Detection of Camera Model Inconsistency and the Existence of Optical Image Stabilization System

Shu-Hao Yeh, Di Wang, Wei Yan, and Dezhen Song

Abstract— Cameras becomes more ubiquitous on mobile devices which pave the way for Augment Reality (AR) applications. AR's enabling technology is the underlying visual-inertial Simultaneous Localization and Mapping (SLAM) package which requires a precise camera model for mapping purpose. Due to manufacturing inconsistency and device aging over time, the preciseness is often hard to maintain over time. On the other hand, those cameras are often equipped with optical image stabilization (OIS) system. OIS changes camera intrinsic parameters and being aware of its existence is important before a high-order SLAM model is applied. Here we present a two-step approach to detect if an image conforms to a given camera model (distortion coefficient and intrinsic matrix) by developing two statistical hypothesis testings. We have implemented the algorithm and test it in physical experiments. Results show that our algorithm successfully detects model inconsistency and the existence of OIS system with 85.4% recall and 100% precision.

I. INTRODUCTION

Modern mobile phones and tablets are often equipped with on-board cameras that support Augmented Reality (AR) applications. AR Application Programming Interfaces (APIs) such as ARKit™ from Apple™ or ARCore™ from Google™ are essentially visual-inertial Simultaneous Localization and Mapping (SLAM) packages. All visual SLAM algorithms require precise camera model to enable accurate environment mapping. However, this is often compromised because 1) camera parameters may change over time due to device aging and 2) optical image stabilization system changes intrinsic parameters when activated. To deal with the issue, we need the ability to detect camera model inconsistency and the existence of OIS before any high-order SLAM model is used.

OIS is a mechanism to improve image quality by moving optical lens to compensate for the image blurring caused by hand jittering motion. Since OIS shifts and/or rotates camera lens to reduce image blurring (Fig. 1), it consequently changes camera projection model. Due to its ubiquitous existence, OIS's impact on the quality of mapping process cannot be ignored. However, most devices ignore its existence by using a low resolution image when applying AR APIs which limits the AR mapping accuracy.

The purpose of this work is to pave the way to the development of an adaptive camera model for OIS enabled

S. Yeh, D. Wang, and D. Song are with Computer Science and Engineering Department, Texas A&M University, College Station, TX 77843, USA. Emails: *ericex1015@tamu.edu*, *ivanwang@tamu.edu* , and *dzsong@cs.tamu.edu*.

W. Yan is with Architecture Department, Texas A&M University, College Station, TX 77843, USA. Email: *wyan@tamu.edu*.
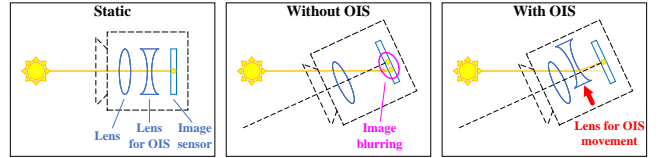
Fig. 1. Illustration of OIS working principle. Left: the scenario without camera movement. Center: The scenario under camera movement, but without lens for OIS resulting in the image blurring. Right: The OIS compensates the camera movement.

devices and also to provide understandings for camera model accuracy for future high resolution AR applications. Here we present a 2-step algorithm to detect model consistency and the existence of OIS system. Using checkerboard grid points as inputs, we compose a statistical hypothesis testing to verify if collinearity property is preserved for collinear grid points under the existing camera distortion removal process. Then we formulate a second statistical hypothesis testing to check if the camera intrinsic matrix matches the image points. We have implemented our algorithm and tested it in physical experiments. Results show that our algorithm successfully detects model inconsistency and the existence of OIS system with 85.4% recall and 100% precision.

II. RELATED WORK

To understand how OIS impacts the imaging process, let us review projective geometry and recent OIS development.

Projective geometry in computer vision describes the relation between 3D points in world coordinate system and the imaged 2D points, and is the foundation of vision-based robotic localization [1]–[3] and 3D reconstruction [4]–[6], etc. Camera perspective projection model [4], [7], also known as Pinhole model, is the most widely adopted geometric model since it employs a robust linear model to describe the perspective mapping from 3D points to 2D image points. The model parameters can be easily interpreted by using the camera intrinsic (e.g. focal length, principle points, etc.) and extrinsic parameters.

