Pattern Analysis and Applications (2024) 27:41
https://doi.org/10.1007/s10044-024-01264-1

ORIGINAL ARTICLE

High-recall calibration monitoring for stereo cameras

Jaroslav Moravec[1] · **Radim Šára**[1]

Received: 30 October 2023 / Accepted: 12 March 2024

Abstract
Cameras are the prevalent sensors used for perception in autonomous robotic systems, but their initial calibration may degrade over time due to dynamic factors. This may lead to a failure of downstream tasks, such as simultaneous localization and mapping (SLAM) or object recognition. Hence, a computationally lightweight process that detects the decalibration is of interest. We describe a modification of StOCaMo, an online calibration monitoring procedure for a stereoscopic system. The method uses robust kernel correlation based on epipolar constraints; it validates extrinsic calibration parameters on a single frame with no temporal tracking. In this paper, we present a modified StOCaMo with an improved recall rate on small decalibrations through a confirmation technique based on resampled variance. With fixed parameters learned on a realistic synthetic dataset from CARLA, StOCaMo and its proposed modification were tested on multiple sequences from two real-world datasets: KITTI and EuRoC MAV. The modification improved the recall of StOCaMo by 25 % (to 91 % and 82 %, respectively), and the accuracy by 12 % (to 94.7 % and 87.5 %, respectively), while labeling at most one-third of the input data as uninformative. The upgraded method achieved the rank correlation between StOCaMo V-index and downstream SLAM error of 0.78 (Spearman).

Keywords Autonomous robots · Stereo cameras · Calibration monitoring

1 Introduction

Visual perception of robotic vehicle platforms such as self-driving cars or aerial vehicles relies on the knowledge of inter-sensor calibrations, especially in systems based on visual stereo matching. Although many accurate camera calibration methods have been developed [1], the accuracy of the calibration parameters can gradually deteriorate over time as the sensors are exposed to environmental conditions and external stresses. A decalibration may degrade the performance of a downstream visual task [2]. In an extreme case, an autonomous vehicle may need to shut down the downstream visual data processing system for critical safety when the sensors fail or decalibrate. However, frequent false alarms, causing operational delays, should be avoided. Therefore, the sensor system should include a self-assessment mechanism that monitors the calibration quality. The most important performance metrics are low false positive (false decalibration alarms) and false negative (not reporting actual decalibration) rates. The self-assessment algorithm should also run in real-time with low demand for computational resources and be optimized for the sequential character of data acquisition, with a short time to detection.

Other use cases include visual algorithm verification and testing or deep learning of perception modules. This is a frequent task in the automotive industry. These procedures employ big data but should avoid uncalibrated inputs while keeping as much data available for testing or learning as possible. Again, false positive and false negative performance metrics are essential, but real-time and online processing is not necessarily required.

By *decalibration*, we mean the mismatch between a reference calibration and the current data. This mismatch could be expressed as an error in some calibration parameters, such as the relative translation and rotation between a pair of stereo cameras. Unlike in automatic calibration, when all parameters need to be found, it is not crucial to monitor all

✉ Jaroslav Moravec
moravj34@fel.cvut.cz

Radim Šára
sara@fel.cvut.cz

[1] Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, Prague, Czech Republic

Published online: 13 April 2024

the parameters, as the decalibration typically manifests in several degrees of freedom.

It is unnecessary to recalibrate or track the values of these parameters explicitly when the goal is purely to *detect* a decalibration, not to correct it as in online infrastructure-based calibration methods. In this paper, we expand on this idea, introduced in [3], and develop a *calibration monitoring* method, an online statistical verification method for a specific sensor setup. Specifically, we focus on a stereoscopic camera pair used in stereoscopic vision or visual odometry (structure-from-motion, SfM) and Simultaneous Localization and Mapping (SLAM) algorithms. Decalibration affects the methods in two ways: (1) sparse correspondences and disparity map quality degrade with epipolar geometry error because the local image descriptors do not align with the epipolar lines, resulting in erroneous matches [4], (2) the disparity value does not correspond to the true (inverse) depth, leading to a distance estimation bias [5] that can quickly accumulate via the incremental character of the visual odometry [6].

2 Related work

The standard approach to camera calibration uses predefined targets of known dimensions, e.g. [1]. The targets are captured from various viewing angles, and parametric constraints are derived from the geometric relationships of the corresponding points. These methods offer high precision at the cost of longer and/or inconvenient execution and are suitable for obtaining the reference calibration. In this work, we focus on monitoring the extrinsic calibration parameters (the relative camera position and orientation [7]).

2.1 Automatic targetless calibration

In contrast to the above mentioned approaches, targetless (or self-) calibrations obtain calibration information from unstructured data. We divide the methods into several groups.

2.1.1 Correspondence-based

These methods use detected (or hand-picked) matched features to optimise the calibration parameters either alone (e.g., by utilising epipolar geometry) [8] or together with the 3D structure of the scene (calibration from infrastructure) [9]. The former is very fast but lacks precision. On the other hand, the latter is very precise, but the optimisation over thousands of parameters (using bundle adjustment [7]) can be computationally costly. Thus, some combination of the two is often used [10].

2.1.2 Odometry-based

These methods employ visual odometry estimation in each sensor separately, and then they find the transformation between them by solving the hand-eye (HE) calibration problem [11]. Although it does not require any common field of view of the sensors or any initial guess, the precision of the parameters is usually low. HE methods require sufficiently complex motions in 6D [12]. This is often infeasible in the automotive domain, where the vehicle is bound to the ground (plane).

2.1.3 End-to-end learning-based

The fast development of deep learning (DL) also considerably impacts the stereo self-calibration task. These methods differ in the way they employ the DL. For example, they can estimate the fundamental matrix [13] or optimise the consistency between monocular depth and inferred stereo depth [14]. They achieve good results at the cost of long training and inference and/or the need for high-performance hardware.

2.2 Online calibration tracking

Precise and stable approaches for targetless calibration are usually computationally expensive and can hardly be run during the operation of the sensor system. Hence, online calibration tracking methods that follow unpredictable changes in calibration parameters were proposed.

Following up on their previous work [10], the authors of [15] studied three geometric constraints for calibration parameters tracking with robust iterated extended Kalman Filter (IEKF). Using a reduced bundle adjustment, they achieved very accurate results in real-time, although dynamic objects hurt the stability of the approach in some environments. They found that combining epipolar constraints for instantaneous coarse calibration with their reduced bundle adjustment method yields stable and accurate results.

Lowering the effect of the dynamic objects in the scene was then studied in [16], where they used CNN to segment the pixels. Only static points were used in the optimisation. Besides epipolar geometry, the homographies induced by the ground plane also provide calibration information [17]. This requires image segmentation, too. Such segmentation could be obtained as a by-product of the downstream task preprocessing.

Online calibration of epipolar geometry parameters was considered in [18]. Detected image keypoints (possibly aggregated over several frames) were used to estimate the parameters by epipolar error minimization. The Kalman filter was then used to track the parameters. Hence, in

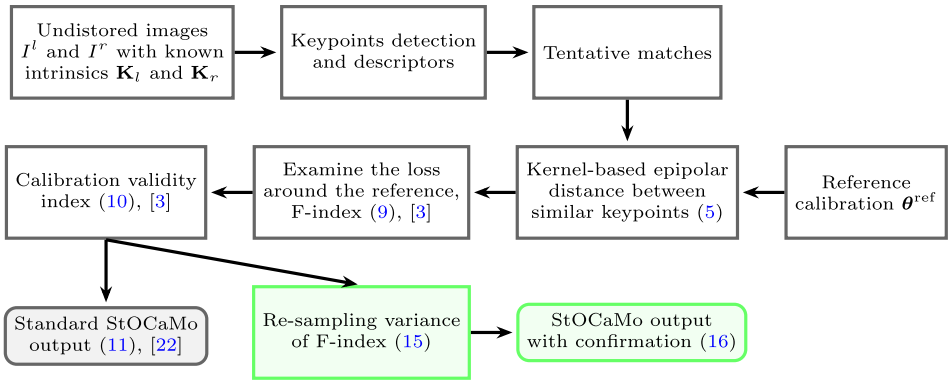


Fig. 1 A diagram of the StOCaMo method, which reports whether the visual system is decalibrated. A modification, which adds a confirmation step for the StOCaMo output, is in green. This confirmation is the main topic of this paper (color figure online)

contrast to [15], they did not require temporal matching of features. They show that the epipolar error is sufficient for accurate online recalibration (tested downstream for visual odometry and reconstruction).

