Generalized Pupil-Centric Imaging and Analytical Calibration for a Non-frontal Camera

Avinash Kumar and Narendra Ahuja
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana, 61801, IL, USA
{avinash,n-ahuja}@illinois.edu

Abstract

We consider the problem of calibrating a small field of view central perspective non-frontal camera whose lens and sensor planes may not be parallel to each other. This can be due to manufacturing defects or intentional tilting. Thus, as such all cameras can be modeled as being non-frontal with varying degrees. There are two approaches to model non-frontal cameras. The first one based on rotation parameterization of sensor non-frontalness/tilt increases the number of calibration parameters, thus requiring heuristics to initialize a few calibration parameters for the final non-linear optimization step. Additionally, for this parameterization, while it has been shown that pupil-centric imaging model leads to more accurate rotation estimates than a thin-lens imaging model, it has only been developed for a single axis lens-sensor tilt. But, in real cameras we can have arbitrary tilt. The second approach based on decentering distortion modeling is approximate as it can only handle small tilts and cannot explicitly estimate the sensor tilt.

In this paper, we focus on rotation based non-frontal camera calibration and address the aforementioned problems of over-parameterization and inadequacy of existing pupil-centric imaging model. We first derive a generalized pupil-centric imaging model for arbitrary axis lens-sensor tilt. We then derive an analytical solution, in this setting, for a subset of calibration parameters including sensor rotation angles as a function of center of radial distortion (CoD). A radial alignment based constraint is then proposed to computationally estimate CoD leveraging on the proposed analytical solution. Our analytical technique also estimates pupil-centric parameters of entrance pupil location and optical focal length, which have typically been done optically. Given these analytical and computational calibration parameter estimates, we initialize the non-linear calibration optimization for a set of synthetic and real data captured from a non-frontal camera and show reduced pixel re-projection and undistortion errors compared to state of the art techniques in rotation and decentering based approaches to non-frontal camera calibration.

1. Introduction and Previous Work

Camera calibration estimates intrinsic (physical) and extrinsic (pose) parameters of a camera with respect to a known world coordinate system. In a camera, occasionally, due to manufacturing limitations, lens and sensor plane may be slightly tilted with respect to each other. In other cases, a significant amount of tilt may be desired for unconventional imaging, e.g. obtaining perfectly focused image of a plane not perpendicular to the direction of viewing, controlling orientation of camera's depth of field profile (Scheimpflug principle), computing omnifocus image [10] or estimating scene depth using depth from focus [9]. Such a camera with tilted sensor is called as *non-frontal* [9], as opposed to an ideal *frontal* camera whose lens and sensor planes are parallel. Typically all lens-sensor configurations can be considered as non-frontal with frontal being a special case. Conventionally, there are two main approaches to model non-frontal sensor:

(1) Implicit: The "effect" of sensor tilt on an image is modeled as decentering distortion [5] about an *effective axis* normal to the sensor and passing through the camera's center of projection. As sensor and lens planes are not parallel, this axis is different from the *optic axis*, about which only radial distortion exists. It has been analytically shown that decentering modeling is approximate and only holds for small amounts of sensor tilt. In fact, for small tilts, a standalone radial distortion model about optic axis is approximately equivalent to a combination of decentering and radial distortion model about the effective axis [13]. Most conventional calibration methods [8, 14, 15, 17] follow this model. We denote all calibration parameters except decentering parameters in this model as the set $\mathtt{U}$.

(2) Explicit: It is assumed that the lens optic axis coincides with the z-axis of lens coordinate system and the sensor coordinate system has its origin at the intersection of optic axis and sensor plane, defined as the center of radial distortion (CoD) henceforth. The sensor non-frontalness is explicitly modeled by a 3×3 rotation matrix $\mathtt{R}$ [7]. Although, this results in increased number of calibration parameters as $(\mathtt{U}, \mathtt{R})$, it is a physically meaningful model of

lens-sensor configuration and can be used for sensor tilt estimation via camera calibration. Such a tilt estimate can be used for depth estimation [9] and omnifocus imaging [10]. Thus, in this paper, we follow this model for non-frontal camera calibration.

Next, we discuss various aspects of non-frontal calibration which are improved upon in this work for achieving higher accuracy in calibration parameter estimates.

Rotation Model: As the lens is planar and symmetric about optic axis, only two Euler angles corresponding to sensor rotation about x and y axis of the lens coordinate system are sufficient to model R.