The projective geometric models in computer vision often assume no nonlinear distortion in the image. Namely, points on a straight line must be imaged as points on a straight line [4]. This conservation of collinearity is the underlying assumption for projective geometric models. The collinearity recovery usually can be achieved through the camera calibration [8]–[11] by removing lens distortion. Wang et al. [12] classify the lens distortion into radial distortion, decentering (tangential) distortion, and thin prism

distortion. The changing of camera parameters over time or the actuation of OIS may destroy the property which results in incomplete removal of lens distortion which leads to incorrect application of the geometric model in parameter estimation. Our work can detect if the lens distortion removal is complete to ensure that the underlying assumption in applying geometric model is still valid.

OIS has been widely adapted in cameras especially for smartphones for better image quality under hand jittering. OIS usually contains actuators such as voice coil motor [13], [14], liquid lens [15], or piezoelectric motor [16], etc. OIS compensates for camera movements by altering optical path which is usually done by shifting or rotating camera lens barrel after detecting vibration. The issues of variable camera intrinsic parameters for cameras with OIS mechanism has not gained much attention even though cameras with OIS have become more and more common. As an early work, CIP-VMobile [17] includes the camera intrinsic parameters into the state variables to achieve robust pose estimation. However, CIP-VMobile assumes known servo actuator measurements which may not be reliable in mobile phone cameras. We do not assume known actuator measurements and only rely on images taken by the camera.

III. Problem Definition

For a given camera, its intrinsic matrix K and distortion coefficients (DCs) are often either provided by manufacturers (e.g. intrinsic matrix is often embedded in AR APIs in mobile phones) or obtained through pre-calibration at stationary position to avoid triggering OIS. The camera images are often pre-rectified to remove nonlinear lens distortion based on K and DCs. However, we need to verify these parameters. Placing a checkerboard pattern in a camera field of view with sufficient resolution, we can use checkerboard points to detect if camera model inconsistency in both lens distortion removal and the camera intrinsic parameters.

A. Assumption and Nomenclature

We assume that the position noises of the checkerboard points follow zero-mean Gaussian distribution with known variance σ^2 in each dimension and the noise in each dimension is independent. Common notations in this paper are defined as follows.

K   the prior intrinsic matrix of the camera obtained from manufacturer or the pre-calibration.

I input image with its 2D coordinate system $\{I\}$, and it is a rectified image after the lens distortion removal using the prior intrinsic parameters.

$\{W\}$ 3D world coordinate system. $\{W\}$ is defined on the checkerboard with its origin located at the top-left inner vertex, where its X-axis parallel to the top edge, and its Y-axis parallel to the left edge.

$\{C\}$ 3D camera coordinate system (CCS) for I where its origin is at the camera center, and its X-axis and Y-axis parallel to the horizontal and vertical axes of $\{I\}$, respectively.

$\mathbf{x}$ is a homogeneous 3-vector describing a 2D point position in $\{I\}$, $\mathbf{x} \in \mathbb{P}^2$, 2D projective space.

$\mathbf{X}$ is a 3-vector describing a 3D point position.

$_W^C\mathbf{R}$ rotation matrix from $\{W\}$ to $\{C\}$, $_W^C\mathbf{R} \in \mathrm{SO}(3)$.

$_W^C\mathbf{t}$ translation vector from $\{W\}$ to $\{C\}$, $_W^C\mathbf{t} \in \mathbb{R}^3$.

All 3D coordinate systems are right-handed system. Symbol ~ on a variable means that it is in inhomogeneous coordinate.

