

Camera Calibration without Camera Access - A Robust Validation Technique for Extended PnP Methods

Emil Brissman^{1,2}[0000-0002-0418-9694], Per-Erik Forssén¹[0000-0002-5698-5983],
and Johan Edstedt¹[0000-0002-1019-8634]

¹ Computer Vision Laboratory, Dept. EE, Linköping University, Sweden

² Saab, Sweden

{emil.brissman, per-erik.forssen, johan.edstedt}@liu.se

Abstract. A challenge in image based metrology and forensics is intrinsic camera calibration when the used camera is unavailable. The unavailability raises two questions. The first question is how to find the *projection model* that describes the camera, and the second is to detect incorrect models. In this work, we use off-the-shelf extended PnP-methods to find the model from 2D-3D correspondences, and propose a method for model validation. The most common strategy for evaluating a projection model is comparing different models’ residual variances—however, this naive strategy cannot distinguish whether the projection model is potentially underfitted or overfitted. To this end, we model the residual errors for each correspondence, individually scale all residuals using a predicted variance and test if the new residuals are drawn from a standard normal distribution. We demonstrate the effectiveness of our proposed validation in experiments on synthetic data, simulating 2D detection and Lidar measurements. Additionally, we provide experiments using data from an actual scene and compare non-camera access and camera access calibrations. Last, we use our method to validate annotations in MegaDepth.

1 Introduction

Intrinsic camera calibration is a fundamental computer vision problem. It involves finding the parameters that allow the conversion of pixel coordinates to bearing angles [12]. It is possible to use the camera for *metrology* using a calibration. In the single-view case, metrology means measuring the lengths and angles of objects depicted in an image. As an extension, it is the underpinning of single view 3D reconstruction [4]. Metrology has many applications, including non-contact measurements, sensor fusion, and forensic analysis.

Traditionally, intrinsic calibration is a semi-automatic process, which involves imaging of calibration objects [30,28]. Such calibration allows controlled accuracy; however, access to the camera is required. In forensic analysis, the camera is only sometimes available, depending on the received material. Therefore, we

aim to facilitate measurements in an image when the camera is unavailable. Using a calibration profile from a camera of the same model often works well, but the accuracy is unknown in this approach and should thus be avoided in forensics.

In the Perspective-n-Point (PnP) problem, the goal is to estimate the camera pose given a set of 2D-3D point correspondences. Early methods assume a calibrated camera, and only estimate translation and rotation parameters [7]. More recent variants of PnP also estimate the intrinsic camera parameters [14,21]. These *extended PnP methods* (xPnP) do not require the camera to be available, in contrast to calibration pattern methods [30,28]. However, they introduce new challenges such as 2D-3D matching and validation.

In this work, we attend to the validation of camera calibration for forensic metrology applications [2]. Usually, a model is assumed to be validated if it, on average, has low residuals. However, this approach will not provide any measure of uncertainty in the image plane. Moreover, deriving the uncertainty is challenging because the amount of distortion scales non-linearly with the distance to the camera centre. Thus, we treat noise modelling as a robust regression problem and predict a residual scaling for each 2D-3D correspondence. When the model is correct, we assume the scaled residuals to follow a standard normal distribution (Figure 1). Next, to verify this assumption, we use a hypothesis test. Simulated data, an indoor scene and MegaDepth [18], with annotated cameras depicting different scenes, demonstrate our proposed validation.

Contributions Our contributions are as follows: **(i)** We propose a method for testing residuals based on variance predictions and standardisation. **(ii)** We suggest using xPnP methods for unavailable cameras as input to our method, given 2D-3D correspondences. **(iii)** An empirical estimate of the variance scales residuals poorly. Instead, we propose a predictive noise model to scale individual residuals over the 2D detector and projected 3D noise. **(iv)** We analyse the effectiveness of our method in quantitative and qualitative experiments and

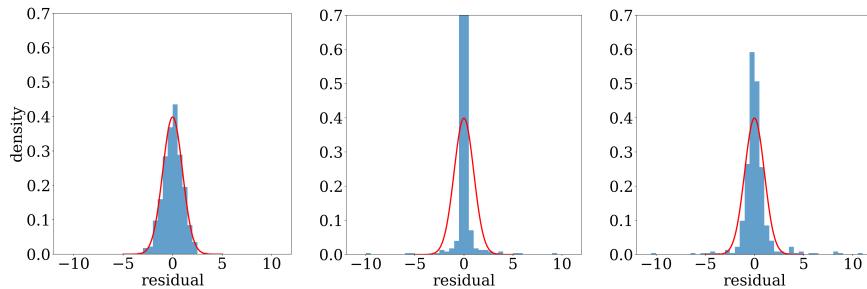


Fig. 1. Left: Standardised residuals for a correct model with one distortion parameter using our robust scale estimate. I.e. the residuals are not affected by the model error. Middle: Standardised residuals from images under more distortion, for an incorrect model using a non-robust scale estimate. Right: Standardised residuals for an incorrect model, using a robust scale estimate.

demonstrate its ability to significantly predict incorrect models, also when the mean of the residuals is low.

