2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Adaptive Outlier Thresholding for Bundle Adjustment in Visual SLAM

Alejandro Fontan[1], Javier Civera[2] and Michael Milford[1]

***Abstract*— State-of-the-art V-SLAM pipelines utilize robust cost functions and outlier rejection techniques to remove incorrect correspondences. However, these methods are typically fine-tuned to overfit certain benchmarks and struggle to adapt effectively to changes in the application domain or environmental conditions. This renders them impractical for many robotic applications in which robustness in a wide variety of conditions is essential. In this paper we introduce a novel distribution-based approach for online outlier rejection that reduces the necessity for scene-specific fine-tuning while simultaneously improving the overall SLAM performance. Through experiments across 3 different public datasets, we show that our approach consistently outperforms state-of-the-art methods in various real-world settings. Our code is available at https://github.com/alejandrofontan/ORB_SLAM2_Distribution**

I. INTRODUCTION

Visual Simultaneous Localization and Mapping (V-SLAM), that utilizes a camera as the primary sensor, has achieved outstanding results in providing accurate real-time positioning while simultaneously creating a map in complex and unstructured environments for almost two decades now [1], [2]. This remarkable progress has been possible through the integration of many techniques, mainly frame-to-frame Visual Odometry (VO) [3], map-to-frame camera tracking [2], local sliding-window Bundle Adjustment (BA) [4], [5], [6], [7] and Loop Closure (LC) [8], [9], [10], [11] for global consistency. As a result, SLAM has found extensive adoption in real-world applications across various industries, including robotics, autonomous driving, and augmented and virtual reality.

The estimation models at the core of V-SLAM methods [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22] are all formulated to account for spurious sensor measurements, likely to appear in visual data. Due to this, V-SLAM has consistently demonstrated a high degree of accuracy and robustness, even in scenarios characterized by data association errors or the emergence of spurious features due to dynamic environment changes [23], [24], [25].

To minimize the impact of outliers and avoid full tracking failure, V-SLAM systems tend to tune their outlier rejection methods for high precision, which results in a low recall and the loss of valid measurements, ultimately degrading the estimation accuracy [26], [27]. Furthermore, such conservative tuning is often scene-specific and fails to generalize effectively to variations in application deployment or environment conditions, resulting in sub-optimal accuracy or, in some cases, even catastrophic failure.

The accuracy and robustness of V-SLAM pipelines has a significant dependence on the specific parameter tuning used for outlier rejection, that often requires extensive training data prior to deployment in specific applications [28]. Reducing

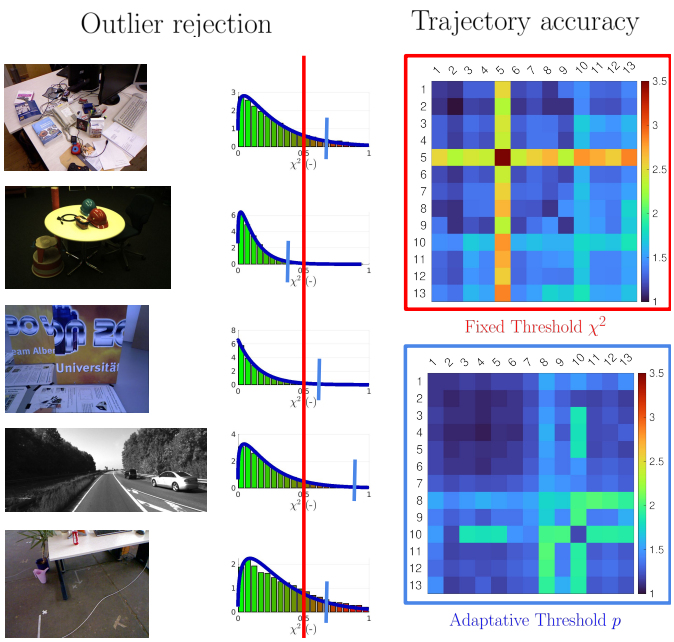


Fig. 1: **Probabilistic outlier rejection for BA.** The central column shows reprojection errors histograms after Bundle Adjustment in the context of feature-based V-SLAM. Note how the varying camera, motion and scene conditions of different datasets (left column) result in very distinct error distribution shapes. Selecting a unique threshold for outlier rejection for them all (red line) can lead to sub-optimal performance. Our approach dynamically selects an appropriate threshold (blue lines) for each error histogram. The inter-sequence accuracy matrices in the right colum (colder colors are better, see Section IV-C for details) show that our approach significantly reduces the need for parameter fine-tuning, a crucial aspect for a wide range of robotics and computer vision applications.

*This research was partially supported by funding from ARC Laureate Fellowship FL210100156 to MM and the QUT Centre for Robotics, the Spanish Government under Grant PID2021-127685NB-I00 and TED2021-131150B-I00 and the Aragon Government under Grant DGA_FSET45_20R.

[1]Alejandro Fontan is with the QUT Centre for Robotics, Queensland University of Technology, 4000 Brisbane, Australia alejandro.fontan@qut.edu.au

[2]Javier Civera is with the School of Engineering, University of Zaragoza, 50009 Zaragoza, Spain jcivera@unizar.es

[2]Michael Milford is with the QUT Centre for Robotics, Queensland University of Technology, 4000 Brisbane, Australia michael.milford@qut.edu.au

the dependency on training data is crucial for achieving robust perception and navigation in a wide range of robotic applications, especially those with limited or costly data collection or operating in environments subject to frequent changes. The aim of our research is to develop an outlier detection technique for feature-based Bundle Adjustment in V-SLAM that improves the overall performance of V-SLAM in *any* scene. This requires that our technique adapts to changes in the application context without hand-tuned parameters.