B. Camera Model Review

Since our goal is to detect the inconsistency of the camera model [4], let us begin with a brief review of camera model. Camera model describes how the camera images the 3D world. Camera model is mainly composed by (1) perspective projection and (2) lens (nonlinear) distortion. Camera projection is often represented by a homogeneous 3×4 matrix which encapsulates the mapping from the 3D points in the world space to the 2D image plane by using the camera extrinsic and intrinsic parameters. The camera extrinsic parameters are the perspective transformation from the 3D world coordinates $\{W\}$ to the 3D camera coordinates $\{C\}$, and contain the rotation matrix $_W^C\mathbf{R}$ and the translation vector $_W^C\mathbf{t}$. The camera intrinsic parameters are related to 2D image projection from the 3D camera coordinates by the focal lengths f_x and f_y, the principal point (p_x, p_y) and skew s. We usually represent the camera intrinsic parameters by a camera intrinsic matrix $\mathrm{K} = \begin{bmatrix} f_x & s & p_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$. Therefore, the camera projection of a 3D world corresponding point $^W\mathbf{X}$ to its corresponding 2D image point $\mathbf{x}$ can be written as

$$\mathbf{x} = \lambda \left(\mathrm{K} \begin{bmatrix} _W^C\mathbf{R} & _W^C\mathbf{t} \end{bmatrix} \right) \begin{bmatrix} ^W\mathbf{X} \\ 1 \end{bmatrix}, \tag{1}$$

where λ is a scalar.

The perspective project, in fact, is a linear model. Thus, it assumes that the 3D world points and the projected 2D points have a linear relationship. The 3D world lines must be imaged as lines. Unfortunately, a camera in reality often introduces lens (nonlinear) distortion, where the linear assumption cannot hold. To remove nonlinear distortion, we use the lens distortion models to describe the mapping between the distorted points and the undistorted points. Radial distortion model is the most common lens distortion model, and it utilizes a polynomial function to capture the effect that a light ray bends more when it passes through camera lens further away from the optical center. Let (x_d, y_d) and (x_u, y_u) be the distorted and undistorted points under normalized image coordinates, respectively, where the origin of its image coordinate system is translated to the principal point where the optical axis intercepts image plane. The radial distortion between (x_d, y_d) and (x_u, y_u) can be written as

$$x_d = x_u(1 + \kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6 + \cdots)$$
$$y_d = y_u(1 + \kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6 + \cdots), \tag{2}$$

where κ_i is the i-th radial distortion coefficients, and r is the radial distance from the center of the radial distortion. The first, second and third radial distortion coefficients are

usually predominant. Therefore, the other orders are usually negligible.

C. Inputs

The first input is the prior camera model obtained from either pre-calibration or its manufacturer. It includes K and DCs as lens distortion model. Input image I has been rectified by using K and DCs.

We then preprocess the rectified image I to obtain the 2D and 3D checkerboard inner vertex correspondences (i.e. checkerboard point correspondences). Denote the i-th 3D checkerboard vertex as ${}^W\mathbf{X}_i \in \mathbb{R}^3$, where $\left({}^W\mathbf{X}_i\right)_3 = 0$ since the checkerboard vertices are residing on the X-Y plane. $\left({}^W\mathbf{X}_i\right)_j$ is denoted as the j-th element of ${}^W\mathbf{X}_i$. We perform the checkerboard point detection [18] to obtain the 2D checkerboard points. Denote the i-th 2D checkerboard point as $\mathbf{x}_i$. The 2D and 3D checkerboard point correspondence set is defined as $\{\mathbf{x}_i \leftrightarrow {}^W\mathbf{X}_i : i = 1, \cdots, n\}$, where n is the number of the checkerboard points.

D. Problem Definition

Our problem is defined as,

Definition 1: Given the prior camera intrinsic matrix K and n checkerboard point correspondences $\{\mathbf{x}_i \leftrightarrow {}^W\mathbf{X}_i\}_{i=1}^n$, first determine if the camera model inconsistency in lens distortion removal exists in I. If the lens distortion removal is consistent, then determine if the camera model inconsistency in the camera intrinsic parameters exists in I.
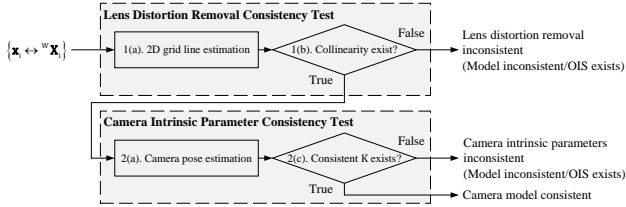
IV. ALGORITHM

Fig. 2.   Detection algorithm flow diagram.