Imaging Model: (Prior) If the sensor is physically tilted only about a single lens axis (e.g. y-axis), [3] showed that tilt estimates were more accurate in a pupil-centric imaging model compared to a thin-lens imaging model. They derived equations, in terms of pupil-centric parameter set L to map calibration results obtained from thin-lens model to pupil-centric model and vice versa. Thus, one could simply do thin-lens calibration, apply the equations and obtain pupil-centric parameters. **(Proposed)** But for real cameras with arbitrary sensor tilt, these equations do not generalize directly. Instead, an affine transformation using L is first applied to known checkerboard world points which are then input to a thin-lens calibration framework. The thin-lens calibration results can then be linearly transformed back to obtain pupil-centric estimates. Thus, we first derive generalized pupil centric imaging for arbitrary sensor tilt and obtain a mapping from pupil-centric to a geometrically equivalent thin-lens imaging model.

Calibration Technique: Typical camera calibration consists of 3 steps: (i) Assuming no lens distortion, compute 3×4 perspective projection matrix whose elements encode calibration parameters via a system of non-linear equations [1]. (ii) Analytically solve these equations to obtain initial calibration parameter estimates [6, 11, 15]. (iii) Non-linearly refine these estimates taking lens distortion into account. For implicitly modeled non-frontal sensors [15, 17, 8], decentering parameters are estimated in step (iii), which renders initial estimation of U in steps (i,ii) an easy task. But, in calibration of explicitly modeled non-frontal sensors using generalized pupil-centric imaging, the number of calibration parameters increases to $(\mathtt{U}, \mathtt{R}, \mathtt{L})$, thus leading to two effects. First, the system of non-linear equations in step (i) becomes under-determined. Second, the non-linearity between elements of $(\mathtt{U}, \mathtt{R}, \mathtt{L})$ encoded in the projection matrix of step (i) (Sec. 4) become too complex causing step (ii) to be difficult and non-trivial. **(Proposed)** To handle these, we append a novel pupil-centric constraint to the system of non-linear calibration equations in step (i) and assume that two parameters in $(\mathtt{U}, \mathtt{R}, \mathtt{L})$ corresponding to CoD are known. This makes step (ii) a well-constrained problem. Then we propose an new analytical solution for finding actual calibration parameters in step (ii) (Sec. 5.1). This solution is then used in a computational framework to find an optimal estimate of CoD, such that a separate novel radial alignment based constraint is minimized (Sec. 6). Given the optimal CoD, the analytical solution is then used to reliably initialize the non-linear minimization step (iii). Finally, we also show that calibrated estimates of entrance pupil location and optical focal length of the camera 6 can also be achieved in the analytical framework via the pupil-centric constraint. Conventionally, these have been done optically [3].

Relationship to Gennery [7] To the best of our knowledge, this is the only work that considers explicit non-frontal parameterization. Following are some major differences of our work with theirs. First, they assume that angles of incidence and exitance of a light ray at the entrance and exit pupil are equal, which is incorrect [3], while we incorporate a generalized pupil-centric imaging which models the exact relationship between these rays. Second, they initialize non-linear refinement of U heuristically and set R as identity matrix, which can lead to local optima, while we analytically derive initial estimates leading to increased confidence in attaining a solution hopefully close to global optima. Third, they include decentering distortion as a part of non-linear optimization, which is redundant for explicit non-frontal model [13] and can lead to instable results. Fourth, we focus on cameras with small field of view and thus assume that entrance-pupil is fixed (central) while they assume a general scenario of varying entrance-pupil (non-central) location (typical to fish-eye lenses).

Another explicit sensor tilt estimation technique was proposed in [2] using image defocus, but they do not show how to compute R from their tilt parameterization.

We now list **the major contributions this paper makes**:
(1) We generalize pupil-centric imaging from single-axis to to arbitrarily rotated non-frontal sensors (Sec 3.1,3.2). This leads to a new mapping from pupil-centric to a geometrically equivalent thin-lens imaging (Sec. 3.3). Thus, non-frontal calibration under pupil-centric imaging can be done in a thin-lens framework.
(2) We derive the linear projection equations in terms of $(\mathtt{U}, \mathtt{R}, \mathtt{L})$ for the equivalent thin-lens imaging and the non-linear relationship among calibration parameters encoded by these projection equations (Sec. 4).
(3) We propose an analytical solution to solve $(\mathtt{U}, \mathtt{R})$ and two parameters of L, namely entrance pupil and optical focal length by incorporating a novel pupil-centric constraint and assuming that CoD parameters in U are known (Sec. 5.1).
(4) We develop a new radial alignment like constraint (similar to [14] based on analytical solution to estimate the center of radial distortion (CoD) (Sec. 6). A similar technique but using a different constraint has earlier been proposed by [12] for catadioptric cameras.