In this paper we introduce a novel approach for outlier filtering in V-SLAM, which is mainly based on an adaptive threshold estimated online from the reprojection errors. Our key contribution (Section III) lies in the online calculation of a **probabilistic threshold** derived from a ***Gamma* distribution**, where the distribution's parameters are continuously and robustly estimated from ***partial-fitting*** (Section III-C) to the reprojection residuals (see Fig. 1).

Our ablation analyses and results on various datasets show that our approach consistently achieves the **highest accuracy** through per-sequence, per-dataset and single-global parameter fine-tuning of our model (Section IV-B) . Additionally, we introduce a **normalized Absolute Trajectory Error** to pinpoint operational settings where parameter selection maximizes overall accuracy across a set of heterogeneous sequences with very different error magnitudes (Section IV-A). Our experiments illustrate how our approach effectively **reduces the dependency on specific training data** (Section IV-C), making it a valuable contribution to a wide range of robotics and computer vision applications where reducing fine-tuning efforts is important.

II. Related Work

Filtering [1], [29], [30], [31] and Bundle Adjustment (BA) approaches [32], [3], [2], [8], [33] represent the two main methodologies for real-time camera motion and structure estimation from a monocular video. BA-based V-SLAM, involving batch optimization over several spatially distributed keyframes, has shown a higher accuracy than filtering-based methods with equivalent computational budget [34]. This section provides a brief review of outlier rejection methods in the context of BA-based SLAM.

RANSAC [35] have been extensively used to detect outliers in feature-based Structure from Motion, in particular for relative motion [5] and visual odometry [36]. Although generally robust, RANSAC is limited by its high computational cost when the outlier ratio or the minimal instantiation set grow.

An alternative approach involves integrating the outlier classification into the alignment optimization. Typically, this is done by reformulating the maximum-likelihood optimization to account for outlier measurements and their solution method, Iteratively Reweighted Least Squares (IRLS) [37]. **Robust cost functions** [38], [14], [17] play a crucial role in BA when dealing with spurious data since error functions exhibiting subquadratic growth can significantly reduce the influence of such outliers. Both approaches classify and reweight the influence of each measurement on the optimization problem based on the current estimate at each iteration assuming that outlier measurements follow a different noise model than inliers.

		Outlier Detection Tecnique
DSO	[17]	Photometric Median-based adaptative thresholding.
ORB-SLAM2	[16]	Chi-squared test over resolution-normalized residuals
BAD-SLAM	[19]	Normalized geo./photo. residuals with Tukey's/Huber.
OKVIS	[12]	Chi-square test using pose from IMU integration.
VINS-MONO	[18]	RANSAC with a fundamental matrix model.
DROID-SLAM	[22]	Deep Learning Approach
OURS		Online threshold from *partial-fitting* of Γ distribution.

TABLE I: Overview of **outlier rejection in popular V-SLAM implementations**. Ours is the first implementation that utilizes a *Gamma* distribution as a prior for outlier rejection.

Drawn from target-tracking literature for culling unlikely associations, **validation gating** [39] applies a statistical test to potential measurements, verifying if they fit the assumed statistical model. The test is applied on individual measurements, via a threshold on the *Gaussian* noise distribution [40]. This reduces the search space of potential associations, providing substantial computational savings. The gating can be extended to account for the uncertainty in the sensor pose [41], considering multiple associations simultaneously (batch gating) [40], or simultaneously consider outliers in multiple measurement types [42].

Robust cost functions and statistical tests comprise the majority of outlier rejection methods in the current literature for feature-based Bundle Adjustment, where they are typically used in a pairwise manner. These techniques rely heavily on the accuracy of the **visual covariances** associated with the features. However, the vast majority of them assume an isotropic *Gaussian* distribution for reprojection errors, despite evidence to the contrary found in real-world data. The work of Yang *et al.* [43] implicitly emphasizes the importance of visual covariances by examining factors such as photometric calibration, motion bias, and rolling shutter, and their impact on direct, feature-based, and semi-direct odometries. Kerl *et al.* [44] introduced weightings for photometric residuals by fitting errors to a t-student distribution. Similarly, Gomez-Ojeda *et al.* [45] developed a robust probabilistic model for projection errors following *Gamma* distributions for Stereo-VO. Finally, Fontan *et al.* [46] derived a model for the covariance of visual residuals, capturing how local 2D patches undergo perspective deformation when imaging 3D surfaces around a point. Table I present a brief review of how state-of-the-art VO/V-SLAM systems implement outlier rejection in their BA optimizations.

In the case of ORB-SLAM2 [16], visual residuals are scaled proportionally to the resolution at which ORB features are detected. A chi-squared test is applied to identify observations as outliers, and they are subsequently discarded during the optimization process, both in the middle and at the end of the optimization. DSO [17] apply a gradient-dependent weighting to reduce the impact of photometric errors in pixels with high gradient. In both approaches, a robustified Huber cost function is utilized, and the noise model follows the standard isotropic *Gaussian* assumption. In BAD-SLAM [19], depth-based geometric and photometric residuals

are normalized using covariance values estimated through uncertainty propagation of the measurement uncertainties. These normalized residuals are then combined into a single cost function, which is weighted using Tukey's biweight and the Huber robust loss function respectively. In Visual-Inertial OKVIS [12], outlier rejection applies a chi-square test in image coordinates, utilizing the uncertain pose predictions obtained through IMU integration. VINS-Mono [18] employs RANSAC with a fundamental matrix model [5]. New learning-based approaches, such as DROID-SLAM [22], are able to estimate 2D-correspondendes together with an associated confidence map.

