.

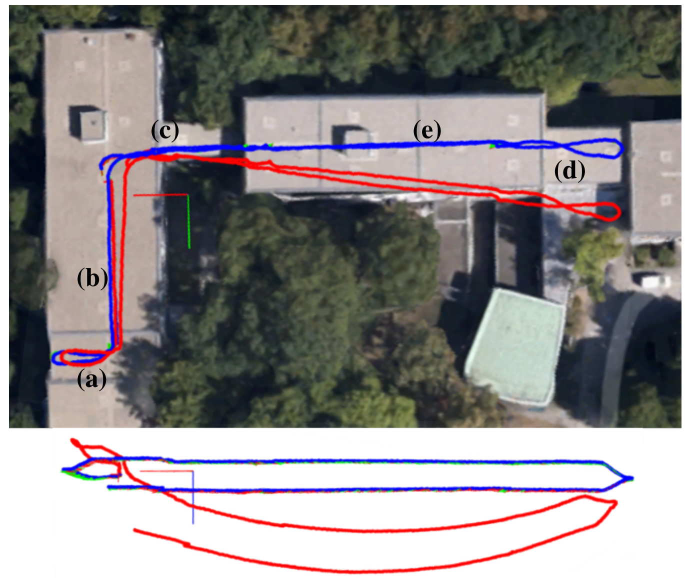


Fig. 4. Comparison of filtered trajectory in blue and plain visual odometry in red. The plotted coordinate system corresponds to the global vanishing directions. A slight drift can be seen in the upper top-view, the side-view in the bottom reveals a large drift in vertical direction.

To evaluate the estimation quantitatively we gathered the ground truth attitude from an Xsens MTi-300 inertial measurement unit which was calibrated externally to the cameras. The IMU corrects drift internally using the vector of gravity and the compass heading, as long as the unit is exposed only to short translational accelerations as in our case, the readings can be considered free of drift and are precise enough to be used as basis for comparison here.

To determine the noise in the measurement of vanishing directions we mapped the observations for all three directions into the global IMU frame, the projections onto the unit sphere are shown in Figure 5. The green and red horizontal vanishing directions show a similar noise pattern, which can be explained by the fact that the camera mainly rotates around the vertical axis during walking. The vertical vanishing direction (blue) varies equally around the gravity vector. We determine the variance for the direction parameters θ and φ from this mapping and use it in the Kalman correction step as measurement noise ($\sigma_{\theta,\varphi} = 1e^{-4}$).

The estimated path is plotted in Figure 4. The horizontal drift of uncorrected visual odometry (red path) becomes obvious on the long straight corridors (top), the vertical drift

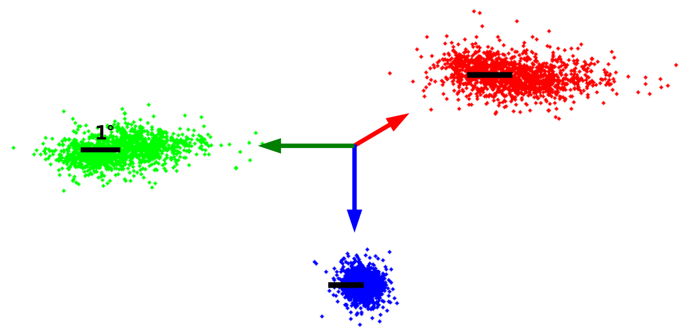


Fig. 5. Local vanishing direction measurements mapped into the global IMU frame.