As shown in Fig. 2, our algorithm consists of two main blocks: (1) lens distortion removal consistency test and (2) camera intrinsic parameter consistency test. If the inconsistency of either lens distortion removal or camera intrinsic parameters exists in the rectified image I, we then determine that OIS has been actuated during the image capturing since the OIS is the only factor able to change the geometric relationship between the lens and the image sensors when altering the optical path.

A. Lens Distortion Removal Consistency Test

First, we want to verify if the lens distortion is removed completely from the rectified image I. It is crucial since projective geometry is defined by assuming the conservancy of collinearity in projection. If so, it also means the image has no lens distortion. The projective geometric model estimation can be corrupted and failed if the lens distortion removal is

inconsistent in I. We want to check if lens distortion removal is complete in the rectified image I by using the collinearity test.

1) 2D Grid Line Estimation: We use the lines from the 2D grid which is formed by the checkerboard points to check the collinearity. Let $\mathcal{I}$ be the point index set, where $\mathcal{I} := \{1, \cdots, n\}$. The point index set of the j-th line in the 2D grid is defined as $\mathcal{I}_j := \{k : k \in \mathcal{I}\}$. The point index set of lines satisfies $\bigcup_j |\mathcal{I}_j| = 2n$, where $|\cdot|$ is the set cardinality. Denote the j-th estimated line of the 2D grid as a 3-vector $\mathbf{l}_j \in \mathbb{R}^3$ with unit norm $\|\mathbf{l}_j\|_2^2 = 1$. $\|\cdot\|_2$ is the L2 norm. The estimated line $\mathbf{l}_j$ can be calculated by

$$\min_{\mathbf{l}_j} \sum_{k \in \mathcal{I}_j} d_\perp(\mathbf{x}_k, \mathbf{l}_j)^2, \qquad (3)$$

where $d_\perp(\mathbf{x}_k, \mathbf{l}_j)$ is the perpendicular distance function between the point $\mathbf{x}_k$ and the line $\mathbf{l}_j$.

2) Collinearity Test: The collinearity of points on the lines from the 2D grid is measured by the average distance between the points to the estimated lines. When the collinearity is preserved in the image, the average distance between the points and the estimated lines must be small. However, if the lens distortion is not removed completely from the image, the lines are bent, which results in the enlarging distance between the points and the estimated lines. Define the distance set between the points and the estimated lines of the 2D grid as

$$\mathcal{D} := \{d_{k,j} = d_\perp(\mathbf{x}_k, \mathbf{l}_j) : \forall \mathbf{l}_j \text{ and } \forall k \in \mathcal{I}_j\}. \qquad (4)$$

Given the distance set $\mathcal{D}$, the average distance and the standard deviation are defined as follow:

$$\mu = \frac{1}{n_d} \sum_{d_{k,j} \in \mathcal{D}} d_{k,j} \text{ and } s = \sqrt{\frac{1}{n_d - 1} \sum_{d_{k,j} \in \mathcal{D}} (d_{k,j} - \mu)^2}, \qquad (5)$$

where $n_d = |\mathcal{D}|$ is the number of distance in $\mathcal{D}$.

To detect the collinearity, we design the following hypothesis testing based on the Z-test:

$$\begin{aligned} \mathbf{H_0} &: \quad \mu > \mu_d, \\ \mathbf{H_1} &: \quad \text{Otherwise.} \end{aligned} \qquad (6)$$

μ_d is the distance threshold determined by the experiments. The test statistic can be calculated by

$$Z = \frac{\mu - \mu_d}{s/\sqrt{n_d}}. \qquad (7)$$

Define $\Phi(x)$ as the cumulative distribution function of the standard normal distribution at value x. By setting the significance level α, the p-value is obtained by $\Phi^{-1}(1 - \alpha)$. We consider that the collinearity exists in the lines of the 2D grid by rejecting $\mathbf{H_0}$ when

$$Z \leq \Phi^{-1}(1 - \alpha). \qquad (8)$$

Otherwise, we consider that the lens distortion removal is not complete in the rectified image I.