In the context of SLAM in dynamic environments, traditional methods often classify points associated with dynamic objects as outliers and exclude them from the optimization due to the assumption of environment staticity [14], [17], [19]. Newer approaches address this challenge by utilizing Convolutional Neural Networks (CNN) for pixel-wise segmentation of dynamic objects like people and cars in the frames. This segmentation allows the SLAM algorithm to avoid extracting features from these dynamic elements [23], [25]. More advanced strategies aim to jointly optimize the structures of the static scene and the dynamic objects, along with the trajectories of both the camera and the moving agents, within a bundle adjustment framework [23], [24].

III. Formulation

BA in V-SLAM jointly optimizes keyframe poses $\mathbf{T}_k \in \mathrm{SE}(3)$ and 3D map point locations $\mathbf{X}_p \in \mathbb{R}^3$ through least-squared-error minimization of the reprojection residuals, weighted by their associated covariances. Outlier rejection is achieved by filtering out the observations of points with the highest normalized residuals. In this section, we provide formulations for both standard BA and outlier rejection. After that, we derive our online adaptive thresholding for outlier rejection, based on a *partial-fitting* of a *Gamma* distribution.

A. Bundle Adjustment (BA)

BA optimizes the 3D positions of the map points $\mathbf{X}_p \in \mathbb{R}^3$ and keyframe poses $\mathbf{T}_k \in \mathrm{SE}(3)$ by minimizing their reprojection error with respect to the matched keypoints $\mathbf{x}_{p,k} \in \mathbb{R}^2$. The error term for the observation of a map point p in a keyframe k is:

$$\mathbf{e}_{p,k} = \mathbf{x}_{p,k} - \Pi(\mathbf{R}_k \mathbf{X}_p + \mathbf{t}_k), \tag{1}$$

where $\Pi(\cdot)$ (determined by the intrinsic camera parameters) projects $\mathbf{X}_p$ expressed in the camera coordinates in the image; and $\mathbf{R}_k \in \mathrm{SO}(3)$ and $\mathbf{t}_k \in \mathbb{R}^3$ are respectively the rotation matrix and translation vector corresponding to the pose $\mathbf{T}_k$.

The cost function to be minimized is:

$$\sum_{p,k} \rho_h(\mathbf{e}_{p,k}^T \mathbf{\Omega}_{p,k}^{-1} \mathbf{e}_{p,k}), \tag{2}$$

where ρ_h is a robust cost function (e.g. Huber) and $\mathbf{\Omega}_{p,k} = \sigma_{p,k}^2 \mathbf{I}_{2\times 2}$ is the covariance matrix associated to the matched keypoint $\mathbf{x}_{p,k}$ [14].

Normalized Mahalanobis residuals. As shown in equation (2), the residual form to be minimized in the optimization commonly has the form of the Mahalanobis norm of the reprojection error $r_{p,k}^2 = \mathbf{e}_{p,k}^T \mathbf{\Omega}_{p,k}^{-1} \mathbf{e}_{p,k}$.

B. Gamma Distribution $\Gamma(r^2|\alpha,\beta)$

We model r^2 (subindices p, k are dropped for simplicity) as a *Gamma* distribution, i.e. $r^2 \sim \Gamma(\alpha,\beta)$, whose Probability Density Function (PDF) is given by

$$\Gamma(r^2|\alpha,\beta) = \frac{1}{\beta^\alpha \gamma(\alpha)} (r^2)^{\alpha-1} e^{\frac{-r^2}{\beta}} \quad , \quad r^2, \alpha, \beta > 0, \tag{3}$$

where α and β are the so-called shape and scale parameters, respectively, and $\gamma(\cdot)$ is the Gamma function. The two parameters of this *Gamma* distribution are estimated through the minimization of the negative log-likelihood, which is given by $-\sum \log \Gamma(r_i^2|\alpha,\beta)$, using the histogram of the residual magnitudes. Previous approaches used robust methods to reduce the effect of high outlier ratios in the estimation of the distribution parameters [45]. We propose the novel approach of a *partial-fitting* of the distribution (as detailed in Section III-C). This online procedure introduces minimal overload to the optimization (see profiling results in Table II). However, it provides two significant benefits: a more precise camera pose estimation and increased robustness against outliers and noisy measurements compared to the standard *Gaussian*-based assumption.