2. Overview

We first present the thin-lens (Sec. 3.1) and the generalized pupil-centric (Sec. 3.2) imaging equations for an arbitrarily rotated non-frontal sensor. Based on these equations,

we derive a set of mappings to transform non-frontal pupil-centric imaging model to a geometrically equivalent non-frontal thin-lens imaging model (Sec. 3.3). For the equivalent thin-lens model, we derive non-frontal projection equations relating the corresponding known world (on a checkerboard) and the image points in terms of 14 calibration parameters (Sec. 4). The solution (3×4 projection matrix) to these equations non-linearly encodes 14 camera calibration parameters in 11 equations ($12 - 1$ for scale of 3×4 projection matrix) (Tab. 1). In order to solve this under- determined system, we first propose an additional pupil-centric constraint and assume that CoD is known. Thus, we have 12 non-linear equations in 12 calibration parameters. We then derive an analytical solution to solve for these 12 parameters (Sec. 5). Given this solution and a proposed radial alignment based cost function, we propose a computational method to find optimal CoD (Sec. 6). The analytical solution corresponding to this CoD is then used to initialize the non-linear minimization of calibration parameters (Sec. 7). Finally, reduced re-projection and undistortion error for calibration of a non-frontal camera based on proposed generalized pupil-centric imaging model and analytical initialization, as compared to traditional techniques, is shown on synthetic (Sec. 8) and real (Sec. 9) images.

3. Pupil-centric and Thin-lens Imaging

We derive the generalized non-frontal thin-lens and pupil-centric imaging equations and the mapping between them. The coordinate systems (CS) used in this paper are defined as:

- World ($\mathtt{C_W}$): The known world points lie in this CS.
- Lens ($\mathtt{C_{H_1}}$,$\mathtt{C_{H_2}}$): The CS located at primary ($\mathtt{H_1}$) and secondary ($\mathtt{H_2}$) principal planes (Fig. 2).
- Entrance Pupil ($\mathtt{C_{E_n}}$): The CS parallel to $\mathtt{C_{H_1}}$ but centered at entrance pupil location (Fig. 3).
- Sensor ($\mathtt{C_S}$): The CS located on the sensor with origin at the intersection of optic axis and sensor plane (Fig. 3).
- Image ($\mathtt{C_I}$): The CS of the observed image (Fig. 3).

As pupil-centric model requires location of principal planes as a parameter, we will model thin-lens as a Gaussian thick-lens since both are geometrically equivalent. The notation $\mathtt{X_Y}$ represents point $\mathtt{X}$ in coordinate system $\mathtt{Y} \in \{\mathtt{C_W}, \mathtt{C_{H_1}}, \mathtt{C_{H_2}}, \mathtt{C_{E_n}}, \mathtt{C_S}, \mathtt{C_I}\}$.

3.1. Gaussian Model of Image Formation

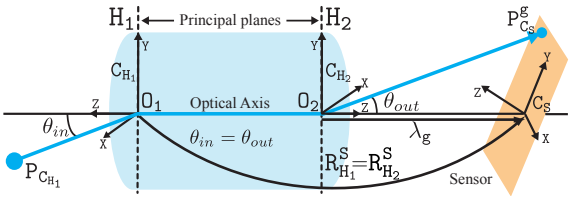

Figure 1. Gaussian (Thin-lens) Model of Image Formation

Consider a world point in $\mathtt{C_{H_1}}$ as $\mathtt{P_{C_{H_1}}} = (\mathtt{x_1}, \mathtt{y_1}, \mathtt{z_1})$. Under the Gaussian thick-lens imaging (Fig. 1), its image $\mathtt{P^g_{C_S}} = (\mathtt{x_g}, \mathtt{y_g})$ is formed on the non-frontal sensor by the light ray from $\mathtt{P_{C_{H_1}}}$ incident on nodal point $\mathtt{O_1}$ and exiting at $\mathtt{O_2}$ at equal angles. If $\mathtt{C_{H_2}}$ and $\mathtt{C_S}$ are related by a rotation $\mathtt{R^S_{H_2}}$ and translation $\mathtt{Tg^S_{H_2}}$ where,