1.1 Background Motivation

At the Swedish National Forensic Center (NFC), the task is to collect information linked to crimes without the possibility of misinterpretations when used in the Swedish court system. At the time of writing, NFC uses the Zhang method [30] for metrology. However, this commonly accepted practice only applies to images where the camera is available. Lidar scanning is a standard technique in forensic investigations, and in many legal cases, the depicted location is revisited for scanning. Using this working methodology, Olsson [22] first investigated the validation of xPnP methods that forms the basis of this work. In [22], model correctness is assessed by checking if the empirical mean of the re-projected sample distances is within the two centre quartiles. However, this decision will prefer incorrect models since outliers will expand the decision range.

1.2 Ethical Consideration

This work does not concern any police investigations or legal cases. Instead, the method we propose analyses residuals using synthetic data, available benchmark data, and a snapshot of a fictional crime scene provided by a police agency. These are all free from apparent ethical dilemmas.

2 Related Work

Semi-automatic Calibration Camera calibration is a broad subject found in many areas of industry and research. However, the most common camera calibration practice is to use a printed pattern on a planar surface. This strategy was proposed by Zhang [30], who suggested using a checkerboard pattern with equidistant squares of black and white. The inner corners of the pattern form unambiguous features that are easy to find. Detecting several of these features, also called saddle points, between different views allows camera parameters to be estimated. Each detected saddle point in each picture is assigned to its corresponding point on the checkerboard. This set of correspondences represents a series of homographies, determining the intrinsic and extrinsic parameters for one or more cameras. In the case of Tsai [28], camera calibration depends only on one view of a co-planar checkerboard pattern. The Zhang method [30], instead depends on at least three pictures of a planar checkerboard pattern. More recently, deep methods like Li *et al.* [17] take a single image as input and jointly learn to predict distortion coefficients and optical flow from images with lens-type annotations. However, this problem only concerns visual quality and provides no model accuracy assessment.

Perspective-n-Point The PnP problem [7] refers to finding a rigid transformation from 2D-3D point matches. That is, to estimate the rotation matrix \mathbf{R}

and the translation vector \mathbf{t} describing the camera pose in the coordinate system of the 3D points, assuming that the intrinsic parameters never changed during the sampling of a scene. The minimal but ambiguous case, P3P [9,19,23] is not considered in this work.

Lepetit *et al.* [16] (EPnP), reduced the computational complexity to $O(n)$ operations. Although the convergence is fast, the solution depends on initialization and global convergence is not guaranteed. Later works recognized the need to include intrinsic parameters to generalize application tasks [21][14]. Nakano [21] extends the PnP problem by including intrinsic parameters and dividing the parameter estimation into different stages. Radial distortion and equal focal length horizontally and vertically are assumed, as well as fixating the principal point to the image center. Larsson *et al.* [14] instead require a minimal correspondence set and add a local optimization step [15]. In our work, we propose a method to validate xPnP methods, for application in forensic analysis, by *Goodness-of-Fit* (GoF) testing between distributions. That is, we do not improve the methods [21] and [14].

Empirical Performance Evaluation

Works by Wang *et al.* [29] and Thai *et al.* [27] are related to our work. Wang *et al.* [29] propose a method to test hypotheses about the effect of conceptual changes in deep classification models. That is, if the difference is the probable reason behind the increase in accuracy. Thai *et al.* [27] propose to identify cameras by raw pixel intensities. Similar to our approach, two quantities parametrize the intensity variance— analog gain, controlled by the camera ISO setting, and electronic noise caused during sensor readout. For an unknown camera the parameters are first estimated and secondly tested against known parameter values (null hypothesis).

Goodness of Fit The goodness of fit testing is one of the fundamental tasks in statistics. In this work, we focus on normality testing due to normal distributions being a good model for uncertainty in projective geometry [13,8]. Still, our approach could easily be generalized to GoF tests for arbitrary distributions.

There are a large variety of proposed statistics for normality testing, of which the Kolmogorov-Smirnov (KS) [20], D’Agostino-Pearson (DAP) [6], and Shapiro-Wilk (SW) [26] tests are well known. These tests all seek to maximize the power, *i.e.*, minimizing the risk of the null hypothesis being accepted, given that an alternative hypothesis is correct. We discuss those further in Section 3.5 and test all three in Section 4.

Single View Metrology Metrology is the study of measurement. In the context of computer vision, single view metrology [5] involves estimating, *e.g.*, angles and lengths, from a single image. In all metrology, an accurate measure of uncertainty is crucial, and in particular in the forensic setting. Previous work in single view metrology has focused on the undistorted (but often uncalibrated) case [5], with model uncertainty assumed to be normally distributed [5,13,8,3]. At inference time, these uncertainties can be propagated by first order propagation or by Monte Carlo simulation. However, in those works, both the uncertainty estimation and propagation requires *a priori* knowledge of the noise levels and

estimation method, and implicitly assume the estimated model is approximately correct. In contrast to those methods, our approach

1. Is estimation agnostic, *i.e.*, we can treat the estimators as black boxes.
2. Generalises to arbitrary projection models.
3. Does not implicitly assume that the estimated model is approximately correct.

In particular, perturbation theory, as used in previous work, does not provide a reliable measure of the trustworthiness of the estimated model, it simply provides an approximate measure of the estimation sensitivity to the input. In contrast, our method directly measures trustworthiness by testing the hypothesis of the matches being generated from the estimated model.