Goodness of Fit (GoF). Once the distribution parameters have been estimated, we evaluate the goodness of fit:

$$GoF = \frac{1}{n} \sum_{i=1}^{n} \left(\Gamma_{cdf}\left(r_i^2|\alpha,\beta\right) - p_i\left(r_i^2\right) \right)^2, \tag{4}$$

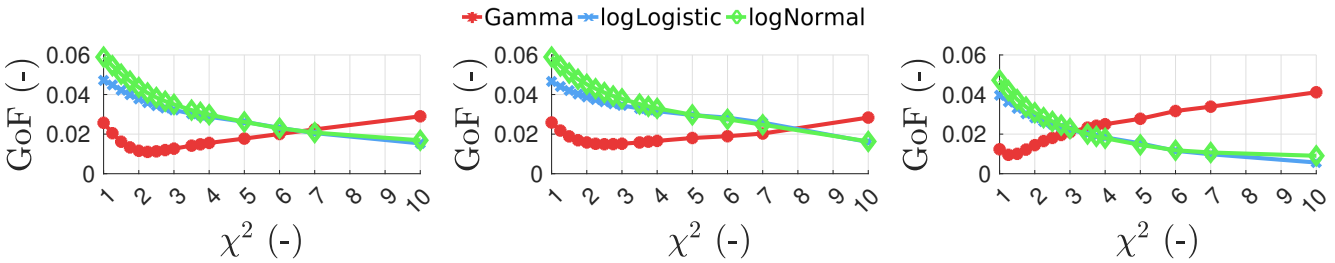
where

$$\Gamma_{cdf}\left(r^2|\alpha,\beta\right) = \int_0^{r^2} \Gamma\left(u|\alpha,\beta\right)\, du \tag{5}$$

is the Cumulative Distribution Function, r_i^2 are n equidistant values in the interval $r_i^2 \in [0, \Gamma_{cdf}^{-1}(0.95|\alpha,\beta)]$, and p_i is the percentage of residuals smaller than r_i^2.

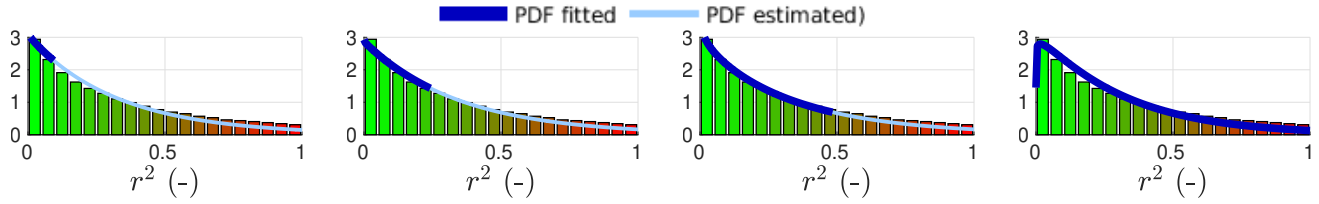
Similar to Kerl et al. [44] demonstrating that the photometric error of an RGB-D SLAM approach is better fitted by a *t-student* distribution than a *Gaussian*, Gomez-Ojeda et al. [45] showed that a *Gamma* distribution is more suitable than a *chi-squared* distribution for modeling the normalized Mahalanobis residuals of a feature-based stereo odometry. We analyze the goodness of fit of the *Gamma* distribution to the residuals of feature-based monocular SLAM and compare it with two other common probabilistic distributions for non-negative random variables with scale and shape parameters, namely, *LogNormal* and *LogLogistic*.

The shape and scale of the distribution of residuals can vary depending on the scenario, scene conditions, or the type of residuals, such as feature-based or photometric. Even different descriptors of 2D image keypoints may lead to changes in the distribution. In Figure 2, we present the GoF for the distributions to the reprojection errors after BA in ORB-SLAM2 [16] across three sequences from three different



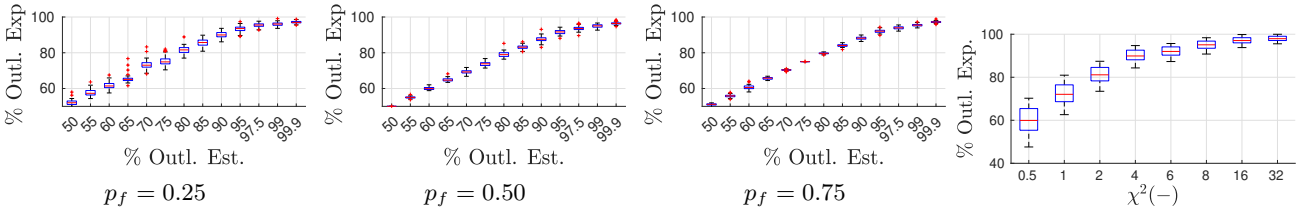
fr3 long office - RGB-D TUM [47] *table 3 - ETH Benchmark* [19] *00 - KITTI Dataset* [48]

Fig. 2: **Goodness of Fitting (GOF).** We assessed the suitability of three continuous probability distributions—namely, *Gamma*, *LogNormal*, and *LogLogistic*—for modeling a non-negative random variable. The accompanying figure illustrates the GoF for these distributions (Section III-B) when the data is filtered based on a threshold χ^2. The *Gamma* distribution provides a superior fit for smaller reprojection error values but deteriorates as χ^2 increases, indicating a higher presence of outliers. In contrast, *LogNormal* and *LogLogistic* exhibit better fitting to errors as the likelihood of outliers increases.