$$\mathtt{R^S_{H_2}} = \begin{bmatrix} \mathtt{r_{11}} & \mathtt{r_{12}} & \mathtt{r_{13}} \\ \mathtt{r_{21}} & \mathtt{r_{22}} & \mathtt{r_{23}} \\ \mathtt{r_{31}} & \mathtt{r_{32}} & \mathtt{r_{33}} \end{bmatrix} \text{ and } \mathtt{tg^S_{H_2}} = \begin{bmatrix} 0 \\ 0 \\ \lambda_g \end{bmatrix}, \quad (1)$$

then, image point $(\mathtt{x_g}, \mathtt{y_g})$ can be obtained as:

$$\mathtt{x_g} = \frac{\lambda_g(-\mathtt{r_{22}x_1} + \mathtt{r_{21}y_1})}{\mathtt{r_{31}x_1} + \mathtt{r_{32}y_1} + \mathtt{r_{33}z_1}}, \quad (2)$$

$$\mathtt{y_g} = \frac{\lambda_g(\mathtt{r_{12}x_1} - \mathtt{r_{11}y_1})}{\mathtt{r_{31}x_1} + \mathtt{r_{32}y_1} + \mathtt{r_{33}z_1}}. \quad (3)$$

Note that because $\mathtt{C_{H_1}}$ and $\mathtt{C_{H_2}}$ are parallel, $\mathtt{R^S_{H_1}} = \mathtt{R^S_{H_2}}$.

3.2. Pupil-centric Model of Image Formation

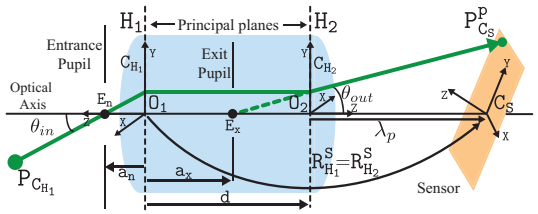

Figure 2. Pupil-centric Model of Image Formation

The pupil-centric model for image formation is shown in Fig. 2, with $\mathtt{P_{C_{H_1}}} = (\mathtt{x_1}, \mathtt{y_1}, \mathtt{z_1})$ imaged on a non-frontal sensor at location $\mathtt{P^p_{C_S}} = (\mathtt{x_p}, \mathtt{y_p})$. Here, the light ray incident at entrance pupil $\mathtt{E_n}$ and exiting at exit pupil $\mathtt{E_x}$ is assumed to be the chief ray responsible for image formation. This model is defined in terms of lens parameters $\mathtt{L} = \{\mathtt{a_n}, \mathtt{a_x}, \mathtt{d}, \mathtt{F}\}$ (Fig. 2), where $\mathtt{a_n}$ and $\mathtt{a_x}$ are distances between the front principal plane ($\mathtt{H_1}$) and the entrance ($\mathtt{E_n}$) and the exit pupil ($\mathtt{E_x}$) respectively, $\mathtt{d}$ is the distance between $\mathtt{H_1}$ and the back principal plane $\mathtt{H_2}$ and $\mathtt{F}$ is the optical focal length in the thin lens equation relating the conjugate points $\mathtt{E_n}$ and $\mathtt{E_x}$:

$$\frac{1}{\mathtt{a_n}(\mathtt{O_1E_n})} + \frac{1}{(\mathtt{d} - \mathtt{a_x})(\mathtt{O_2E_x})} = \frac{1}{\mathtt{F}}. \quad (4)$$

Here, $\mathtt{a_n}, \mathtt{a_x}, \mathtt{d}$ are defined such that positive z-axis of $\mathtt{C_{H_1}}$ is towards incoming light. If $\mathtt{C_S}$ is rotated by $\mathtt{R^S_{H_2}}$ (same as in Sec. 3.1) and translated by $\mathtt{tp^S_{H_2}} = (0, 0, \lambda_p)$ with respect to $\mathtt{C_{H_2}}$, then $(\mathtt{x_p}, \mathtt{y_p})$ can be obtained as:

$$\mathtt{x_p} = \frac{\left(\lambda_p + \frac{\mathtt{a_n}}{\alpha}\right)\left(-\mathtt{r_{22}}(\alpha\mathtt{x_1}) + \mathtt{r_{21}}(\alpha\mathtt{y_1})\right)}{\mathtt{r_{31}}(\alpha\mathtt{x_1}) + \mathtt{r_{32}}(\alpha\mathtt{y_1}) + \mathtt{r_{33}}(\mathtt{z_1} - \mathtt{a_n})}, \quad (5)$$