2.3 Online calibration monitoring

Online calibration tracking methods provide good results, but changing the calibration during the system operation might be unsafe. For example, in the automotive domain, if the data fusion stops working due to a loss of multi-sensor calibration, the vehicle should undergo some authorised service rather than rely on parameters tracked during the drive. Online tracking is also quite expensive and cannot be run constantly as the computing resources are needed elsewhere. A more lightweight system could instead detect a miscalibration and trigger the calibration procedure only when needed.

Calibration monitoring as a research problem (specifically in the automotive domain) was introduced in [3]. They detected the LiDAR-Camera system extrinsic miscalibration by examining the alignment between image edges and projected LiDAR corners (points with large radial distance differences with neighbours). As the method had no memory, it could validate the calibration parameters on a single frame without any temporal tracking of the parameters.

The detection of decalibration for intrinsic parameters of a single camera was studied in [19]. They employed end-to-end learning of the average pixel position difference (APPD) using convolutional neural networks (CNN). A similar method for extrinsic calibration monitoring of stereo cameras was introduced in [20]. They trained CNN to output the extrinsic parameters' actual miscalibration. They presented the effectiveness of their monitoring system on ORB-SLAM2 [21] failure detection. Even though the method showed promising results, it had a low recall (many undetected SLAM failures). We use this method as a baseline in our downstream experiment (see Sect. 4.4).

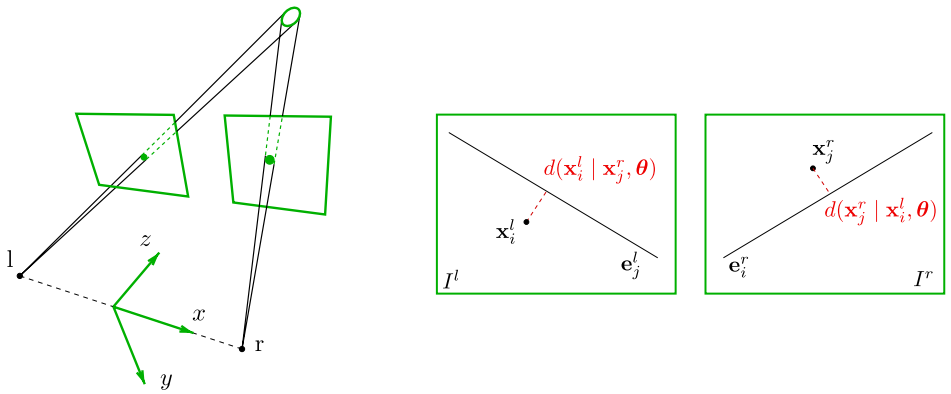
2.4 Contributions and structure of the paper

This paper is an extended version of [22], describing the StOCaMo method for online calibration monitoring for stereo cameras. It is a single-frame approach that minimizes the epipolar error with robust kernel correlation [23]. This paper proposes an additional confirmation technique for the StOCaMo output, significantly improving recall on small (borderline) decalibrations. In addition to the conventional statistical performance metrics, which we enrich for the specificity results, we also measure and report the confirmation efficiency. This metric reflects the percentage of input data labelled uninformative at a given precision level. To evaluate the performance of our method, we conducted experiments on a larger subset of the two real-world datasets. These datasets have been chosen for their reproducibility and comparability.

The rest of the paper is organized as follows: Section 3 gives a detailed description of the StOCaMo method, including a new resampling mechanism designed to improve the algorithm's recall. The experimental Sect. 4 starts with discussing the shape of the proposed loss function in comparison with a previous approach. In Sect. 4.2, we discuss the behaviour of the intermediate robustification function called the F-index. Sect. 4.3 reports calibration monitoring results on synthetic decalibrations of two real-world datasets. A comparison with the results from [20] is done in Sect. 4.4. Section 5 summarizes the results and outlines some topics for further research.

3 Methods

StOCaMo is based on the examination of epipolar error between similar keypoints. Instead of using some robust optimisation technique (RANSAC [24], LMedS [25]), we employ the kernel correlation principle, which is implicitly robust [23]. Because of the low time-to-detection requirement, we use the method from [3] to validate the calibration on a single frame. Figure 1 summarises the method:



(a) orthogonal world system xyz (b) epipolar lines

Fig. 2 Coordinate system (**a**) and epipolar geometry (**b**) for a pair of cameras

First, we extract keypoints [26] and their descriptors [27] from each stereo image. Second, we find tentative left-to-right and right-to-left matches guided by the descriptor similarity, which we then use in the kernel-based epipolar error [23]. Third, we evaluate the error function over a grid of small parameter perturbations around the reference. These are the primary measurements for the monitoring task. A probability distribution maps the primary measurements to the probabilities of correct and incorrect calibration. These are combined into what we term the calibration validity index (V-index in short), as in [3]. In this paper, we add one more step: A confirmation of the binary statistical decision at StOCaMo 's output, with the possibility of giving an 'unconfirmed' answer. The output is then one of three outcomes: 'calibrated system', 'unconfirmed', 'decalibrated system'. The 'unconfirmed' label is meant for low-quality data that contains too weak information on the system's calibration status. The proposed modification thus implements a data rejection mechanism.

The method is designed for a stereoscopic camera pair that relies on an overlap in the field of view to perform tasks based on inter-image correspondences. Therefore, our method assumes such a camera configuration. We further assume that sensors are global-shutter and that both cameras' intrinsic parameters (calibration matrices and distortion coefficients) and the reference extrinsic calibration [7] are known. The sensors should also be synchronized well so that there is a minimal effect of the relative latency in the data.

3.1 The StOCaMo method

Let us assume we are given two undistorted images, I^l and I^r, captured by two cameras with known camera matrices, $\mathbf{K}_l$ and $\mathbf{K}_r$. Our goal is to decide whether the given extrinsic calibration parameters $\boldsymbol{\theta}^{\text{ref}}$ are correct.

3.1.1 Keypoint detection and feature extraction

As already stated, our method minimises the epipolar distance of tentative matches. Therefore, we need to detect suitable keypoints and extract descriptors for matching.

Let $\mathcal{I}^l$ be the set of keypoint indices in the left image and $\mathcal{I}^r$ be the set of keypoints in the right image. Each left-image keypoint $i \in \mathcal{I}^l$ has a location $\mathbf{p}_i^l \in \mathbb{P}^2$ (hence, it has three 'homogeneous' coordinates in the projective space $\mathbb{P}^2$) and descriptor $\mathbf{f}_i^l \in \mathbb{R}^c$. Analogically, we have $\mathbf{p}_j^r, \mathbf{f}_j^r$ for $j \in \mathcal{I}^r$ in the right image.

In our experiments, we use the STAR keypoint detector [26] and the BRIEF descriptor [27], which provides descriptor vectors of dimension $c = 32$. These choices are not critical for the statistical performance of StOCaMo.

3.1.2 Epipolar geometry

As intrinsically calibrated cameras capture both images, we first transform the keypoints by the 3×3 calibration matrices $\mathbf{K}_l, \mathbf{K}_r$:

$$\mathbf{x}_i^l = \mathbf{K}_l^{-1}\mathbf{p}_i^l \quad \text{and} \quad \mathbf{x}_j^r = \mathbf{K}_r^{-1}\mathbf{p}_j^r, \quad \mathbf{x}_i^l, \mathbf{x}_j^r \in \mathbb{P}^2. \tag{1}$$

Let now $\mathbf{R} \in \mathbb{R}^{3\times3}$ be the matrix of the relative rotation between the cameras, and $\mathbf{t} \in \mathbb{R}^3$ be the relative translation vector. The rotation matrix is represented via Rodrigues' formula by the axis-angle vector $\boldsymbol{\omega} \in \mathbb{R}^3$, which describes a rotation by angle $||\boldsymbol{\omega}||$ around the rotation axis $\boldsymbol{\omega}$. The extrinsic parameters of dimension six are then composed of two three-element vectors $\boldsymbol{\theta} = (\mathbf{t}, \boldsymbol{\omega})$.