(a) $p_f = 0.25$ (b) $p_f = 0.5$ (c) $p_f = 0.75$ (d) $p_f = 1.0$

Fig. 3: **Partial fitting of $\Gamma(\alpha, \beta)$ (*Gamma* distribution).** The *PDF fitted line* (Probability Density Function) represents the portion of the data used to estimate the parameters α, β of $\Gamma(\alpha, \beta)$. The *PDF estimated line* displays the distribution estimation for the remaining data. It is noteworthy that even with a small amount of the data (i.e., $p_f = 0.25$), we can obtain accurate estimates of the distribution parameters, as demonstrated in Figure 4. **These histograms have been truncated at $r^2 = 1.0$ for visualization purposes.*



$p_f = 0.25$ $p_f = 0.50$ $p_f = 0.75$

Fig. 4: **Outlier estimation.** The subfigures depict the percentage of outliers found (*% Outl. Exp.*) when varying the targeted percentage (*% Outl. Est.*) while considering different amounts of fitted data (p_f). In our subsequent experiments, we have chosen to use 50% of the data ($p_f = 0.5$). This choice strikes a balance between accurately estimating the outlier percentage and using a relatively small amount of data to prevent overfitting in the estimation of the $\Gamma(\alpha, \beta)$ distribution. The figure in the right side illustrates the *% Outl. Exp.* when varying the static threshold χ^2 used in the BA optimization of ORB-SLAM2 [16] for outlier rejection. Notice the increased variability in the estimated percentage compared to our results shown in the figures on the left.

datasets when filtering the residuals r^2 using a threshold χ^2 (**we use the threshold χ^2 over the residuals r^2 as a parameter for our experiments**). Figure 3 illustrates how a *Gamma* distribution accurately captures the behavior of the residual magnitude, as it exhibits a certain bias and a heavy tail.

C. Partial-fitting of a Gamma distribution

In Figure 4, we can observe the accuracy of our approach in estimating outliers when fitting the *Gamma* distribution using 25%, 50%, and 75% of the lowest residuals. Clearly, the accuracy improves as we use a larger amount of data for fitting the distribution. However, even when using a small amount of data (e.g., $p_f = 0.25$), the accuracy remains high. Moreover, as shown on the right in Figure 4, when using a fixed threshold χ^2 for outlier filtering, the variability in the estimated outlier percentages is significantly larger compared to our method. Note the variability between 50% and 70% for a χ^2 value of 0.5. These significant variability differences between our online method and the standard fixed χ^2 threshold illustrates a higher accuracy in the outlier rejection, that will in turn improve the SLAM accuracy as shown in Section IV.

	mean	std	mean + std	median	rmse
Tracking	29.81	9.21	39.02	28.44	31.19
Mapping	453.43	197.75	651.18	371.32	493.68
- Dist. Est.	4.56	0.45	5.01	4.32	4.58

TABLE II: **Timing (ms)**: our implementation of the distribution estimation occurs within the BA in the mapping thread, adding negligible time overhead to the full pipeline.

					Hz	Res.
RGB-D TUM [47]	Indoor	Real	Office	Handheld	30	640 x 480
ETH [19]	Indoor	Real	Living Room	Handheld	40	739 x 458
KITTI [48]	Outdoor	Real	Road	Car	10	1241 x 376

TABLE III: **Dataset characteristics.** We used sequences from 3 datasets, utilizing ORB-SLAM2 as underlying pipeline [16]. These sequences show significant variations in terms of image noise, resolution, scene content, and application domain.

Once the distribution is fitted, we can use the estimated parameters to calculate an adaptive threshold for a given confidence value p. **We use the probabilistic threshold p as a parameter for our experiments**.

IV. Experiments

The primary goal of our distribution-based outlier rejection approach is to enhance the accuracy and versatility of SLAM. To achieve this, we conducted a comprehensive set of experiments encompassing a diverse range of scenarios.

A. Methodology

We implemented our methods within the local BA in the mapping thread of the state-of-the-art feature-based monocular ORB-SLAM2 [16]. We also employ the original ORB-SLAM2 implementation as the baseline method for our comparisons. To minimize the influence of real-time constraints in SLAM, which could potentially impact our evaluation results, we utilize a sequential version of ORB-SLAM2[1]. To account for the influence of other non-deterministic effects, our accuracy metrics are calculated over 10 runs of each sequence.

Real-time estimation is a critical requirement for SLAM. Table II presents the timing performance of the tracking and mapping threads. Our implementation of the outlier distribution estimation occurs within the local bundle adjustment optimization in the mapping thread, adding negligible time overhead to the complete pipeline.

We evaluate our approach using the **Absolute Trajectory Error** (ATE) [47], which calculates the root-mean-squared of the translation errors. In order to find the operating points where the parameter selection maximizes the overall accuracy of a pipeline in a set of heterogeneous sequences, such as the set of values where the trajectory estimation is better for a car sequence (e.g. [48]) compared to an office scenario (e.g. [47]), where errors are in meters and centimeters, we define a normalized ATE.

Let E_j be the set of n ATE values obtained from one evaluation in one sequence j, normalized by the absolute minimum of the ATE found in that sequence from all experiments, configurations, and approaches (both ours and vanilla). We define the hat operator as the mean plus one standard deviation over one set of values: $\hat{x} = \mu(x) + \sigma(x)$. Finally, we define a general **normalized ATE** $\hat{E}$ over a set of N sequences being $E = \{\hat{E}_j | \hat{E}_j \text{ with } j \in \{0, N\}\}$ the set of normalized ATE values of each sequence. Note that this normalization procedure can be applied to any other evaluation metric to aid selecting operational points for SLAM algorithms running in various applications and scenarios.