3 Method

We propose a method that compares observed and expected noise levels. The method takes residual values as input, given a calibration computed from an xPnP method and 2D-3D correspondences. We decompose the residual error for each correspondence as three additive terms: **(i)** 2D detector noise, **(ii)** 3D detector noise projected into the image, and **(iii)** model noise. The expected model noise is zero for the correct model, not affecting the residual distribution in any direction. We describe this in Section 3.3. To handle unexpected model noise, Section 3.4 details robust regression over **(i)** and **(ii)** to obtain a scale value for each 2D point. Finally, we assume the scaled residuals are drawn from a standard normal distribution and test this using a GoF test. We motivate our preferred choice of test in Section 3.5. We consider all points to influence the validation decision and believe this to improve applications in forensic analysis. We begin with an example to get a good intuition of our approach.

3.1 Motivating Example

Consider a correct data model $y = x$ and an (incorrect) hypothesis $h_{\text{bad}} : y = x + 0.5x^5$. Under the assumption that y is observed with some Gaussian noise, the residuals r of the true model will be distributed as $\mathcal{N}(0, \sigma_y^2)$. In contrast, the residuals of the incorrect hypothesis are typically *significantly* different from the expected distribution (as shown in Figure 1). Thus, if σ_y is known, a simple hypothesis test is whether $\frac{r}{\sigma_y} \sim \mathcal{N}(0, 1)$. However, in real world scenarios σ_y is typically not known and needs to be estimated. Since incorrect hypotheses typically contain outliers, it is important with a robust estimate of the noise level. We show these steps in Figure 1. It is clear that h_{bad} produces a tailed residual distribution that does not follow the expected Gaussian curve. Hence we can use the KS test [20] to validate the produced models.

Underfitting and Overfitting It is common for a complex model to be optimized to fit the data y perfectly. We can describe overfitting and underfitting as a

constant multiplication of σ_y^2 , yielding residuals distributed as $\mathcal{N}(0, a\sigma_y^2)$. When $a < 1$ the model is overfitted, and when $a > 1$ it is incorrect (underfitting). The following sections describe how we can apply this intuition to validate a camera calibration.

3.2 Camera Calibration

Calibration fundamentally depends on correspondences of point coordinates. An arbitrary camera, c , observes a set of K 3D points $\{\mathbf{X}_k\}_{k=1}^K$, and a set of corresponding image points $\{\mathbf{x}_k^c\}_{k=1}^K$. Point sets and correspondences are known $\forall k$, and for each camera. In this work, we consider xPnP based camera calibration using the methods proposed by [21] and [14]. Both extrinsic (rotation and translation) and intrinsic (focal length and distortion) parameters in (3) are computed to enable measurement of length and angles in the camera image.

Distortion Depending on the optical system of a camera, small or large displacements of image coordinates can be introduced, called image distortion. Unlike the focal length, which scales the image uniformly, distortion is characterised as scaling the image differently depending on the distance to a distortion centre. The farther the pixels are from the centre of distortion, the more they are distorted. We let the same point represent the distortion and optical centre, which is assumed to be fixed and in the centre of the image.

$$\mathbf{y}' = g(\mathbf{y}, \boldsymbol{\theta}) \quad (1)$$

We model the distortion as in (1), and let $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$ specify the non-linear distortion terms. When g uses one distortion term, it will be denoted as $\text{D}(1, 0)$ and as $\text{D}(3, 0)$ when all three terms are used, according to [14].

Correspondences The calibration uses Lidar measurements, which map physical features with high precision by emitting narrow laser beams that are reflected back. Even if the image, whose camera we want to calibrate, and the lidar map are recorded at separate times, there should be enough overlapping features left for calibration. That is, consistent physical properties. Such properties, which are more likely to be consistent, are, for example, those found on buildings, vegetation, paintings, furniture, etc. In practice, correspondences can be of varying quality, making robust estimation a critical importance, when computing an xPnP solution [14]. Therefore, we use only the residuals from correspondences marked as inliers by the model estimator for model validation.

3.3 Residual error model

Regardless of whether the corresponding coordinates are found by an interest point detector, or whether they are manually annotated by a human, they will suffer from *detection noise*. This means that a location estimate $\tilde{\mathbf{x}}$ has a residual $\boldsymbol{\epsilon}_{detector}$, compared to the ideal point location $\hat{\mathbf{x}}$. This residual is typically modelled as a 2D normal distribution:

$$\boldsymbol{\epsilon}_{detector} = \tilde{\mathbf{x}} - \hat{\mathbf{x}} \sim \mathcal{N}(\mathbf{0}, \sigma_d^2 \mathbf{I}). \quad (2)$$

For a successful calibration, the residual between a detected point, and the projection of the corresponding 3D point using the estimated parameters, should also satisfy (2). In other words:

$$\hat{\mathbf{x}} = \text{proj}_{\Theta, \mathbf{P}}(\hat{\mathbf{X}}) = \mathbf{K}g(\pi(\mathbf{R}\hat{\mathbf{X}} + \mathbf{t}), \boldsymbol{\theta}). \quad (3)$$

Here $\hat{\mathbf{X}}$ is the ideal 3D point, and π is the pinhole projection. The *intrinsic calibration*, $\Theta = (\mathbf{K}, \boldsymbol{\theta})$, and the *extrinsic calibration*, $\mathbf{P} = (\mathbf{R}, \mathbf{t})$ (the camera pose) are of course estimates in practice. We summarize the error caused by the estimation in an additive *modelling noise* term ϵ_{model} . We intend to explain the residuals by detection noise in the image and in the Lidar, and test whether the explanation holds using a test on the residual data, *e.g.* by the DAP test [6], or by the KS test [20], testing the GoF.