Due to the epipolar constraint [7], given a left-image keypoint $i \in \mathcal{I}^l$, the corresponding right-image keypoint $j \in \mathcal{I}^r$ needs to lie on the epipolar line $\mathbf{e}_i^r$ in the right image I^r, $\mathbf{x}_j^r \in \mathbf{e}_i^r$ (see Fig. 2b), such that

$$\mathbf{e}_i^r = \mathbf{E}(\boldsymbol{\theta})\mathbf{x}_i^l, \qquad \mathbf{e}_i^r = \left(e_{i,1}^r, e_{i,2}^r, e_{i,3}^r\right) \in \mathbb{P}^2,$$

and, analogically, given the right-image keypoint $j \in \mathcal{I}^r$, the left-image keypoint $i \in \mathcal{I}^l$ must lie on the line

$$\mathbf{e}_j^l = \mathbf{E}^\top(\boldsymbol{\theta})\mathbf{x}_j^r, \qquad \mathbf{e}_j^l = \left(e_{j,1}^l, e_{j,2}^l, e_{j,3}^l\right) \in \mathbb{P}^2,$$

where $(\cdot)^\top$ is matrix transposition. As the intrinsic parameters are known, the map $\mathbf{E}(\boldsymbol{\theta})$ is given by a 3×3 rank-deficient essential matrix

$$\mathbf{E}(\boldsymbol{\theta}) = [\mathbf{t}]_\times \mathbf{R}, \tag{2}$$

in which $[\cdot]_\times$ is the 3×3 skew-symmetric matrix composed of the elements $\mathbf{t}$ [7].

If the locations $\mathbf{x}_i^l$, $\mathbf{x}_j^r$ are imprecise, one expresses a pair of *epipolar errors*, each defined as the distance of a point from the corresponding epipolar line it belongs to:

$$\begin{aligned} d\left(\mathbf{x}_j^r \mid \mathbf{x}_i^l, \boldsymbol{\theta}\right) &= \frac{\left|\left(\mathbf{x}_j^r\right)^\top \mathbf{E}(\boldsymbol{\theta})\, \mathbf{x}_i^l\right|}{\sqrt{\left(e_{i,1}^r\right)^2 + \left(e_{i,2}^r\right)^2}}, \\ d\left(\mathbf{x}_i^l \mid \mathbf{x}_j^r, \boldsymbol{\theta}\right) &= \frac{\left|\left(\mathbf{x}_i^l\right)^\top \mathbf{E}^\top(\boldsymbol{\theta})\, \mathbf{x}_j^r\right|}{\sqrt{\left(e_{j,1}^l\right)^2 + \left(e_{j,2}^l\right)^2}}. \end{aligned} \tag{3}$$

where $\mathbf{a}^\top\mathbf{b}$ stands for the dot product of two vectors. Note that the numerators are the same, but the denominators differ.[1] Note also, that due to the calibration (1), this error is essentially expressed in angular units.

3.1.3 Average epipolar error

The standard approach to estimate the quality of a calibration (used also as a baseline in [20]) would minimise the average epipolar error over the nearest neighbours in the descriptor space:

$$\begin{aligned} \mathrm{AEE}\,(\boldsymbol{\theta}) = &\frac{1}{n} \sum_{i \in \mathcal{I}^l} \sum_{j \in \mathrm{NN}_i^r} d\left(\mathbf{x}_j^r \mid \mathbf{x}_i^l, \boldsymbol{\theta}\right) \\ &+ \frac{1}{n} \sum_{j \in \mathcal{I}^r} \sum_{i \in \mathrm{NN}_j^l} d\left(\mathbf{x}_i^l \mid \mathbf{x}_j^r, \boldsymbol{\theta}\right), \end{aligned} \tag{4}$$

where $n = |\mathcal{I}^l| + |\mathcal{I}^r|$ and NN_i^r is the nearest neighbour of $\mathbf{f}_i^l$ in the set $\{\mathbf{f}_j^r\}_{j \in \mathcal{I}^r}$ and NN_j^l is the nearest neighbour of $\mathbf{f}_j^r$ in $\{\mathbf{f}_i^l\}_{j \in \mathcal{I}^l}$. Lower AEE $(\boldsymbol{\theta})$ is better.

To reduce false matches, one can use the so-called Lowe's ratio [28]: The descriptor distance to two nearest neighbours is computed, and the match (i, j) is removed from (4) if their ratio is lower than a given threshold.

3.1.4 Robust loss function

Even though Lowe's ratio lowers the number of outliers to the epipolar geometry, there may still be undetected false matches (e.g., due to repetitive objects in the scene). Hence, we use a different model, which considers the uncertainty of a particular match. Specifically, we employ the kernel correlation principle, studied in the context of registration problems, e. g. in [23].

Let us define a kernel loss function for a point and a line (between subspaces of dimension zero and one). As the distance between these is symmetric, the kernel will also be a symmetric function of the distance. The uncertainty of the keypoint location is expressed through a predefined variance of the kernel. Hence, keypoints that have a large distance from their corresponding epipolar line (see (3)) will have a small effect on the resulting loss. Specifically, we use the Gaussian kernel and evaluate the loss function on the $k\,n$ tentative matches in the feature space:

$$\begin{aligned} KC(\boldsymbol{\theta}) = &-\frac{1}{n} \sum_{i \in \mathcal{I}^l} \sum_{j \in \mathrm{kNN}_i^r} \exp\left[-\frac{d^2\left(\mathbf{x}_j^r \mid \mathbf{x}_i^l, \boldsymbol{\theta}\right)}{2\sigma^2}\right] \\ &-\frac{1}{n} \sum_{j \in \mathcal{I}^r} \sum_{i \in \mathrm{kNN}_j^l} \exp\left[-\frac{d^2\left(\mathbf{x}_i^l \mid \mathbf{x}_j^r, \boldsymbol{\theta}\right)}{2\sigma^2}\right], \end{aligned} \tag{5}$$

where kNN_i^r are k nearest neighbours of $\mathbf{f}_i^l$ in the set $\{\mathbf{f}_j^r\}_{j \in \mathcal{I}^r}$ and kNN_j^l are the k nearest neighbours of $\mathbf{f}_j^r$ in $\{\mathbf{f}_i^l\}_{j \in \mathcal{I}^l}$ and n is as in (4). Again, lower $KC(\boldsymbol{\theta})$ is better due to the negative sign.

There are two model hyper-parameters: k and σ. Using $k > 1$ nearest neighbours helps the performance, as non-matching descriptors usually have larger distances and do not contribute to the loss too much. We set $k = 5$ in our experiments. The σ parameter depends on the *calibration tolerance* δ and we set $\sigma = \delta$. In this work, we assume

$$\delta = 0.005\,\mathrm{rad} \tag{6}$$

in the rotation around the x axis (Fig. 2a). In general, this value should be provided by the user of the monitoring

[1] Alternatively, we could use a single (symmetric) distance of a correspondence $\left(\mathbf{x}^l, \mathbf{x}^r\right)$ from the epipolar manifold $\mathbf{x}^r\mathbf{E}\mathbf{x}^l = 0$. Since there is no closed-form solution for this distance, it can be approximated with the Sampson error [7]. We prefer the epipolar error that needs no approximation.

system, and it should represent the tolerable deviation of the calibration parameters from the true ones.