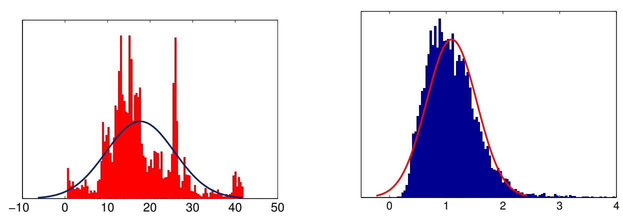


Fig. 6. Distribution of attitude error in degrees without (left) and with (right) vanishing direction reference compared to IMU ground truth. Mind the differently scaled x-axis of the plots.

appears even stronger in this sequence (bottom). The filter is able to correct this drift (blue path) by using the three orthogonal vanishing directions (corresponding to the axis of the plotted coordinate system) as environment fixed reference. The overall position error after closing the loop decreases from 3.6% to 1.2% using the filtered estimate, but more expressive is the estimated attitude error in the global frame. Compared to IMU ground truth the uncorrected incremental visual odometry deviates by up to 40° from the true attitude, after correction this reduces to a mean deviation of around 1° (Figure 6). The position increment benefits accordingly.

In our experimental scene the vanishing directions are orthogonal to each other. Note that this is not a requirement as e.g. assumed in [5] – any frame of vanishing directions can be applied as reference. At this point we do not treat the problem of finding these directions but assume them to be known a priori, which is often the case in constrained indoor navigation applications.

Compared to related approaches in [4] and [6] we benefit from all vanishing directions, including those currently not well visible. By tracking the whole frame of vanishing directions instead of re-detecting the current frontal direction we avoid to re-initialize the reference direction after taking a turn. Our filtered orientation estimate stays consistent with the global scene model also in these situations.

IV. Conclusion

We introduced a method to minimize the accumulating drift in odometric position and orientation estimation using visual scene references. We use a camera to measure the scene vanishing directions which allows us to correct the absolute orientation estimate of the system w.r.t. the environment.

Instead of detecting single vanishing points we proposed to track the whole frame of vanishing directions. This exploits knowledge about the scene already in the measurement process and enables us to robustly track all vanishing points, including those which are currently not visible.

An orientation filter was introduced which corrects the estimated attitude of odometric sensor systems with these reference measurements. This in turn greatly improves the incrementally estimated position of the system w.r.t. the scene. The evaluation in a typical office building shows that the approach can handle drifting odometric data of a free moving camera and reach an accuracy close to a referenced inertial measurement unit. Thus, it provides an option to augment or replace such additional reference sensors in platforms equipped with vision sensors. The remaining small drift in the estimated trajectory can be tied to incrementing errors in the translation component of the odometry, which can principally not be corrected using the vanishing directions.

References

[1] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.

[2] V. Nguyen, A. Harati, A. Martinelli, R. Siegwart, and N. Tomatis, "Orthogonal slam: a step toward lightweight indoor autonomous navigation," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, Oct 2006, pp. 5007–5012.

[3] Y. H. Lee, C. Nam, K. Y. Lee, Y. S. Li, S. Y. Yeon, and N. L. Doh, "Vpass: Algorithmic compass using vanishing points in indoor environments," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, Oct 2009, pp. 936–941.

[4] C. Kessler, C. Ascher, N. Frietsch, M. Weinmann, and G. Trommer, "Vision-based attitude estimation for indoor navigation using vanishing points and lines," in *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION*, May 2010, pp. 310–318.

[5] V. Huttunen and R. Pich, "A monocular camera gyroscope," *Gyroscopy and Navigation*, vol. 3, no. 2, pp. 124–131, 2012.

[6] L. Ruotsalainen, J. Bancroft, and G. Lachapelle, "Mitigation of attitude and gyro errors through vision aiding," in *2012 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2012, Sydney, Australia, November 13-15, 2012*, 2012, pp. 1–9.

[7] J.-P. Tardif, "Non-iterative approach for fast and accurate vanishing point detection," in *Computer Vision, 2009 IEEE 12th International Conference on*, Sept 2009, pp. 1250–1257.

[8] M. Antone and S. Teller, "Automatic recovery of relative camera rotations for urban scenes," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 2, 2000, pp. 282–289 vol.2.

[9] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *Robotics Automation Magazine, IEEE*, vol. 18, no. 4, pp. 80–92, Dec 2011.

[10] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, June 2011.