For 2D-3D matches, the 3D points are also affected by noise ϵ_{3D} . Thus, (3) should be replaced by:

$$\hat{\mathbf{x}} - \epsilon_{lidar} - \epsilon_{model} = \text{proj}_{\Theta, \mathbf{P}}(\hat{\mathbf{X}} - \epsilon_{3D}), \quad (4)$$

where ϵ_{3D} is the detection noise in 3D, and ϵ_{lidar} its projection. By combining (4) with (2) we obtain the following residual model:

$$\boldsymbol{\epsilon} = \epsilon_{detector} + \epsilon_{lidar} + \epsilon_{model}. \quad (5)$$

We model the detection error as in (2), and describe the lidar error model in detail below.

Lidar error model For a Lidar sensor, the 3D noise has both angular and depth components. However, when the camera and 3D-sensor are close to being co-axial, and point in roughly the same direction (*i.e.* \mathbf{t} is small, and $\mathbf{R} \approx \mathbf{I}$ in (3)), the depth error becomes irrelevant, and the projection in the image ϵ_{lidar} is dominated by \mathbf{K} , which is affine. This means that the shape of ϵ_{lidar} is a simple, but location dependent scaling.

We thus model the projection of the Lidar error ϵ_{lidar} as:

$$\epsilon_{lidar} \sim \mathcal{N}(0, \sigma_l^2 \text{diag}(a_x^2, a_y^2)), \quad (6)$$

where σ_l^2 is a noise variance, and a_x, a_y are the noise scalings in horizontal and vertical directions. These depend on the location in the image. To estimate a_x, a_y , we can project the current 3D point and its neighbours in pan and tilt directions to obtain:

$$a_{x,k} = \|\text{proj}(\mathbf{X}_k) - \text{proj}(\mathbf{X}_k^P)\| \quad (7)$$

$$a_{y,k} = \|\text{proj}(\mathbf{X}_k) - \text{proj}(\mathbf{X}_k^T)\|, \quad (8)$$

where \mathbf{X}_k is the current 3D point, and \mathbf{X}_k^P , and \mathbf{X}_k^T are its neighbours in pan and tilt directions.

3.4 Noise Estimation

The parameters of the detector errors in the model (5) can be fitted to the observed residuals using robust linear regression. However, estimating the variance of ϵ_{model} is neglected since its observed values are those that will remain in order to test whether the model is incorrect. I.e. we assume:

$$E \{ \epsilon^2 \} = \sigma_d^2 + \sigma_l^2. \quad (9)$$

By using a common σ_d for x, and y image residuals, and the aspect ratio model in (6) we obtain:

$$E \left\{ \begin{pmatrix} \epsilon_x^2 \\ \epsilon_y^2 \end{pmatrix} \right\} = \begin{pmatrix} 1 & a_x^2 \\ 1 & a_y^2 \end{pmatrix} \begin{pmatrix} \sigma_d^2 \\ \sigma_l^2 \end{pmatrix}. \quad (10)$$

We fit these to the observed K residuals, for each camera, c , separately, to obtain the regressor parameters (σ_d, σ_l) .

$$\begin{pmatrix} \epsilon_{x,1}^2 & \epsilon_{y,1}^2 & \dots & \epsilon_{x,K}^2 & \epsilon_{y,K}^2 \end{pmatrix}^\top = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ a_{x,1}^2 & a_{y,1}^2 & \dots & a_{x,K}^2 & a_{y,K}^2 \end{pmatrix}^\top \begin{pmatrix} \sigma_d^2 \\ \sigma_l^2 \end{pmatrix} \quad (11)$$

In practice, we do not use linear regression by solving the normal equations to (11), but use robust regression using IRLS [11] with initial weights $1/p(\epsilon_k|\sigma = 5.0)$. We can now obtain standardised residuals:

$$\tilde{\epsilon}_k = \begin{pmatrix} \epsilon_{x,k}/\sigma_{k,x} \\ \epsilon_{y,k}/\sigma_{k,y} \end{pmatrix} = \begin{pmatrix} \epsilon_{x,k}/\sqrt{\sigma_d^2 + a_{x,k}^2\sigma_l^2} \\ \epsilon_{y,k}/\sqrt{\sigma_d^2 + a_{y,k}^2\sigma_l^2} \end{pmatrix}. \quad (12)$$

3.5 Hypothesis Testing

When the modelling error is low, the standardised residuals in (12) should pass a statistical test, such as, *e.g.*, the KS test. We can thus use the test to check whether the calibration worked for a particular set of 2D-3D correspondences. More formally, we test the **null hypothesis** \mathcal{H}_0 : *The standardised residuals (12) are distributed explicitly according to a standard normal*, against \mathcal{H}_1 : *at least one value does not match that distribution*. Related to this classical approach is that the data we are testing is random, so the test decision is random too, which means there is still a small probability of an incorrect decision. Nevertheless, tests are useful to detect low model errors and thus further validate the calibration.