3.1.5 Calibration monitoring

As the percentage of inlier correspondences changes from frame to frame (not to say from dataset to dataset), we cannot simply decide the calibration validity based on the loss function value (5) itself. The time-to-detection requirement also prefers to make the decision on a single frame, without any previous memory needed, as in [20]. Therefore, we use the approach introduced in [3], based on examining the loss function around the reference in a small perturbation grid defined in the parameter space.

Each of the six extrinsic parameters will have its own 1D grid step constant. For example, in the case of the translation in x it is defined as follows:

$$\boldsymbol{\Theta}_{\mathrm{tx}}^{\mathrm{pert}} = \left\{\theta_{\mathrm{tx}}^{\mathrm{ref}} - e_{\mathrm{tx}}, \theta_{\mathrm{tx}}^{\mathrm{ref}}, \theta_{\mathrm{tx}}^{\mathrm{ref}} + e_{\mathrm{tx}}\right\}, \tag{7}$$

where e_{tx} is the grid constant; it is defined analogically for the remaining parameters. The loss function (5) will then be evaluated on the Cartesian product of such 6D decalibrations:

$$\boldsymbol{\Theta}^{\mathrm{pert}} = \boldsymbol{\Theta}_{\mathrm{tx}}^{\mathrm{pert}} \times \boldsymbol{\Theta}_{\mathrm{ty}}^{\mathrm{pert}} \times \boldsymbol{\Theta}_{\mathrm{tz}}^{\mathrm{pert}} \times \boldsymbol{\Theta}_{\mathrm{rx}}^{\mathrm{pert}} \times \boldsymbol{\Theta}_{\mathrm{ry}}^{\mathrm{pert}} \times \boldsymbol{\Theta}_{\mathrm{rz}}^{\mathrm{pert}}, \tag{8}$$

which we call the *perturbation grid*. This can yield up to $3^6 = 729$ evaluations when all the parameters are perturbed. If the calibration is correct, the decalibrations should yield a higher loss value than in $\boldsymbol{\theta}^{\mathrm{ref}}$. Hence, inspired by [3], we define a quality measure called an *F-index*:

$$F\left(\boldsymbol{\theta}^{\mathrm{ref}}\right) = \frac{1}{|\boldsymbol{\Theta}^{\mathrm{pert}}|} \sum_{\boldsymbol{\theta} \in \boldsymbol{\Theta}^{\mathrm{pert}}} \mathbb{1}\left[KC\left(\boldsymbol{\theta}^{\mathrm{ref}}\right) \leq KC(\boldsymbol{\theta})\right], \tag{9}$$

where $\mathbb{1}[\cdot] \in \{0, 1\}$ is the indicator function. If the calibration parameters are correct, the F-index should be close to the unit value and smaller (about 0.5) otherwise.

The authors of [3] proposed to learn two different probability distributions over the $F\left(\boldsymbol{\theta}^{\mathrm{ref}}\right)$ random values for: (1) small noise within the calibration tolerance $\boldsymbol{\theta}^{\mathrm{ref}} \pm \delta$, denoted as p_c, and (2) large decalibration well behind the tolerance $\boldsymbol{\theta}^{\mathrm{ref}} \pm \Delta$, denoted as p_d, with $\Delta > \delta > 0$. Then, they defined a validity index (called *V-index* here) as a posterior probability with equal priors:

$$V\left(\boldsymbol{\theta}^{\mathrm{ref}}\right) = \frac{p_c\left(\boldsymbol{\theta}^{\mathrm{ref}}\right)}{p_c\left(\boldsymbol{\theta}^{\mathrm{ref}}\right) + p_d\left(\boldsymbol{\theta}^{\mathrm{ref}}\right)}. \tag{10}$$

In the original paper [3], the authors suggested using the normal distribution for p_c and p_d. This selection is not optimal, as the random values F are from the (discretised) interval [0, 1]. Instead, we use the empirical distributions (histograms) of F for p_c and p_d. As the V-index is the posterior probability, we set the threshold for the StOCaMo outcome to 0.5. The standard StOCaMo [22] output is then:

$$\begin{cases} \text{'decalibrated'} & \text{if } V\left(\boldsymbol{\theta}^{\mathrm{ref}}\right) < 0.5, \\ \text{'calibrated'} & \text{if } V\left(\boldsymbol{\theta}^{\mathrm{ref}}\right) \geq 0.5. \end{cases} \tag{11}$$

3.1.6 Increasing StOCaMo recall

The standard StOCaMo method [22] suffers from low recall (correctly detected decalibration) on borderline (small) decalibrations. Hence, in this work, we propose a confirmation method for the 'calibrated' decision of StOCaMo.

After some experimentation, we chose a resampling method that proved effective for variance estimation. If the system is calibrated, then the F-index (9) should be very high. But this should hold for any (reasonable) subsample of the keypoints. Thus, we examine the variance of F-indices over subsets of keypoints. This has a negligible cost, as all the errors are already precomputed in (5).

We divide the permuted keypoint indices into m subsets for each image as follows:

$$S_i^l = \left\{\mathrm{perm}(\mathcal{I}^l)_j\right\}_{j=i\cdot\frac{|\mathcal{I}^l|}{m}}^{(i+1)\cdot\frac{|\mathcal{I}^l|}{m}} \quad \text{and} \quad S_i^r = \left\{\mathrm{perm}(\mathcal{I}^r)_j\right\}_{j=i\cdot\frac{|\mathcal{I}^r|}{m}}^{(i+1)\cdot\frac{|\mathcal{I}^r|}{m}}, \tag{12}$$

where perm$(\cdot)$ is a random permutation of the index array. The loss function is then derived for the subsets pair $\left(S_k^l, S_k^r\right)$ as follows:

$$\begin{aligned} KC\left(\boldsymbol{\theta} | S_k^l, S_k^r\right) = &-\frac{1}{n} \sum_{i \in S_k^l} \sum_{j \in \mathrm{kNN}_i^r} \exp\left[-\frac{d^2\left(\mathbf{x}_j^r \mid \mathbf{x}_i^l, \boldsymbol{\theta}\right)}{2\sigma^2}\right] \\ &-\frac{1}{n} \sum_{j \in S_k^r} \sum_{i \in \mathrm{kNN}_j^l} \exp\left[-\frac{d^2\left(\mathbf{x}_i^l \mid \mathbf{x}_j^r, \boldsymbol{\theta}\right)}{2\sigma^2}\right]. \end{aligned} \tag{13}$$

Note that all the errors and their exponential function values are already pre-computed in (5). The F-index for the same subsets pair is then defined analogically to (9) as:

$$\begin{aligned} F\left(\boldsymbol{\theta}^{\mathrm{ref}} | S_k^l, S_k^r\right) = &\frac{1}{|\boldsymbol{\Theta}^{\mathrm{pert}}|} \\ &\sum_{\boldsymbol{\theta} \in \boldsymbol{\Theta}^{\mathrm{pert}}} \mathbb{1}\left[KC\left(\boldsymbol{\theta}^{\mathrm{ref}} | S_k^l, S_k^r\right) \leq KC\left(\boldsymbol{\theta} | S_k^l, S_k^r\right)\right]. \end{aligned} \tag{14}$$

We estimate the variance of the F-index on the subsets:

$$\sigma_F^2\left(\boldsymbol{\theta}^{\mathrm{ref}}\right) = \frac{1}{m} \sum_{i=1}^{m} \left(F\left(\boldsymbol{\theta}^{\mathrm{ref}} | S_i^l, S_i^r\right) - \frac{1}{m} \sum_{k=1}^{m} F\left(\boldsymbol{\theta}^{\mathrm{ref}} | S_k^l, S_k^r\right) \right)^2. \tag{15}$$

StOCaMo with confirmation then returns one of three outcomes:

$$\begin{cases} \text{'decalibrated'} & \text{if } V\left(\boldsymbol{\theta}^{\text{ref}}\right) < 0.5, \\ \text{'calibrated'} & \text{if } V\left(\boldsymbol{\theta}^{\text{ref}}\right) \geq 0.5 \ \wedge \ \sigma_F^2(\boldsymbol{\theta}^{\text{ref}}) \leq \tau_F^2, \\ \text{'unconfirmed'} & \text{otherwise,} \end{cases} \quad (16)$$

where $V(\boldsymbol{\theta}^{\text{ref}})$ is computed as in (9) and (10). Besides the standard statistical accuracy of (16), we also examine the data loss this rule causes (i.e., the fraction of data that has the 'unconfirmed' outcome). We use $m = 10$ subsets in this work and learn the threshold τ_F on the synthetic dataset with other parameters. This value of m is a tradeoff dictated by the limited data available to us: It should not be too small to have good variance estimates in (15), and it should not be too large because then there would be too few samples in the subset from which the F-index (14) is computed.