$$\mathtt{y_p} = \frac{\left(\lambda_p + \frac{\mathtt{a_n}}{\alpha}\right)\left(\mathtt{r_{12}}(\alpha\mathtt{x_1}) - \mathtt{r_{11}}(\alpha\mathtt{y_1})\right)}{\mathtt{r_{31}}(\alpha\mathtt{x_1}) + \mathtt{r_{32}}(\alpha\mathtt{y_1}) + \mathtt{r_{33}}(\mathtt{z_1} - \mathtt{a_n})}, \quad (6)$$

where, α is a pupil centric constraint given as

$$\alpha = \frac{\mathtt{F} - \mathtt{a_n}}{\mathtt{F}}. \quad (7)$$

3.3. Mapping Pupil-centric to Thin-lens

The thin-lens and pupil-centric model are geometrically equivalent if $\mathtt{P^g_{C_S}} = \mathtt{P^p_{C_S}}$ (Fig. 1,2). Comparing Eq. (2,5) and

Eq. (3,6), we get the following mapping. If

$$\lambda_g = \lambda_p + \frac{a_n}{\alpha}, \quad \text{where } \left[\frac{a_n}{\alpha} = O_2E_x(\text{Eq. } 7, 4)\right] \tag{8}$$

and $P_{C_{H_1}}=(x_1, y_1, z_1)$ is transformed by A_{pg} where,

$$A_{pg} = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{9}$$

and translated along the optic axis by $-a_n$ resulting in new world coordinates of $P_{C_{E_n}} = (x_1^g, y_1^g, z_1^g)$ as

$$x_1^g = \alpha x_1, y_1^g = \alpha y_1, z_1^g = z_1 - a_n. \tag{10}$$

then, the pupil-centric model is geometrically equivalent to a thin-lens model. This implies that thin-lens calibration with transformed world point $P_{C_{E_n}}$ (Eq. 10) will yield estimates for rotation R and parameters λ_g, a_n, α. Given these parameters and from Eq. 8, λ_p can be obtained. Next, we obtain projection equations for equivalent thin-lens model which will be used for analytical calibration.

4. Non-frontal Projection Equations

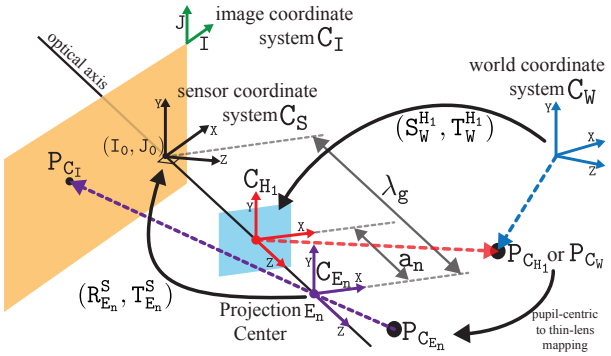

Figure 3. Coordinate Systems for non-frontal camera calibration.

In this section, we derive the non-frontal projection equations for a thin-lens imaging model derived from a pupil-centric imaging model (Sec. 3.3). These equations encode the calibration parameters which we want to estimate. See Fig. 3 for the imaging setting. Here, a world point $P_{C_W}=(X, Y, Z)$ is imaged at pixel location $P_{C_I}=(I, J)$. These two points can be related in terms of calibration parameters $U=[s_{ij} : 1 \leq (i,j) \leq 3, t_x, t_y, t_z, \lambda_g, s_x, I_0, J_0]$, sensor rotation $R = R_{H_2}^S$ (Eq. 1) parameterized by 2 Euler angles and lens parameters (a_n, F) as follows.