B. Camera Intrinsic Parameters Consistency Test

Next, we want to examine if the prior camera intrinsic parameters K is consistent with the rectified image I. We begin with the camera pose estimation.

1) Camera Pose Estimation: Given the 2D and 3D checkerboard point correspondences $\{\mathbf{x}_i \leftrightarrow {}^W\mathbf{X}_i\}$, we estimate the initial camera pose ${}^C_W\mathbf{R}$ and ${}^C_W\mathbf{t}$ by using using PnP method [19]. We then employ the MLE [4] to refine the camera pose ${}^C_W\mathbf{R}$ and ${}^C_W\mathbf{t}$. Define the measurement vector as

$$\mathbf{e} := \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} \in \mathbb{R}^{2n} \quad \text{and} \quad \mathbf{e}_i = \widetilde{(\hat{\mathbf{x}}_i)} - \tilde{\mathbf{x}}_i \in \mathbb{R}^2, \quad (9)$$

where $\mathbf{e}_i$ is the reprojection error vector, $\hat{\mathbf{x}}_i$ is the reprojected point using (1) and $\widetilde{[x, y, z]}^\mathsf{T} = [x/z, y/z]^\mathsf{T}$ is vector dehomogenizing. The MLE solves camera pose ${}^C_W\mathbf{R}$ and ${}^C_W\mathbf{t}$ by minimizing

$$\min_{{}^C_W\mathbf{R}, {}^C_W\mathbf{t}} \mathbf{e}^\mathsf{T}\Sigma_e^{-1}\mathbf{e}. \quad (10)$$

$\Sigma_e \in \mathbb{R}^{2n \times 2n}$ is the covariance matrix of $\mathbf{e}$, where $\Sigma_e = \text{Diag}(\cdots, \Sigma_x, \cdots)$. $\Sigma_x \in \mathbb{R}^{2 \times 2}$ is the covariance matrix of $\tilde{\mathbf{x}}_i$, and is determined by the experiments.

2) Camera Intrinsic Parameters Consistency Test: We use the reprojection error of the checkerboard point to measure if K is consistent with I. We model the reprojection error vector $\mathbf{e}_i \in \mathbb{R}^2$ in (9) as the zero-mean Gaussian distribution with the covariance Σ_x. For each checkerboard point and its reprojected point $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$, we design the following hypothesis testing:

$$\begin{aligned} \mathbf{H_0}: & \quad \mathbf{x}_i \text{ and } \hat{\mathbf{x}}_i \text{ do not fit K}, \\ \mathbf{H_1}: & \quad \text{Otherwise.} \end{aligned} \quad (11)$$

The test statistic can be calculated by

$$\mathbf{e}_i^\mathsf{T}\Sigma_x^{-1}\mathbf{e}_i \quad (12)$$

Since we approximate $\mathbf{e}_i$ as the normal distribution with zero mean vector and the covariance Σ_x, (12) is a χ^2 distribution with 2 DoFs. We use the distance threshold $F_2^{-1}(1-\alpha)$ by setting the significance level α to determine if $\mathbf{H}_0$ is rejected. We consider $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ agree with the prior camera intrinsic matrix K by rejecting $\mathbf{H}_0$ when

$$\mathbf{e}_i^\mathsf{T}\Sigma_x^{-1}\mathbf{e}_i \le F_2^{-1}(1-\alpha). \quad (13)$$

We evaluate the camera intrinsic parameters consistency by checking the ratio of the checkerboard points and its reprojected points which agrees with the prior camera intrinsic matrix K by using (11). Let $1_k(\mathbf{x}_i, \hat{\mathbf{x}}_i)$ be an indicator function to determine if $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ agree with the K

$$1_k(\mathbf{x}_i, \hat{\mathbf{x}}_i) := \begin{cases} 1, & \text{when} \quad \mathbf{e}_i^\mathsf{T}\Sigma_x^{-1}\mathbf{e}_i \le F_2^{-1}(1-\alpha) \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

The ratio of the checkerboard points and its reprojected points which agree with the prior camera intrinsic matrix

K can be calculated by

$$\frac{1}{n}\sum_{i=1}^n 1_k(\mathbf{x}_i, \hat{\mathbf{x}}_i), \quad (15)$$

We consider the camera intrinsic parameters of I is consistent with the prior camera intrinsic matrix K if

$$\frac{1}{n}\sum_{i=1}^n 1_k(\mathbf{x}_i, \hat{\mathbf{x}}_i) > \gamma_k, \quad (16)$$

where γ_k is a ratio threshold determined by the experiments. Otherwise, we claim that OIS actuated when image I is captured.