3.2 Discussion

After we have described the proposed method, it calls for a comparison with the method proposed by Zhong et al. [20] and for a discussion of the algorithmic differences. First, Zhong et al. learned a function that maps the input image pair to the calibration error. The main contribution of their paper is the *Weighted Overall Disturbance Effect* (WODE), which acts as a 'teacher' used to learn the parameters of the neural network. Specifically, the weight $w_i(d_i)$ definition in their equation (1) is interesting. Instead of expressing the calibration error in rotation and translation parameters, the $w_i(d_i)$ expresses it as the rotation angle needed to re-rectify the images to a common image plane. In contrast, we use epipolar error expressed in the image plane, where the primary measurements occur. Second, unlike in the method of Zhong et al., we do not estimate error on any parameters but directly verify the reference parameters on data and then perform a statistical test on the calibration validity. Apriori, these two methods do not have clear advantages over each other, and an experimental evaluation (will be provided in Sect. 4.4) is needed.

4 Experiments

In this work, experiments are performed on one synthetic (CARLA [29]) and two real-world datasets (KITTI [30], EuRoC [31]). All the parameters are learned on the synthetic dataset and then used on the real-world ones without any modification. This illustrates a good generalisation of StOCaMo.

CARLA is a simulator based on the Unreal Engine, hence it provides highly realistic scenes in several pre-created worlds [29]. Although we use it in this work to simulate the stereo pair of cameras, it provides a plethora of other sensors (LiDARs, depth cameras, Radars, etc.). We simulate a stereo pair of two cameras with 70° horizontal and 24° vertical fields of view with a resolution of 1241×376 pixels. Both cameras are front-facing with parallel optical axes (or, equivalently, image planes) and have a baseline equal to 0.4 m. We use the default map Town10HD_Opt with 100 autopilot vehicles to simulate the traffic. We recorded 155 sequences (from different spawn points) with 200 frames each.

KITTI is one of the most popular public datasets in the automotive domain [30]. It contains data for several different tasks, such as odometry estimation, optical flow estimation, object recognition, or object tracking. We will use the rectified and synchronised data from the training part of the odometry evaluation dataset (Sequences 00-10). There are two grayscale, global-shutter cameras with 70° horizontal and 29.5° vertical field of view and a resolution of 1226×370 pixels.

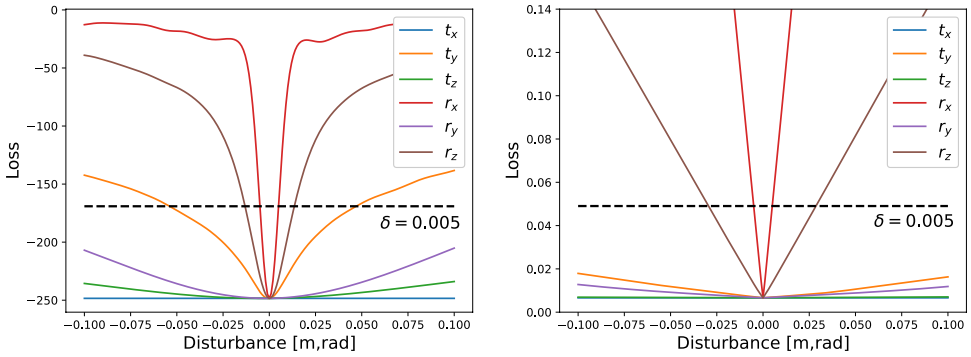
EuRoC MAV dataset was captured on-board a micro aerial vehicle (MAV) [31]. Hence, it provides unique fast movements that are not present in the automotive data. In the experiments, we use ten sequences (all five from the Machine Hall and five from the Vicon Room,[2]) with the stereo, global-shutter cameras. The horizontal field of view of unrectified images is 79°, the vertical field of view is 55°, the resolution is 752×480, and the baseline is 0.11 m.

4.1 Loss function shape

We first examined the shape of the proposed KC loss (5) in a similar way as in [20]. We took one sequence from the CARLA dataset (200 frames) and evaluated the average loss. The calibration perturbations were selected as in [20], from $[-0.1, 0.1]$ rad in rotation and $[-0.1, 0.1]$ m in translation. The loss function should be minimal for the reference parameters and quickly increase for large perturbations.

The result of the KC loss (5) is shown in Fig. 3a. All the parameters except translations in the x direction have clear minima in the reference, but some have a higher impact on the KC loss than others. Specifically, the x-axis rotation r_x has the most prominent influence because it corrupts all the epipolar lines identically (either moves them up or down). The rotation r_z around the z (optical) axis and the y-axis translation t_y have a similar effect on the epipolar lines. One can see (black dashed line) that the decalibration of $\delta = 0.005$ rad (6) in r_x has a similar effect on the KC loss as 0.012 rad decalibration in r_z and 0.05 m decalibration in t_y. The other two observable degrees of freedom (y-axis

[2] We did not use sequence *Vicon Room 2 03* which contained too much motion blur.



(a) Proposed kernel correlation loss (b) Average epipolar error

Fig. 3 Evaluation of the proposed KC loss (5) (**a**) and the standard AEE loss (4) (**b**) on one sequence from the CARLA dataset. Translations in the x direction (blue) are unobservable in epipolar error. The translation plot in the y direction (orange) (almost) coincides with the rotation around the y axis (purple) in (**b**). This comparison shows increased relative sensitivity in the translation in y by our proposed loss (**a**) (color figure online)

rotation r_y and z-axis translation t_z) are less apparent than the others. To increase the sensitivity to those, one would probably need to control the distribution of keypoints in the image (preferring the periphery, for instance).

Note that the change in the baseline length (translation in x) and the focal length value of the intrinsic calibration have no impact: The translation length does not change the $\mathbf{E}(\boldsymbol{\theta})$ in (3). The scaling effect of the focal length or image resolution cancels out in the normalization of (3) when the optical axes are parallel.

Using the BRIEF descriptor, we also performed the same experiment with the average epipolar error (4) with Lowe's ratio on matched STAR features. The result is shown in Fig. 3b, and it replicates the corresponding results from [20], where they used SIFT instead of our combination STAR+BRIEF. The KC loss (5) has a greater sensitivity to translation in y than the AEE loss (4). Moreover, the KC loss also exhibits robustness to larger errors in the rotation in the x axis (red).

In order to estimate the F-index (9), we need to set the grid step e in $\boldsymbol{\Theta}^{\text{pert}}$ (7). In this work, we will detect the decalibration on the three most observable degrees of freedom (DoFs) from Fig. 3a, i.e., rotations around the x and z axes and the translation in y. Decalibration in these parameters will have the largest impact on the epipolar geometry (and all downstream computer vision tasks). The choice of the grid step for rotations around x is based on the calibration tolerance δ (6), so that it is outside of the basin of attraction: $e_{\text{rx}} = 3 \cdot \delta = 0.015\,\text{rad}$. The perturbations of the other two DoFs are set so that their relative change of the loss with respect to the rotation around the x axis is the same. Based on Fig. 3a, they are: $e_{\text{ry}} = 3 \cdot 0.012 = 0.036\,\text{rad}$ and $e_{\text{ty}} = 3 \cdot 0.015 = 0.045\,\text{m}$. With this rule, the corresponding decalibrations in the remaining DoFs would be unrealistically large to be considered.