Evidence The approach involves comparing the samples (residuals) with a statistical model under \mathcal{H}_0 , where a test statistic measures the discrepancy between the data and the model. To this end, we use the KS test [20] and compute a *p-value*, measuring the error size of rejecting \mathcal{H}_0 . Commonly, when the *p-value* is below 5%, \mathcal{H}_0 can be rejected in favour of \mathcal{H}_1 . That is, the error probability is sufficiently low. However, this probability does not directly infer confidence for the data distributed as a standard normal.

Other tests also calculate a *p-value* to test the normality of data. For example, the DAP test [6] sums the discrepancies from a skewness test and a kurtosis test

into a single p -value. Skewness is the asymmetry about the mean, and kurtosis is the measure of the "tailedness". Although parametric tests are preferable to non-parametric ones, and the SW test is one of the more powerful [10], we believe their null hypotheses to be non-directional, where a broader chance of normality is possible, leading to unstable decisions.

4 Experiments

We first evaluate the proposed method for testing a calibration using 2D-3D correspondences on synthetic data simulating detection and Lidar errors. Next, we provide results on a real scenario using Lidar measurements and compare this with a semi-automatic calibration. Last, we analyse a large-scale dataset using our method.

4.1 Synthetic data

We implemented the simulator in OpenCV [1] and will provide code upon publication. We aim to render a fictitious checkerboard pattern with equidistant squares of black and white into a camera c with a small angular rotation maintaining the image centre as its viewpoint at a distance t . The pattern contains equally many saddle points (inner checkerboard corners) vertically as horizontally (15×15). The simulator iterates three main tasks to render each image, which is presented next. **(i)** The projection model has fixed intrinsic parameters according to $D(3,0)$ [14]. That is, $f = 800$ and $\theta = [-0.0684, 0.0100, 0.0006]$. We let the distortion centre coincide with the image centre. Rotation parameters are randomly sampled in the range $\pm 15^\circ$ relative to the z-axis. The translation can also be random, but in our generated synthetic dataset, we move the pattern closer and closer to the camera. **(ii)** Next, we smooth the image (to avoid aliasing), add image noise and interpolate it to size 1600×1600 . A saddle point detector [1] locates the 2D position of these features with sub-pixel precision.

Model	f_x	f_y	c_x	c_y	k_1	k_2	k_3	$\text{std}(\text{axis223m})$	$\text{std}(\text{axisp3364})$
M ₁	✓				✓			2.99	2.91
M ₂	✓		✓	✓	✓			2.96	2.73
M ₃	✓				✓	✓	✓	3.03	2.95
M ₄	✓		✓	✓	✓	✓	✓	3.19	2.52
M ₅	✓	✓	✓	✓	✓	✓	✓	0.81	0.19

Table 1. Projection models used with the Zhang method [30]. Models M₁ and M₃ use the same number of parameters as the PnP methods $D(1,0)$ [14], $D(3,0)$ [14] and $D(3,0)$ [21]. The standard deviation on a set of test images, determines how accurately the two cameras have been calibrated, *axis223m* and *axisp3364*. M₅ is the most accurate model. Models M₁-M₄ (approximately) share the same standard deviation, although models M₄ and M₅ only differ with one parameter. Our method instead decides M₁ as incorrect but M₃ as a plausible model, still usable for metrology.