As explained in Section I, our goal is to explore the ability of SLAM pipelines to generalize to different applications and scenarios, finding approaches with generalizable parameters that minimize fine-tuning efforts. In Section IV-C, we evaluate **inter-sequence accuracy matrices**. We fill these matrices M_{ij} by selecting the normalized ATE value obtained in a sequence i with the set of settings corresponding to the configuration that yielded the minimal normalized ATE in sequence j. By analyzing the values collected in these matrices, we are able to study the generalization capabilities

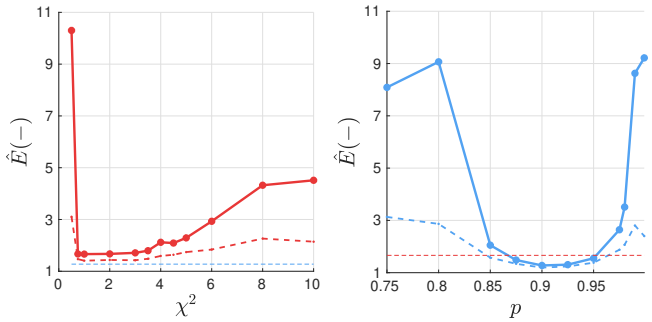
[1]The sequential version of ORB-SLAM2 used in our experiments can be found at: https://github.com/alejandrofontan/ORB_SLAM2_Deterministic.

			Per-sequence		Per-dataset		Global	
			χ^2	p	χ^2	p	χ^2	p
RGB-D TUM	fr1 desk	†(cm)	1.61 (3.0)	**1.60 (0.98)**	1.76	**1.73**	**1.64**	1.73
	xyz 1		**0.94 (1.25)**	1.00 (0.95)	**0.94**	1.03	**1.02**	1.03
	desk 2		0.99 (1.25)	**0.84 (0.9)**	1.03	**0.84**	1.03	**0.84**
	xyz 2		0.30 (0.75)	**0.26 (0.9)**	0.33	**0.26**	0.40	**0.26**
	long office		3.02 (0.25)	**1.06 (0.92)**	3.27	**1.10**	4.75	**1.10**
	nstr text near		1.23 (2.0)	**1.17 (0.9)**	1.24	**1.17**	1.36	**1.17**
	str text far		1.00 (0.75)	**0.97 (0.95)**	1.26	**1.24**	1.24	**1.22**
KITTI	03	†(m)	**0.69 (3.0)**	1.08 (0.95)	1.12	**1.08**	**0.69**	1.95
	04		**0.15 (0.75)**	0.20 (0.87)	**0.15**	0.22	0.48	**0.20**
	07		2.76 (10.0)	**1.86 (0.99)**	**3.23**	4.72	**3.56**	6.43
ETH	large loop 1	†(cm)	1.60 (2.0)	**1.43 (0.92)**	1.60	**1.43**	1.97	**1.62**
	repetitive		0.65 (2.0)	**0.60 (0.92)**	0.65	**0.60**	0.70	**0.61**
	table 3		0.71 (2.0)	**0.54 (0.92)**	0.71	**0.54**	0.96	**0.63**

TABLE IV: **Accuracy results** (*better* ↓): we evaluate trajectory estimation accuracy using the ATE metric across 13 sequences spanning 3 datasets. Our experiments involve ablation studies where we vary both the vanilla ORB-SLAM2 threshold χ^2 for outlier rejection and our probabilistic parameter p. We discern the best accuracy achieved when selecting best parameters per sequence (χ^2 and p in brackets on the table). Our approach achieves the highest accuracy in 10 out of 13 sequences. Furthermore, when employing the normalized ATE to determine the best parameter on both a per-dataset and global basis (as indicated by the corresponding values in Table V), our approach yet again achieves the highest accuracy in 10 and 9 out of 13 sequences, respectively. †
Please note the varying units for the ATE depending on the dataset.

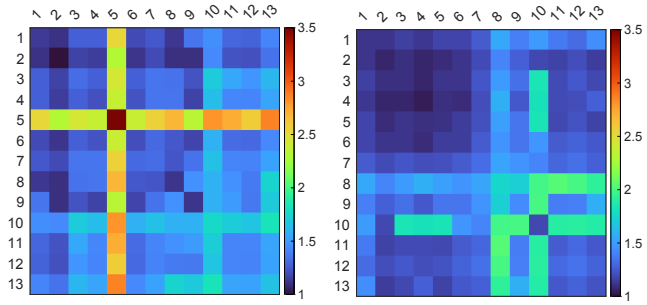
RGB-D TUM		KITTI		ETH		Global	
χ^2	p	χ^2	p	χ^2	p	χ^2	p
1.28	**1.16**	2.04	**2.02**	1.52	**1.26**	1.64	**1.21**
(1.0)	**(0.9)**	(0.75)	**(0.95)**	(2.0)	**(0.92)**	(3.0)	**(0.9)**

TABLE V: **Normalized ATE** (*better* ↓): We identify the best accuracy on a per-dataset and global basis employing the normalized ATE, $\hat{E}$ (Section IV-A). Our approach yields overall the highest accuracy in all cases. Table IV shows the absolute ATE values independently for each sequence for the best parameter (in brackets on the table).



(a) Vanilla (χ^2) (b) Ours (p)

Fig. 5: **Accuracy results with global threshold selection.** We present the combined normalized ATE, $\hat{E}$ (Section IV-A) for all sequences in Table IV while varying both the vanilla ORB-SLAM2 threshold χ^2 for outlier rejection (left) and our probabilistic parameter p (right). It's noteworthy that 1) our method achieves significantly lower trajectory errors at the point of highest accuracy, and 2) our approach exhibits less sensitivity to disturbances around the optimal parameter, as evident from the smoother derivative of the curve.