4.2 Decalibration detection

By the derivations in Sect. 3.1, a decalibration should manifest in the F-index (9). Nevertheless, the magnitude of the change is also important. A small perturbation should not deviate the $F(\theta^{\text{ref}})$ values much from the unit value, while larger decalibrations should make it smaller, with high variance. This sensitivity depends on the selected calibration tolerance δ and thus on the σ and $\boldsymbol{\Theta}^{\text{pert}}$ (8).

In this experiment, we evaluate the F-index on all 155 CARLA sequences (200 frames each) with random (uniform) extrinsic decalibration of several magnitudes:

$$\{0,\ 0.0025,\ 0.005,\ 0.01,\ 0.02,\ 0.05,\ 0.075\}, \tag{17}$$

meters or radians in all six degrees of freedom. Due to the selected calibration tolerance δ, the F-index should be close to the unit value up to this decalibration magnitude. After that, it should quickly decrease to about 0.5. This behaviour can be seen in Fig. 4, where the bar corresponds to the

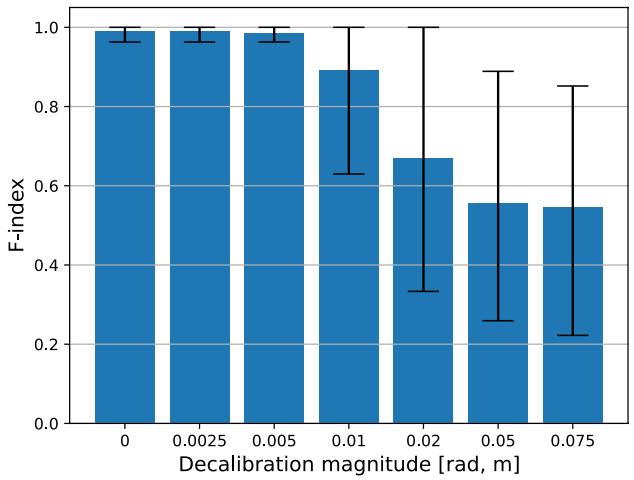


Fig. 4 F-index (9) as a function of decalibration on the CARLA dataset

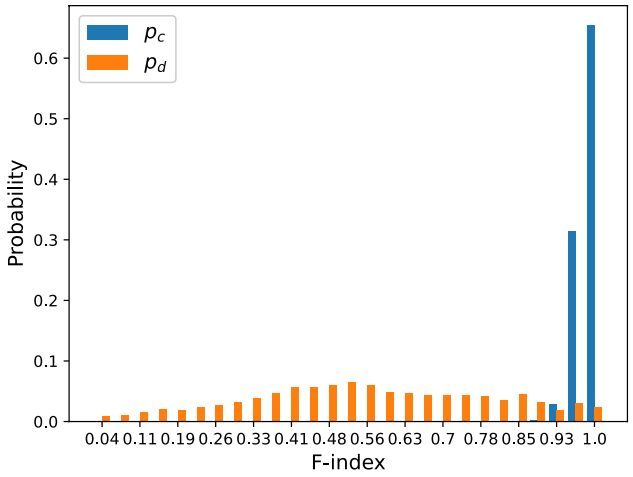


Fig. 5 Histograms of p_c and p_d from (10)

as the 0.01 and 0.02 are too close to the reference, and 0.075 yields similar values. One can see both histograms in Fig. 5.

We also learn the standard deviation of the F-index (9) on this data with the decalibration magnitude 0.005. The deviation equals $\sigma_F^{0.005} = 0.021$, which will be used for the deviation's threshold τ_F.

4.3 Calibration monitoring on synthetic decalibration

In the following experiment, we evaluate the performance of StOCaMo in detecting extrinsic decalibrations on the two real-world datasets. Based on the results from Sect. 4.2, small decalibrations should be reported as valid and larger ones as decalibrated. Hence, we investigate two scenarios: (1) a small decalibration within the calibration tolerance δ (6) from $[-\delta, \delta]$ m or rad, and (2) a borderline decalibration

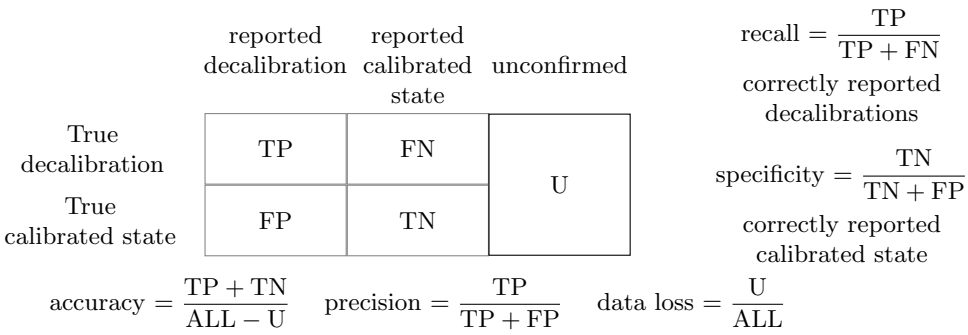


Fig. 6 Studied metrics for synthetic decalibration detection. The standard StOCaMo method had a problem detecting borderline decalibrations, so we focused mainly on recall and accuracy improvement of the proposed modification. ALL = TP + FN + FP + TN + U

average $F(\theta^{\text{ref}})$ over all frames and the errorbars show the 15 % and 85 % quantiles for that decalibration magnitude. Small decalibrations up to 0.005 (m or rad) have high $F(\theta^{\text{ref}})$ values around 0.98 of small variance. With increasing decalibration magnitude, this value drops to 0.55 for the 0.05 and 0.075 magnitudes with a large variance.

To estimate the posterior probability of the reference calibration parameters (see the V-index in (10)), we need to learn the distribution p_c for a small noise within the calibration tolerance δ around the reference and p_d for large decalibrations Δ. We use the actual histograms (as opposed to the normal distribution in [3]) of the F-index values on two magnitudes. The p_c is learned from the magnitude equal to the calibration tolerance, i.e. 0.005 (m or rad). For the p_d, we pick the magnitude

$$\Delta = 0.05, \tag{18}$$

from $[-2\delta, -\delta] \cup [\delta, 2\delta]$ m or rad. This borderline decalibration corresponds to about 5–9 px in the reprojection error for datasets used in this paper. The first decalibration is within the tolerance; hence, the monitoring method should label it as 'calibrated'. Therefore, it examines the metrics of true negative (no decalibration) and false positive (false alarm). The second decalibration magnitude was chosen already large enough to be detected. We use it to estimate the true positive (detected decalibration) and false negative (undetected decalibration) rates of StOCaMo. This second decalibration magnitude is two times smaller than in [22] to investigate borderline decalibrations, with the aim to push the boundary of detectable decalibrations. Unlike in [22], we do not consider larger decalibrations $[2\delta, 4\delta]$ in this paper. Figure 6 summarises the studied metrics for this synthetic decalibration experiment. We report recall, which represents the ratio of correctly detected decalibrations. The original StOCaMo method [22] suffered from low recall (high number of

Table 1 Results of StOCaMo on the borderline synthetic decalibration detection without and with confirmations of decreasing threshold τ_F (i.e., increasing strength)

	Recall	Specificity	Accuracy	Data loss	Error improvement
(a) KITTI					
w/o confirm	63.8 % (±3.4)	99.33 % (±0.28)	81.6 % (±1.6)	0 %	–
$\tau_F = 3\sigma_F^{0.005}$	69.7 % (±3.2)	99.32 % (±0.29)	85.0 % (±1.4)	5.1 % (±1.5)	23.1 % (±7.6)
$\tau_F = 2\sigma_F^{0.005}$	77.1 % (±3.4)	99.28 % (±0.31)	88.8 % (±1.4)	12.1 % (±2.8)	66.3 % (±18.9)
$\tau_F = \sigma_F^{0.005}$	91.0 % (±2.4)	98.75 % (±0.73)	94.7 % (±0.8)	35.4 % (±6.4)	255.0 % (±75.7)
(b) EuRoC					
w/o confirm	56.6 % (±3.6)	95.99 % (±5.23)	76.3 % (±4.3)	0 %	–
$\tau_F = 3\sigma_F^{0.005}$	63.1 % (±5.2)	95.57 % (±5.91)	79.6 % (±5.3)	8.6 % (±3.7)	18.4 % (±10.7)
$\tau_F = 2\sigma_F^{0.005}$	69.1 % (±5.6)	95.09 % (±6.61)	82.4 % (±5.8)	16.1 % (±4.4)	40.8 % (±20.3)
$\tau_F = \sigma_F^{0.005}$	82.0 % (±5.5)	93.30 % (±8.92)	87.5 % (±6.7)	33.0 % (±5.6)	123.3 % (±68.2)

undetected decalibrations) on borderline decalibrations; hence it is the main focus of this work. To show that the improvement of the low recall is not achieved at the cost of labelling many truly calibrated states as 'unconfirmed' (i.e., decreasing true negatives), we also report the specificity metric (the ratio of correctly reported calibrated states). Besides these statistical metrics, we investigate the data loss due to 'unconfirmed' outcomes and the error improvement (1 − accuracy) with respect to standard StOCaMo.