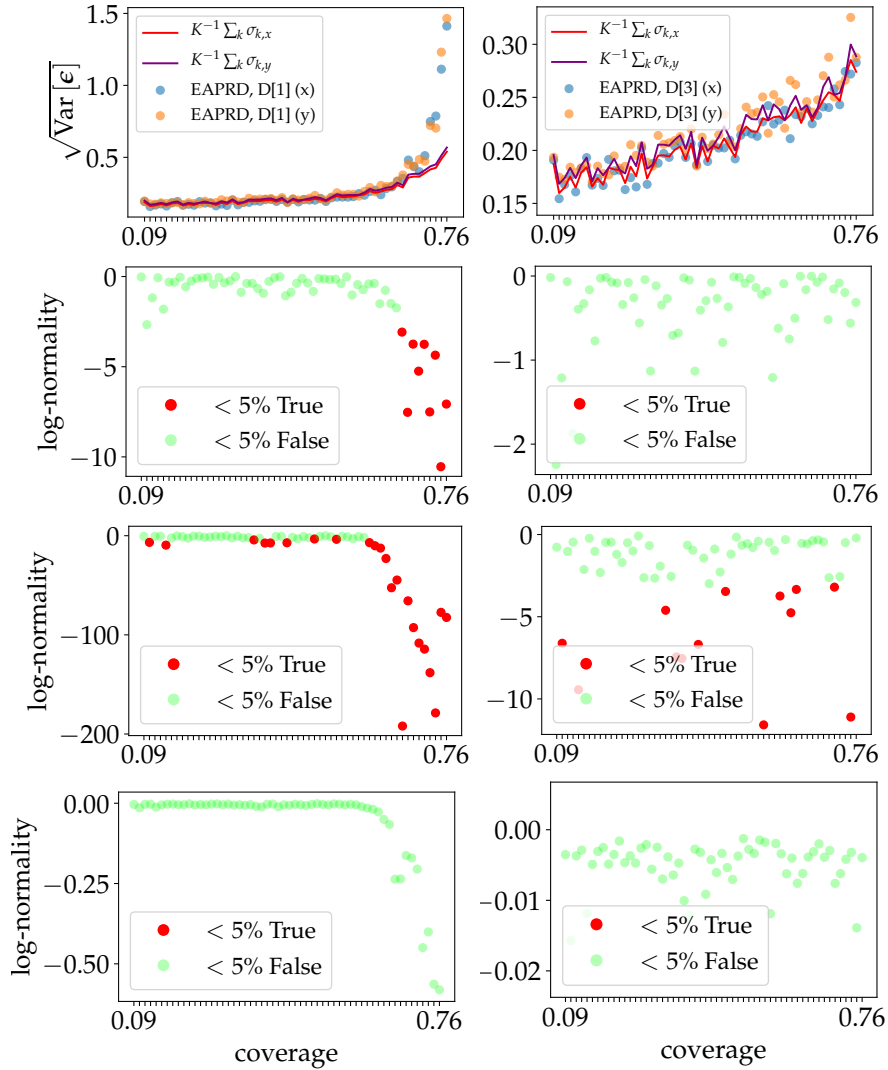


Fig. 2. Using residuals computed from 56 sets of simulated 2D-2D correspondences, we show the results for the incorrect $D(1, 0)$ model (left) and the correct $D(3, 0)$ model (right). We sort the correspondence sets in ascending order, using the area of the 2D points' convex hull (*coverage*). The second, third and fourth rows show the outcome of the KS [20], DAP [6] and SW [26] tests at level 5%. When the model is correct, the standard deviation is low (first row), and our predicted variance follows the corresponding empirical value. The KS test rejects images under an incorrect model, while accepting images under the correct model. In contrast, [6] is too strict, while [26] is too permissive.

We observed the position error to lie within a small range of 0.03 pixels. This corresponds to σ_d in (9).**(iii)** To simulate the Lidar, we add noise on the corresponding 3D points in all images. We transform the noise such that it lives on the sphere with origin \mathbf{t}^c and radius $\|\mathbf{t}^c\|$. On the sphere, the noise magnitude is dependent on \mathbf{t}^c , and in the vertical direction, the noise is always 90% lower compared to the horizontal direction. This reweighing aims to simulate the resolution in the Lidar array, and it replaces (7) and (8), which are used on real datasets. This gives us the resolution aspect, which is used as *explanatory variables* in the estimate, (11).

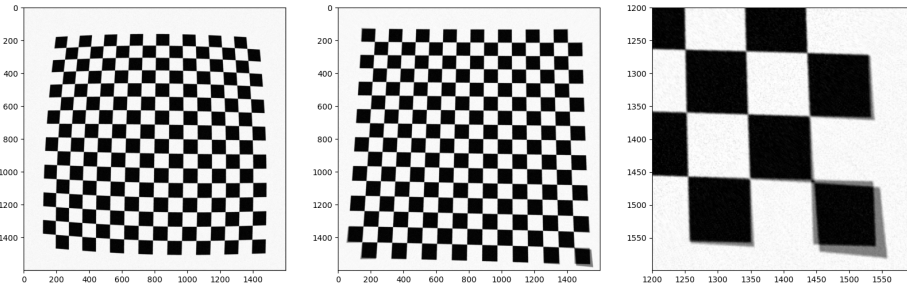


Fig. 3. Incorrect models can have small residual errors, but *Godness-of-Fit* testing exposes them. Left: A distorted image from which the model is estimated. Middle: Undistorted image using true model (black) and wrong model (grey) overlaid. Even though the estimated model is incorrect, when detector errors are small the residual errors are often small, hence simply checking the standard deviation of the residuals is insufficient. Right: Outliers indicate model failure. Given known outlier-free correspondences, deviations from the expected noise distribution expose incorrect models.

Results In Figure 2, we show the output of our method on models $D(1,0)$ [14] and $D(3,0)$ [14] for a set of synthetic 2D-3D correspondences, generated under the $D(3,0)$ model. These are sorted according to the increasing spatial spread of correspondences. The closer the points are to the image edge, the more they are affected by the distortion (Figure 3). The first row in Figure 2 shows the empirical and estimated standard deviations over all points per image, and rows 2-4 whether the scaled residuals are sufficient evidence to reject the null hypothesis for three different tests. Given in the second row of Figure 2 is the output of the KS test for both $D(1,0)$ [14] and $D(3,0)$ [14]. To the left, in the same row, our method rejects the images with correspondences more uniformly spread over the entire image at level 5% using $D(1,0)$ [14]. Making the same test using a projection model with more parameters fits the data more accurately, as indicated by both the low standard deviation and the test. We also test the scaled residuals in the third and fourth rows using [6], and [26], respectively. While these tests are

PnP Method	DT	$\sqrt{\text{Var}[\epsilon]}$	KS[20]	DAP[6]	SW[26]
EAPRD [14]	D(1,0)	2.23	✓	✓	
EAPRD [14]	D(3,0)	0.98		✓	
PNPRF [21]	D(3,0)	1.05		✓	

Table 2. The results for the *axis223m* camera. For each PnP method, with distortion type DT, we report the empirical standard deviation and whether \mathcal{H}_0 can be rejected at level 5% using the KS, DAP or SW test.