(a) Vanilla (χ^2) (b) Ours (p)

Fig. 6: **Inter-sequence accuracy matrix** (*better* ↓). We present the normalized ATE for each sequence (rows), evaluated using thresholds corresponding to the highest accuracy performance of the remaining sequences (columns). It's worth noting that our online adaptive thresholding (on the right) exhibits less sensitivity to disturbances compared to the fixed thresholding of ORB-SLAM2 (on the left). Numerical evaluations of these matrices are provided in Table VI.

of different approaches.

Our extensive evaluation covers 13 sequences across 3 publicly available **datasets** for the evaluation of V-SLAM. It should be noted that to assess the versatility of our approach and its ability to work across complementary datasets with minimal parameter fine-tuning, our evaluation benchmark is significantly larger than previous state-of-the-art SLAM works [16], [17], [49], [19]. Table III provides an overview of the dataset characteristics, including factors like image resolution, frequency, indoor/outdoor scenes, and real/synthetic images.

B. Accuracy evaluation

Table IV provides a comprehensive evaluation of estimated trajectory errors using the ATE metric across sequences

	mean	std	mean + std	median	rmse
Vanilla (χ^2)	1.55	0.44	2.00	1.77	1.98
Ours (p)	**1.25**	**0.13**	**1.35**	**1.35**	**0.72**

TABLE VI: **Numerical evaluations of inter-sequence accuracy matrix** (*better* ↓). As depicted in Figure 6, our adaptive approach demonstrates lower sensitivity to disturbances when compared to fixed thresholding.

spanning 3 datasets. Our ablation experiments involve varying the vanilla ORB-SLAM2 threshold χ^2 for outlier rejection and our probabilistic parameter p. We identify the best accuracy when selecting parameters on a per-sequence basis, and our approach consistently outperforms others, achieving the highest accuracy in 10 out of 13 sequences.

Furthermore, when employing the normalized ATE $\hat{E}$ to determine the best parameter on both a per-dataset and global basis (as indicated by the corresponding values in Table V), our approach yet again achieves the highest accuracy in 10 out of 13 sequences and 9 out of 13 sequences, respectively.

Figure 5 complements Table V by visualizing the combined normalized ATE (as detailed in section IV-A) across all sequences in Table IV. We vary both the vanilla ORB-SLAM2 threshold χ^2 for outlier rejection (on the left) and our probabilistic parameter p (on the right). Notable observations include the fact that our method achieves significantly lower trajectory errors at the point of highest accuracy. Moreover, our approach exhibits less sensitivity to disturbances around the optimal parameter, as evidenced by the smoother derivative of the curve.

C. Inter-sequence accuracy evaluation

In Figure 6, each row represents the normalized ATE for a specific sequence, evaluated using thresholds optimized based on the highest accuracy observed in other sequences (columns). Notably, our online adaptive thresholding (depicted on the right) displays remarkable resilience to disturbances, in stark contrast to the fixed thresholding of ORB-SLAM2 (depicted on the left). We provide precise numerical assessments of these matrices in Table VI, reaffirming the superiority of our adaptive approach in achieving lower inter-sequence accuracy variations (*better* ↓).

V. CONCLUSIONS

In summary, we have introduced a novel framework for outlier rejection in feature-based Bundle Adjustment for Visual SLAM. Our approach utilizes a *Gamma* distribution as a prior to estimate an adaptive threshold in real-time. Through extensive experiments and ablation studies across 13 sequences spanning 3 complementary datasets, we have provided validation for the proposed approach, which has been shown to be effective, and provides a considerable improvement over the current state-of-the-art methods.

Significantly, our approach excels in generalizing to changes in the application domain or environmental conditions without the need for specific parameter training. This makes it particularly relevant for robotic applications with limited access to training data or operating in dynamic environments.

REFERENCES

[1] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

[2] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces," in *2007 6th IEEE and ACM international symposium on mixed and augmented reality*, pp. 225–234, IEEE, 2007.

[3] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Real time localization and 3d reconstruction," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 363–370, IEEE, 2006.

[4] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings*, pp. 298–372, Springer, 2000.

[5] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[6] N. Demmel, C. Sommer, D. Cremers, and V. Usenko, "Square root bundle adjustment for large-scale reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11723–11732, 2021.

[7] S. Weber, N. Demmel, T. C. Chan, and D. Cremers, "Power bundle adjustment for large-scale 3d reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 281–289, 2023.

[8] H. Strasdat, J. Montiel, and A. J. Davison, "Scale drift-aware large scale monocular slam," *Robotics: science and Systems VI*, vol. 2, no. 3, p. 7, 2010.

[9] D. Gálvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.

[10] K. A. Tsintotas, L. Bampis, and A. Gasteratos, "Visual place recognition for simultaneous localization and mapping," *Autonomous Vehicles Volume 2: Smart Vehicles*, pp. 47–79, 2022.

[11] K. A. Tsintotas, L. Bampis, and A. Gasteratos, "The revisiting problem in simultaneous localization and mapping: A survey on visual loop closure detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 19929–19953, 2022.

[12] S. Leutenegger, P. Furgale, V. Rabaud, M. Chli, K. Konolige, and R. Siegwart, "Keyframe-based visual-inertial slam using nonlinear optimization," *Proceedings of Robotis Science and Systems (RSS) 2013*, 2013.