On each frame of the KITTI and EuRoC sequences, we perform ten decalibrations of each of the kinds mentioned above. We then evaluated StOCaMo on these stereo images with perturbed parameters, as shown in Table 1. Results are averaged over all sequences for both datasets, and they also show the standard deviation for each metric over these sequences.

The recall of standard StOCaMo method is quite low on the smaller borderline decalibration magnitude used here (the first rows in Table 1a, b). This corresponds to our original observation that StOCaMo has problems with borderline decalibrations. With the decreasing threshold of the confirmation rule in (16), both the recall and the accuracy increase considerably for both datasets (the other rows in Table 1a, b). An increase of 'unconfirmed' calibrated states (decrease in specificity) is visible with the decreasing threshold, but it is quite negligible compared to the boost in the recall. The statistical precision is 99.0 % (±0.4) on KITTI and 93.6 % (±7.9) on EuRoC datasets (not shown in the table[3]). The data loss (the penultimate column) is similar for both datasets (about one-third of the data for the strongest threshold), but the improvement (i.e., the accuracy and recall) is higher on the KITTI dataset. We attribute this behaviour to the different characteristics of each stereo system, as discussed in [22]. As all the parameters were learned on simulated data from CARLA [29], these are probably more suitable for the KITTI data with similar vertical and horizontal fields of view. Still, these results show a good generalization of the original StOCaMo method and of the proposed confirmation technique, which has shown a substantial improvement in recall.

4.4 Long-time SLAM stability with calibration monitoring

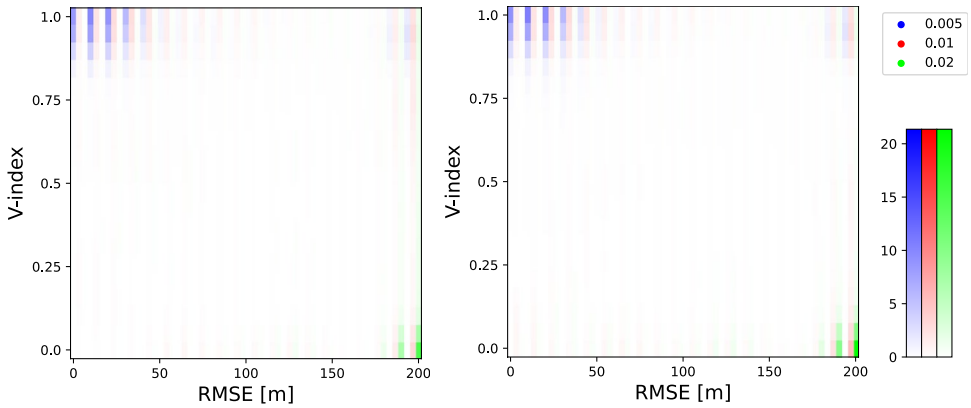
Incorrect calibration parameters may have a negative effect on the downstream computer vision tasks. In [20], they investigated the reliability of their calibration monitoring system in detecting the ORB-SLAM2 [21] failure. Given a sequence from a dataset, the SLAM was considered failing if the SLAM's root mean squared error (RMSE) from the ground truth was higher than some threshold (or if the SLAM did not finish at all). To avoid the selection of an arbitrary threshold needed to calculate accuracy, only the correlation between the RMSE and our V-index (10) is estimated.

We performed a similar experiment to the one in [20], but used more sequences and synthetic decalibration of six magnitudes from (17) to fully investigate the monitoring method's behaviour. All the studied sequences from both datasets were corrupted by 20 perturbations of each magnitude, and these decalibrations were then given to the ORB-SLAM2 to perform trajectory estimation. This gave us RMSE (infinity if diverged) for each perturbation and magnitude (120 values) for each sequence. For StOCaMo, we randomly sampled ten frames (for more precise statistics) from each sequence to predict a validity index for each perturbation and each magnitude (1200 values). For StOCaMo with confirmation, we sampled frames until ten frames were confirmed as 'calibrated' or 'decalibrated'. This gave us 1200 pairs of RMSE and V-indices per sequence, from which we estimated Spearman's and Kendall's correlation

[3] Note that the precision metric does not change with the proposed confirmation technique. Only 'calibrated' outcomes from the original StOCaMo can be labelled as 'unconfirmed' (compare (11) and (16)); hence it does not change the true and false positive rates in the precision.

Table 2 Evaluation of the correlation metrics of StOCaMo without and with confirmation on a downstream experiment on KITTI and EuRoC

	KITTI		EuRoC	
	Spearman	Kendall	Spearman	Kendall
w/o confirm	−0.75 (±0.05)	−0.61 (±0.04)	−0.75 (±0.03)	−0.6 (±0.03)
$\tau_F = \sigma_F^{0.005}$	−0.78 (±0.04)	−0.63 (±0.04)	−0.78 (±0.04)	−0.63 (±0.04)
[20]	0.44	N/A	0.59	N/A



(a) w/o confirmation (b) $\tau_F = \sigma_F^{0.005}$

Fig. 7 A heatmap visualization of ORB-SLAM2 RMSE w.r.t. StOCaMo V-index on the KITTI dataset. The results are shown for three decalibration magnitudes as alternating vertical stripes in three colour channels: Within calibration tolerance δ (in blue), and small (red) and large (green) borderline decalibrations, respectively. Colour saturation corresponds to the density of outcomes. The numbers on the scale bar on the right give the percentage of outcomes (color figure online)

ranks that are shown in Table 2 with their standard deviations over all sequences.

In our work, higher V-index (10) should result in lower RMSE; hence our correlation should be negative (lower is better), as opposed to [20], where they predict the decalibration magnitude. In other words, the sign does not matter in Table 2.

In [20], they achieved Spearman rank correlation of 0.59 on one sequence from the EuRoC dataset and 0.44 on one sequence from the KITTI dataset. StOCaMo shows better correlations with ORB-SLAM2 RMSE, even when taking more sequences into account. These results demonstrate a better ability to predict SLAM error based on the V-index (10) on both datasets. Using the confirmation technique proposed in this work, we enhanced the predictions of the ORB-SLAM2 error given by StOCaMo by about 0.03 in both correlation coefficients and datasets. This improvement comes with no additional computational cost.