PnP Method	DT	$\sqrt{\text{Var}[\epsilon]}$	KS[20]	DAP[6]	SW[26]
EAPRD [14]	D(1,0)	1.71	✓	✓	
EAPRD [14]	D(3,0)	1.47			
PNPRF [21]	D(3,0)	1.08			

Table 3. The results for the *axisp3364* camera. For each PnP method, with distortion type DT, we report the empirical standard deviation and whether \mathcal{H}_0 can be rejected at level 5% using the KS, DAP or SW test.

parametric, they test for any normal distribution. Compared to the proposed KS test, this leads to false positives and negatives, see rows 3-4 of Figure 2. In the second column, the stronger D(3,0) projection model is tested. In most images, \mathcal{H}_0 can not be rejected as expected due to the simulated data conforming to D(3,0) [14]. However, the output does not reveal the tests’ differences in this case. Thus, [20] generally leads to more accurate decisions.

4.2 Lidar Measurements

Next, we compare our method using images from two real cameras, *axis223m* and *axisp3364*, respectively, and a 3D point cloud from a *Leica RTC360* scanner, with semi-automatic calibration. The second camera offers lower-quality images than the first, which is visible in Figure 4. There is also no verified annotation for the cameras; in practice, there is none, and the camera can be inaccessible. The cameras instead depict a scene such that their optical axes have a relatively small angle to the 3D point cloud coordinate system’s z-axis, similar to the simulations in Section 4.1.

Semi-automatic For semi-automatic calibration, we collect images for both cameras using a checkerboard pattern. The pattern has 6×6 saddle points. We split the set of images and estimate the model parameters on the first set using the Zhang method [30] implemented in [1]. Then, we calibrate using different projection models where M_5 , in Table 1, achieves the lowest standard deviation on the second set of images (test). A factor of almost 4 differs between the accuracy of models M_4 and M_5 . In Table 1, the number of model parameters differs by one.

Without Camera Access We show in Tables 2 and 3 that D(1,0) [14] is incorrect compared to D(3,0) [14] and D(3,0) [21] using the KS test on 24 manually annotated correspondences. Each of the annotated 3D points is visible



Fig. 4. The first column shows two images of two cameras, *axis223m* and *axisp3364*, respectively. In the second and third columns, the undistortion looks to be visually removed for both $D(1,0)$ [14], and $D(3,0)$ [14]. Our method correctly detects $D(1,0)$ [14] as incorrect for both cameras (Tables 2 and 3). The undistorted images in the fourth column are visually similar to their original, but this is not detected. For more details, see Section 4.2.

in both cameras and projected to consistent features, *e.g.* corners. Similar to simulation, we observe that the parametric tests contradict each other and are thus infeasible for our application. While models M_1 to M_4 , in Table 1, obtain higher standard deviations using [30], models M_1 and M_3 are equivalent to the models used from [14] and [21], and thus there is possibility that M_5 is overfitted.

Finally, we found that the computed distortion parameters of $D(3,0)$ [21] were all zero, shown in the rightmost column of Figure 4. To our knowledge, [21] divides the xPnP problem into subproblems. In the subproblem that solves distortion, we can't find a condition on θ preventing the *normal equations* from giving the trivial solution. Thus, our method can not make the correct decision to either reject \mathcal{H}_0 or not based on residuals from $D(3,0)$ [21].

4.3 Structure-from-Motion

In this experiment, we use our proposed method on annotations computed from a Structure-from-Motion (SfM) pipeline to get a broader insight into its effectiveness. To this end, we use 1000 images from each scene of MegaDepth [18]. This dataset contains many scenes with 2D-3D correspondences and camera intrinsic and extrinsic parameters given. The SfM pipeline, COLMAP [24][25], estimates the annotation parameters of the widely used benchmark for state-of-the-art comparison. In the dataset, the assumed projection model, a *simple radial*, models a single focal length, one distortion parameter, the distortion centre, rotation and translation. The histogram to the left in Figure 5 shows that residuals are overall low. However, in 70 out of 100 images, our method rejects the null hypothesis at level 5%. The two images in the middle and to the right, in Figure 5, show when \mathcal{H}_0 can not be rejected at level 5% and when \mathcal{H}_0 is rejected in favour of \mathcal{H}_1 . We can thus assume mostly overfitted projection models in [18].



Fig. 5. Left: Density plot of residuals from 1000 images in all scenes, on which the annotation in MegaDepth depends. It is unlikely residuals will be high for images in [18] measuring a good performance. However, our proposed method tests each image and rejects the null hypothesis, \mathcal{H}_0 , on 70 out of 100 images. Middle: Example of when \mathcal{H}_0 can not be rejected, and the *simple radial* projection model is suitable. Right: Example of when our method rejects \mathcal{H}_0 . As can be seen, *e.g.* on the flagpole to the right, the images are distorted.