[13] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13*, pp. 834–849, Springer, 2014.

[14] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[15] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "Svo: Semidirect visual odometry for monocular and multicamera systems," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, 2016.

[16] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[17] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017.

[18] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[19] T. Schops, T. Sattler, and M. Pollefeys, "BAD SLAM: Bundle Adjusted Direct RGB-D SLAM," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[20] V. Usenko, N. Demmel, D. Schubert, J. Stückler, and D. Cremers, "Visual-inertial mapping with non-linear factor recovery," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 422–429, 2019.

[21] A. Rosinol, A. Violette, M. Abate, N. Hughes, Y. Chang, J. Shi, A. Gupta, and L. Carlone, "Kimera: From slam to spatial perception with 3d dynamic scene graphs," *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1510–1546, 2021.

[22] Z. Teed and J. Deng, "Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras," *Advances in neural information processing systems*, vol. 34, pp. 16558–16569, 2021.

[23] B. Bescos, J. M. Fácil, J. Civera, and J. Neira, "Dynaslam: Tracking, mapping, and inpainting in dynamic scenes," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4076–4083, 2018.

[24] B. Xu, W. Li, D. Tzoumanikas, M. Bloesch, A. Davison, and S. Leutenegger, "Mid-fusion: Octree-based object-level multi-instance dynamic slam," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 5231–5237, IEEE, 2019.

[25] I. Ballester, A. Fontán, J. Civera, K. H. Strobl, and R. Triebel, "Dot: Dynamic object tracking for visual slam," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11705–11711, IEEE, 2021.

[26] Y. Zhao and P. A. Vela, "Good feature selection for least squares pose optimization in vo/vslam," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1183–1189, IEEE, 2018.

[27] Y. Zhao and P. A. Vela, "Good feature matching: Toward accurate, robust vo/vslam with low latency," *IEEE Transactions on Robotics*, vol. 36, no. 3, pp. 657–675, 2020.

[28] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[29] A. Azarbayejani and A. P. Pentland, "Recursive estimation of motion, structure, and focal length," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 6, pp. 562–575, 1995.

[30] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from motion causally integrated over time," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 4, pp. 523–535, 2002.

[31] J. Civera, O. G. Grasa, A. J. Davison, and J. Montiel, "1-point ransac for ekf-based structure from motion," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3498–3504, IEEE, 2009.

[32] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1, pp. I–I, Ieee, 2004.

[33] J. Lim, J.-M. Frahm, and M. Pollefeys, "Online environment mapping," in *CVPR 2011*, pp. 3489–3496, IEEE, 2011.

[34] H. Strasdat, J. M. Montiel, and A. J. Davison, "Visual slam: why filter?," *Image and Vision Computing*, vol. 30, no. 2, pp. 65–77, 2012.

[35] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[36] F.-A. Moreno, J.-L. Blanco, and J. González-Jiménez, "Erode: An efficient and robust outlier detector and its application to stereovisual odometry," in *2013 IEEE International Conference on Robotics and Automation*, pp. 4691–4697, IEEE, 2013.

[37] Z. Zhang, "Parameter estimation techniques: A tutorial with application to conic fitting," *Image and vision Computing*, vol. 15, no. 1, pp. 59–76, 1997.

[38] A. Concha and J. Civera, "An evaluation of robust cost functions for rgb direct mapping," in *2015 European Conference on Mobile Robots (ECMR)*, pp. 1–8, IEEE, 2015.

[39] Y. Bar-Shalom, T. E. Fortmann, and P. G. Cable, "Tracking and data association," 1990.

[40] J. Neira and J. D. Tardós, "Data association in stochastic mapping using the joint compatibility test," *IEEE Transactions on robotics and automation*, vol. 17, no. 6, pp. 890–897, 2001.

[41] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (slam): Part ii," *IEEE robotics & automation magazine*, vol. 13, no. 3, pp. 108–117, 2006.

[42] C. H. Tong and T. D. Barfoot, "Batch heterogeneous outlier rejection for feature-poor slam," in *2011 IEEE International Conference on Robotics and Automation*, pp. 2630–2637, IEEE, 2011.

[43] N. Yang, R. Wang, X. Gao, and D. Cremers, "Challenges in monocular visual odometry: Photometric calibration, motion bias, and rolling shutter effect," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 2878–2885, 2018.

[44] C. Kerl, J. Sturm, and D. Cremers, "Robust odometry estimation for rgb-d cameras," in *2013 IEEE international conference on robotics and automation*, pp. 3748–3754, IEEE, 2013.

[45] R. Gomez-Ojeda, F.-A. Moreno, and J. Gonzalez-Jimenez, "Accurate stereo visual odometry with gamma distributions," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1423–1428, IEEE, 2017.

[46] A. Fontan, L. Oliva, J. Civera, and R. Triebel, "Model for multi-view residual covariances based on perspective deformation," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1960–1967, 2022.

[47] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A Benchmark for the Evaluation of RGB-D SLAM Systems," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

[48] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[49] Y. Zhou, H. Li, and L. Kneip, "Canny-vo: Visual odometry with rgb-d cameras based on geometric 3-d–2-d edge alignment," *IEEE Transactions on Robotics*, vol. 35, no. 1, pp. 184–199, 2018.