To better illustrate the effect of the proposed confirmation (16) and to give an insight into the correlation results, we show the RMSE (clipped to 200 m) with respect to the StOCaMo V-index in Fig. 7 over all sequences from the KITTI dataset and three decalibration magnitudes without and with confirmation. The plots are multi-colour heatmaps, showing the Gaussian-smoothed density of the individual (RMSE, V-index) measurement points. For a small decalibration within the calibration tolerance δ (blue), a method should report a high V-index with low RMSE (top-left corner), which holds for both StOCaMo versions without (Fig. 7a) and with (Fig. 7b) confirmation. In the case of larger borderline decalibration (green), the V-index should be low with high RMSE (bottom-right corner). The proposed confirmation has an edge in this scenario, as the original StOCaMo method (see Fig. 7a) has a much larger extent of green colour on the right side of the graph. This corresponds to a higher number of undetected decalibrations and, thus, lower recall in the original method (11). Results on a small borderline decalibration (red) are ambiguous. On the one hand, the confirmation (see Fig. 7b) has a higher recall, i.e., greater density in the bottom-right corner than in the top-right corner (w.r.t. Fig. 7a). On the other hand, it also exhibits more false alarms (bottom-left corner), which corresponds to a somewhat lower specificity (discussed in previous Sect. 4.3). Overall, StOCaMo with confirmation provides much better detection of these major (more than 200 m RMSE) odometry estimation failures. These results correspond to the improvement in rank correlations from Table 2.

4.5 Algorithmic efficiency

Our current implementation of StOCaMo runs on a laptop CPU,[4] and our Python implementation needs only 70 ms per frame. The most time-consuming parts are the kNN search in the KDTree (30 ms) and the keypoint detection and descriptor construction (17 ms), which can be recycled from the image preprocessing for the SfM downstream task. A direct comparison with neuronal nets [20] is difficult to make because of the different hardware architecture.

[4] AMD Ryzen 7 5800 H.

5 Conclusion

We have described an efficient, robust and single-frame method for assessing the calibration quality of a stereoscopic vision system, with particular emphasis on achieving a high recall rate, which signifies the system's ability to detect decalibrations accurately. Our study demonstrates that the integration of F-index variance within the StOCaMo decision process resulted in a 25 % improvement in recall performance across two real-world datasets. Simultaneously, this modification led to a 12 % increase in the overall statistical accuracy, as detailed in Table 1.

The SLAM error prediction, expressed as a rank correlation value, achieved 0.78 with confirmation (over several sequences from two datasets). This is a better result than in [20], by a margin of 0.26. Nevertheless, the WODE idea from [20] is interesting, and we plan to explore its possible combination with the KC definition. This could solve the problem of non-uniform observability of the calibration parameters apparent in Fig. 3a and could result in the design of an optimal calibration monitoring solution for stereoscopic perception systems.

Instead of resampling the variance, an alternative method would consider the resampled F-index distributions, for both classes ('calibrated' and 'decalibrated'). Our experiments have shown that the variance estimate of the 'uncalibrated' class is unstable, the exact cause is unclear and more research on this topic is needed.

Acknowledgements This research has been supported by the Technology Agency of the Czech Republic under the National Competence Centres II Programme (NCC-II), Project #TN02000054 Božek Vehicle Engineering NCC-II (BOVENAC) and by the Czech Technical University in Prague grant SGS22/111/OHK3/2T/13.

Author Contributions JM: Methodology, Software, Data curation, Writing - Original Draft, Writing - editing, Visualization; RŠ: Conceptualization, Methodology, Writing - review & editing, Supervision

Funding Open access publishing supported by the National Technical Library in Prague. Technology Agency of the Czech Republic (project number TN02000054); Czech Technical University in Prague (Grant Number SGS22/111/OHK3/2T/13).

Availability of data and materials Datasets are publicly available from [30] and [31]. The synthetic dataset was generated in a publicly available software CARLA [29].

Declarations

Ethical approval Not applicable.

References

1. Zhang Z (2000) A flexible new technique for camera calibration. IEEE Trans Pattern Anal Mach Intell 22(11):1330–1334
2. Cvišić I, Marković I, Petrović I (2023) Soft2: stereo visual odometry for road vehicles based on a point-to-epipolar-line metric. IEEE Trans Robot 39(1):273–288
3. Levinson J, Thrun S (2013) Automatic online calibration of cameras and lasers. In: Proceedings robotics: science and systems conference, art. no. 29. https://doi.org/10.15607/RSS.2013.IX.029
4. Hirschmuller H, Gehrig S (2009) Stereo matching in the presence of sub-pixel calibration errors. In: IEEE conference on computer vision and pattern recognition, pp 437–444
5. Golkowski AJ, Handte M, Roch P, Marrón PJ (2020) Quantifying the impact of the physical setup of stereo camera systems on distance estimations. In: Fourth IEEE international conference on robotic computing (IRC), pp 210–217
6. Dubbelman G, Groen FCA (2009) Bias reduction for stereo based motion estimation with applications to large scale visual odometry. In: IEEE conference on computer vision and pattern recognition, pp 2222–2229
7. Hartley R, Zisserman A (2003) Multiple view geometry in computer vision. Cambridge University Press, Cambridge
8. Zhang Z, Luong Q-T, Faugeras O (1996) Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction. IEEE Trans Robot Autom 12(1):103–113
9. Dang T, Hoffmann C (2004) Stereo calibration in vehicles. In: IEEE intelligent vehicles symposium, pp 268–273
10. Dang T, Hoffmann C, Stiller C (2006) Self-calibration for active automotive stereo vision. In: IEEE intelligent vehicles symposium, pp 364–369
11. Brookshire J, Teller S (2013) Extrinsic calibration from per-sensor egomotion. In: Proceedings robotics: science and systems conference, pp 504–512
12. Daniilidis K (1999) Hand-eye calibration using dual quaternions. Int J Robot Res 18(3):286–298
13. Poursaeed O, Yang G, Prakash A et al (2018) Deep fundamental matrix estimation without correspondences. In: Proceedings of the European conference on computer vision workshops, pp 485–497
14. Gil Y, Elmalem S, Haim H et al (2021) Online training of stereo self-calibration using monocular depth estimation. IEEE Trans Comput Imaging 7:812–823
15. Dang T, Hoffmann C, Stiller C (2009) Continuous stereo self-calibration by camera parameter tracking. IEEE Trans Image Process 18(7):1536–1550
16. Mueller GR, Wuensche H-J (2017) Continuous stereo camera calibration in urban scenarios. In: International conference on intelligent transportation systems, pp 1–6
17. Mueller RG, Burger P, Wuensche H-J (2018) Continuous stereo self-calibration on planar roads. In: IEEE intelligent vehicles symposium, pp 1755–1760
18. Hansen P, Alismail H, Rander P et al (2012) Online continuous stereo extrinsic parameter estimation. In: IEEE conference on computer vision and pattern recognition, pp 1059–1066

19. Cramariuc A, Petrov A, Suri R et al (2020) Learning camera miscalibration detection. In: IEEE International Conference on robotics and automation, pp 4997–5003
20. Zhong J, Ye Z, Cramariuc A et al (2021) CalQNet-detection of calibration quality for life-long stereo camera setups. In: IEEE intelligent vehicles symposium, pp 1312–1318
21. Mur-Artal R, Tardós JD (2017) ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras. IEEE Trans Robot 33(5):1255–1262
22. Moravec J, Šára R (2023) StOCaMo: online calibration monitoring for stereo cameras. In: IbPRIA 2023: pattern recognition and image analysis, vol LNCS 14062, pp 336–350
23. Tsin Y, Kanade T (2004) A correlation-based approach to robust point set registration. In: European conference on computer vision, pp 558–569
24. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395
25. Rousseeuw PJ, Leroy AM (2005) Robust regression and outlier detection. Wiley, Toronto
26. Agrawal M, Konolige K, Blas MR (2008) Censure: center surround extremas for realtime feature detection and matching. In: European conference on computer vision, pp 102–115
27. Calonder M, Lepetit V, Strecha C et al (2010) Brief: binary robust independent elementary features. In: European conference on computer vision, pp 778–792
28. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60:91–110
29. Dosovitskiy A, Ros G, Codevilla F et al (2017) CARLA: An open urban driving simulator. In: Proceedings of the annual conference on robot learning, pp 1–16
30. Geiger A, Lenz P, Urtasun R (2012) Are we ready for autonomous driving? The KITTI vision benchmark suite. In: IEEE conference on computer vision and pattern recognition, pp 3354–3361
31. Burri M, Nikolic J, Gohl P et al (2016) The EuRoC micro aerial vehicle datasets. Int J Robot Res 35(10):1157–1163