V. EXPERIMENTS

We have implemented our system in Matlab. We evaluate the camera model consistency detection accuracy. We classify datasets into two categories: (1) OIS set which represents the image is affected by OIS actuation, and (2) $\overline{\text{OIS}}$ set which represents images without OIS effect, and evaluate the camera pose accuracy between them.
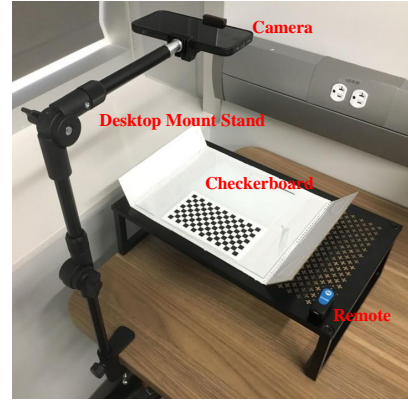

Fig. 3. Experiment setup for $\overline{\text{OIS}}$ image collection.

A. Experiment Setup

We deploy a printed checkerboard on a planar glass surface to ensure flatness (See the checkerboard in Fig. 3). The checkerboard contains 8×16 inner vertices, a sufficient number of checkerboard points for OIS detection. Since most current AR applications are desktop applications or close range indoor applications, our experiment setup also chose the close range working distance. The cell side length of the checkerboard is chosen to be 8 mm to ensure sufficient resolution under the working distance.

In experiments, we use the iPhone 12 Pro camera (see Fig. 3), and we resize the image to a resolution of 1440×1080 pixels to maintain the good image quality since the image quality in the original image resolution 4032×3024 is poor due to low signal-to-noise ratio and aggressive compression in hardware. We calibrate the camera, and collect two datasets: OIS dataset and $\overline{\text{OIS}}$ dataset. We have turned off the autofocus function when collecting both dataset. The OIS dataset is collected by hand-holding the camera with different pose.

Since OIS is a built-in mechanism in the iPhone camera and cannot be turned off beforehand, it is necessary to avoid in triggering OIS when collecting for the $\overline{\text{OIS}}$ dataset. Fig. 3 shows the experiment setup for collecting $\overline{\text{OIS}}$ images. To avoid triggering the OIS, we fix the camera on a rigid desktop mount stand to maintain the same camera pose, and move the checkerboard to collect images under different relative poses. Besides, we use a remote control to capture the images to avoid the hand jittering motion that may trigger OIS during the image capturing. To validate OIS is not triggered under our setup, we perform a camera calibration using the method for collecting $\overline{\text{OIS}}$ dataset. The calibration result shows that the average reprojection error is about 0.13 pixels, which indicates constant camera intrinsic parameters and OIS is not actuated under the setup.

B. Camera Model Consistency Detection Tests

In the tests, we first want to verify if our lens distortion removal consistency test is able to detect the incomplete lens distortion removal from images (Sec. V-B.2). We are also interested in whether our method is able to detect OIS using the two-step camera model consistency tests (Sec. V-B.3).

1) Evaluation Metrics: We use the standard classification results (true positive, true negative, false positive and false negative), and measure the detection accuracy by precision, recall and true negative rate (TNR).