5 Conclusion

We suggested that metrology applications in forensic analysis use xPnP methods and use our proposed method to validate the calibration without camera access. The method formulation processes a single image, estimating a robust scaling of each correspondence and tests if the scaled set of residuals is drawn from a standard normal distribution. We demonstrate via qualitative and quantitative experiments that the KS test is most suitable and provide further insight from an extensive collection of annotated cameras.

Although we are sufficiently confident that the test can determine models as incorrect with a small margin of error, the challenge remains to infer confidence in the image measurements. A test is not a classification, and the *p-value* does not imply measurement confidence. However, when rejection of the null hypothesis is not possible at the acceptable error level, our error model explicitly provides the expected measurement errors over the image. Depending on the number of correspondences, we can get local estimates of expected measurement error from our assumptions of normally distributed residuals. Therefore, our method is a useful tool for xPnP camera calibration.

Acknowledgement

This work was partially supported by the Wallenberg AI, Autonomous Systems, and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation; and the computations were enabled by the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre; and a point cloud of a realistic scene was provided by the Swedish National Forensic Centre (NFC).

References

1. Bradski, G.: The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000)
2. Bramble, S., Compton, D., Klasén, L.: Forensic image analysis. In: *Proceedings of the 13th INTERPOL Forensic Science Symposium* (2001)
3. Brandner, M.: Bayesian uncertainty evaluation in vision-based metrology. In: Gallegos-Funes, F. (ed.) *Vision Sensors and Edge Detection*, chap. 5. IntechOpen, Rijeka (2010). <https://doi.org/10.5772/10135>, <https://doi.org/10.5772/10135>
4. Criminisi, A.: Single-view metrology: Algorithms and applications. In: *Pattern Recognition, 24th DAGM Symposium*. Zurich, Switzerland (January 2002)
5. Criminisi, A., Reid, I., Zisserman, A.: Single view metrology. *International Journal of Computer Vision* **40**(2), 123–148 (2000)
6. D'Agostino, R., Pearson, E.S.: Tests for departure from normality. *Biometrika* **60**, 613–622 (1973)
7. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**(6), 381–395 (jun 1981)
8. Förstner, W.: Uncertainty and projective geometry. In: *Handbook of Geometric Computing*, pp. 493–534. Springer (2005)
9. Gao, X.S., Hou, X.R., Tang, J., Cheng, H.F.: Complete solution classification for the perspective-three-point problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Volume: 25, Issue: 8, Aug. 2003) (2003)
10. Ghasemi, A., Zahediasl, S.: Normality tests for statistical analysis: a guide for non-statisticians. *Int. J. Endocrinol. Metab.* **10**(2), 486–489 (April 2012)
11. Green, P.J.: Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)* **46**(2), 149–192 (1984)
12. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press (2003)
13. Heuel, S.: Uncertain projective geometry: statistical reasoning for polyhedral object reconstruction, vol. 3008. Springer (2004)
14. Larsson, V., et al.: Revisiting radial distortion absolute pose. *International Conference on Computer Vision (ICCV)* (2019)
15. Lebeda, K., Matas, J., Chum, O.: Fixing the locally optimized ransac. *British Machine Vision Conference (BMVC)* (2012)
16. Lepetit, V., Moreno-Noguer, F., Fua, P.: Epnnp: An accurate $o(n)$ solution to the pnp problem. *International Journal Of Computer Vision* **81**, 155–166 (2009)
17. Li, X., Zhang, B., Sander, P.V., Liao, J.: Blind geometric distortion correction on images through deep learning. *Conference on Computer Vision and Pattern Recognition (CVPR)* (2019)
18. Li, Z., Snavely, N.: Megadepth: Learning single-view depth prediction from internet photos (2018)
19. Lu, X.X.: A review of solutions for perspective-n-point problem in camera pose estimation. *Journal of Physics: Conf. Ser.* 1087 052009 (2018)
20. Massey, F.J.: The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association* **46**(253), 68–78 (1951)
21. Nakano, G.: A versatile approach for solving pnp, pnpf and pnpfr problems. *European Conference on Computer Vision (ECCV)* (2016)
22. Olsson, E.: *Lens Distortion Correction Without Camera Access*. Master's thesis, Linköping University, Sweden (2022)

23. Persson, M., Nordberg, K.: Lambda twist: An accurate fast robust perspective three point (p3p) solver. In: Proceedings of the European Conference on Computer Vision (ECCV) (September 2018)
24. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
25. Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.M.: Pixelwise view selection for unstructured multi-view stereo. In: European Conference on Computer Vision (ECCV) (2016)
26. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples)†. *Biometrika* **52**(3-4), 591–611 (1965). <https://doi.org/10.1093/biomet/52.3-4.591>
27. Thai, T.H., Cogranne, R., Retraint, F.: Camera model identification based on the heteroscedastic noise model. *IEEE Transactions on Image Processing* **23**(1), 250–263 (2014)
28. Tsai, R.Y.: A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE* (1987)
29. Wang, Q., Alexander, W., Pegg, J., Qu, H., Chen, M.: Hypoml: Visual analysis for hypothesis-based evaluation of machine learning models. *IEEE Transactions on Visualization and Computer Graphics* **27**(2), 1417–1426 (2021)
30. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(11), 1330–1334 (2000)