2) Lens Distortion Removal Consistency Tests: We want to verify if our lens distortion removal consistency test is able to detect if lens distortion is not removed completely from images. We collect two image datasets: DC dataset which represents the images with incomplete lens distortion removal by setting $\kappa_1 = 0$ in (2) when we rectify the images since the first order κ_1 dominates the radial distortion coefficient, and $\overline{\text{DC}}$ dataset which represents the images with complete lens distortion removal.

a) Experimental Results: Tab. I(a) shows the confusion matrix of lens distortion removal consistency tests. The precision, recall and TNR all can achieve 100%, which means that our lens distortion removal consistency test can successfully detect if the collinearity of images is maintained. It is expected since the collinearity can be easily violated if there is incomplete lens distortion removal, and the corrupted collinearity can be directly reflected on the distance between the points and the lines.

TABLE I

CONFUSION MATRICES OF CAMERA MODEL CONSISTENCY DETECTION TESTS.

		(a) DC				(b) OIS	
		DC	$\overline{\text{DC}}$			OIS	$\overline{\text{OIS}}$
Class	DC	184	0	Class	OIS	134	23
	$\overline{\text{DC}}$	0	184		$\overline{\text{OIS}}$	0	144

3) OIS Detection Accuracy Tests: We are interested in whether our method is able to detect the OIS using our two-step camera model consistency tests. We test our method on both the OIS dataset and the $\overline{\text{OIS}}$ dataset.

a) Experimental Results: Tab. I(b) shows the confusion matrix of OIS detection tests. There are 144 images in the $\overline{\text{OIS}}$ dataset, and 157 image in the OIS dataset. It is worth noting that 134 images in OIS set (the number of true positive in Tab. I(b)) are all detected by the camera intrinsic parameter consistency test. All the images in both OIS and $\overline{\text{OIS}}$ datasets pass the lens distortion removal consistency test, which is possible that the distortion coefficients still can match the image even with OIS actuating since our camera originally does not contain severe lens distortion. The number of false positive n_{FP} in Tab. I(b) is equal to 0, which indicates that our OIS detection method can always correctly identify the images from the $\overline{\text{OIS}}$ dataset. It is expected since the geometric relationship between the lens and the image sensor does not change without OIS effect. Besides, $n_{\text{FP}} = 0$ in Tab. I(b) results in 100% for both precision and TNR. The result of Precision $= 100\%$ and TNR $= 100\%$ indicate that our OIS detection method only classify the images from the OIS dataset into the OIS set, and it does not misidentify the images from the $\overline{\text{OIS}}$ dataset into the OIS set. Recall in Tab. I(b) can achieve about 85.4%, and it shows that our OIS detection method has been very effective. Besides, 85.4% recall is a not negligible rate, and it means that we are highly likely to face the images with inconsistent camera model when the camera is hand-held which is common in such AR applications. It is worth noting that the number of false negative in Tab. I(b) is not equal to 0, and it, in fact, means that images with OIS actuating still sometimes can be consistent with the prior camera model. This is expected because OIS effect may/may not introduces a noticeable change in camera model.

C. Camera Pose Estimation Tests

We are also interested in whether the camera pose accuracy is different between OIS set and $\overline{\text{OIS}}$ set according to the OIS detection result in Sec. V-B.3. Besides, we want to verify if the inconsistency of camera intrinsic parameters result from OIS. Therefore, we include the camera intrinsic matrix K into the parameter vector of MLE for the images in OIS set

$$\min_{\text{K}, {}^C_W\text{R}, {}^C_W\mathbf{t}} \mathbf{e}^{\mathsf{T}}\Sigma_e^{-1}\mathbf{e}. \tag{17}$$

Let "OIS $+$ K*" represent the set of images from OIS and estimate both camera pose and camera intrinsic matrix using (17).

1) Evaluation Metrics: To measure the camera pose accuracy, we use the standard reprojection error [4] in computer vision e for checkerboard corners using the estimated camera pose (with estimate camera intrinsic matrix for OIS+K* set).

2) Experimental Results: Tab. II shows the experimental results of camera pose accuracy. The average reprojection error of $\overline{\text{OIS}}$ set is about 0.13 pixels, and is close to our calibration accuracy. It is expected since the images from the $\overline{\text{OIS}}$ set are consistent with the prior camera intrinsic matrix K. The average reprojection error of the OIS set,

on the other hand, is about 0.45 pixels. The significant difference of the average reprojection errors between the $\overline{\text{OIS}}$ set and OIS set show that our OIS detection method using the camera model consistency tests has been very effective. In addition to the average reprojection error, the standard deviation of reprojection error for $\overline{\text{OIS}}$ is the minimum with 0.07, which again proves that the $\overline{\text{OIS}}$ set contain the consistent camera model with the calibration. For the $\text{OIS} + \text{K}^*$ set, both the average reprojection error and the standard deviation reduce significantly compared with the OIS set, where the average reprojection error is reduced from 0.45 to 0.23 pixels. It again confirms that the camera intrinsic parameters vary when OIS is actuating. Besides, the large average reprojection error and high standard deviation for OIS set indicate that the camera model inconsistency introduced by OIS is not negligible.

TABLE II

EXPERIMENTAL RESULTS OF CAMERA POSE ESTIMATION TESTS. BEST RESULTS ARE HIGHLIGHTED IN BOLDFACE.

	$\overline{\text{OIS}}$	OIS	$\text{OIS} + \text{K}^*$
Avg(e)	**0.13**	0.45	0.23
Std(e)	**0.07**	0.56	0.35

VI. CONCLUSION AND FUTURE WORK

We reported an algorithm in detecting the camera model consistency and the existence of OIS systems. Combining two statistical hypothesis testings in detecting the camera model (distortion coefficient and intrinsic matrix) inconsistency, we are able to successfully detect the model inconsistency and the existence of OIS as shown in the experimental results.

In the future, we will look deep into the relationship between the camera model and the OIS system, and develop a camera model tracking algorithm to further improve the robustness of AR applications.

ACKNOWLEDGMENT

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 *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. IEEE, 2007, pp. 225–234.

[3] 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.

[4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge Univ Pr, 2003.

[5] 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. IEEE, 2006, pp. 363–370.

[6] Y. Zhou, G. Gallego, H. Rebecq, L. Kneip, H. Li, and D. Scaramuzza, "Semi-dense 3d reconstruction with a stereo event camera," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 235–251.

[7] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 8, pp. 1335–1340, 2006.

[8] F. Devernay and O. Faugeras, "Straight lines have to be straight," *Machine vision and applications*, vol. 13, no. 1, pp. 14–24, 2001.

[9] J. Heikkila and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in *Proceedings of IEEE computer society conference on computer vision and pattern recognition*. IEEE, 1997, pp. 1106–1112.

[10] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[11] J. Wang, F. Shi, J. Zhang, and Y. Liu, "A new calibration model of camera lens distortion," *Pattern recognition*, vol. 41, no. 2, pp. 607–615, 2008.

[12] J. Weng, P. Cohen, M. Herniou *et al.*, "Camera calibration with distortion models and accuracy evaluation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 10, pp. 965–980, 1992.

[13] C.-S. Liu, L. Kuo, and B.-J. Tsai, "New electromagnetic design of miniature af vcm actuator with low cost," *Journal of Mechanics*, vol. 32, no. 4, pp. 421–426, 2016.

[14] C.-L. Hsieh, C.-S. Liu, and C.-C. Cheng, "Design of a 5 degree of freedom–voice coil motor actuator for smartphone camera modules," *Sensors and Actuators A: Physical*, vol. 309, p. 112014, 2020.

[15] J. Lee, Y. Park, and S. K. Chung, "Multifunctional liquid lens for variable focus and aperture," *Sensors and Actuators A: Physical*, vol. 287, pp. 177–184, 2019.

[16] Y. Zou, W. Zhang, F. Tian, F. S. Chau, and G. Zhou, "Miniature tunable alvarez lens driven by piezo actuator," *International Journal of Nanotechnology*, vol. 12, no. 10-12, pp. 818–828, 2015.

[17] L. Jin, H. Zhang, and C. Ye, "Camera intrinsic parameters estimation by visual–inertial odometry for a mobile phone with application to assisted navigation," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 4, pp. 1803–1811, 2020.

[18] A. Geiger, F. Moosmann, Ö. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *2012 IEEE international conference on robotics and automation*. IEEE, 2012, pp. 3936–3943.

[19] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, p. 155, 2009.