Transformation between C_W to C_{H_1}: Given a 3×3 rotation matrix $S_W^{H_1} = (s_{ij} : 1 \leq (i,j) \leq 3)$ and 3×1 translation vector $T_W^{H_1} = (t_x, t_y, t_z)$ between C_W and C_{H_1}, P_{C_W} can be represented in C_{H_1} as $P_{C_{H_1}} = (x_1, y_1, z_1)$,

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} = \begin{bmatrix} S_W^{H_1} & T_W^{H_1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_{C_W} \\ 1 \end{bmatrix} = \begin{bmatrix} s_{11}X+s_{12}Y+s_{13}Z+t_x \\ s_{21}X+s_{22}Y+s_{23}Z+t_y \\ s_{31}X+s_{32}Y+s_{33}Z+t_z \\ 1 \end{bmatrix} \tag{11}$$

Pupil-centric to Gaussian (C_{H_1} to C_{E_n}): Applying the equivalence relationship from Sec. 3.3, $P_{C_{H_1}}$ is transformed to $P_{C_{E_n}} = (x_1^g, y_1^g, z_1^g)$, thus converting a pupil-centric calibration problem to a thin-lens one as:

$$\begin{bmatrix} x_1^g \\ y_1^g \\ z_1^g \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \\ 0 & 0 & 1 & -a_n \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha y_1 \\ z_1 - a_n \\ 1 \end{bmatrix} \tag{12}$$

Transformation between C_{E_n} to C_S: Given a 3×3 rotation matrix $R_{E_n}^S = R_{H_2}^S = (r_{ij} : 1 \leq (i,j) \leq 3)$ and 3×1 translation vector $T_{E_n}^S = (0, 0, \lambda_g)$ between C_{E_n} and C_S, $P_{C_{E_n}}$ can be represented in C_S as $P_{C_S} = (x_s, y_s, z_s)$ where

$$\begin{bmatrix} x_s \\ y_s \\ z_s \\ 1 \end{bmatrix} = \begin{bmatrix} R_{E_n}^S & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & T_{E_n}^S \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_{C_{E_n}} \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11}x_1^g+r_{12}y_1^g+r_{13}z_1^g+r_{13}\lambda_g \\ r_{21}x_1^g+r_{22}y_1^g+r_{23}z_1^g+r_{23}\lambda_g \\ r_{31}x_1^g+r_{32}y_1^g+r_{33}z_1^g+r_{33}\lambda_g \\ 1 \end{bmatrix} \tag{13}$$

where, $R_{E_n}^S = R_{H_2}^S$ (Eq. 1) since xy planes of $C_{H_1}, C_{H_2}, C_{E_n}$ are parallel to each other and the optic axis is the common z-axis (Fig. 2 and Fig. 3), I is a 3×3 identity matrix. λ_g is the distance between the origin of C_{E_n} and C_S measured along the optic axis/z-axis of C_{E_n}. The center of projection is located at E_n. Let its coordinates in C_S can be obtained as $O_{C_S} = (r_{13}\lambda_g, r_{23}\lambda_g, r_{33}\lambda_g)$.

Central Perspective Projection: Given the world point P_{C_S}(Eq. 13) and location O_{C_S}, the intersection of the image ray connecting these two points on the non-frontal sensor can be computed as $P_{nf} = (x_{nf}, y_{nf})$.

Metric to Pixels (C_S to C_I): Next P_{nf} can be converted to image coordinates $P_{C_I} = (I, J)$ via pixel sizes (s_x, s_y) and CoD (I_0, J_0) as:

$$I = \frac{x_{nf}}{s_x} - I_0; \quad J = \frac{y_{nf}}{s_y} - J_0. \tag{14}$$

where, the skew in C_I is assumed to be 0 [8, 4]. Simplifying Eq. (11-14) results in a linear equation $\mathbf{GQ} = \mathbf{h}$ similar to DLT [1], relating i^{th} observed world and image point:

$$\underbrace{\begin{bmatrix} X_i & Y_i & Z_i & 0 & 0 & 0 & I_iX_i & I_iY_i & I_iZ_i & 1 & 0 \\ 0 & 0 & 0 & X_i & Y_i & Z_i & J_iX_i & J_iY_i & J_iZ_i & 0 & 1 \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} Q_1 \\ \vdots \\ Q_{11} \end{bmatrix}}_{\mathbf{Q}} = \underbrace{\begin{bmatrix} -I_i \\ -J_i \end{bmatrix}}_{\mathbf{h}} \tag{15}$$

where $\mathbf{Q}$ is the 3×4 projection matrix (upto scale with Q_{12}=1) non-linearly encoding 14 unknown calibration parameters expressed all together as $\mathbf{U} = [s_{ij}(3 \text{ angles}), r_{ij}(2 \text{ angles}) : 1 \leq (i,j) \leq 3, t_x, t_y, t_z, \lambda_g, s_x, I_0, J_0, a_n, \alpha]$ via 11 equations shown in Tab. 1 where $\lambda_{gx} = \frac{\lambda_g}{s_x}$, $\lambda_{gy} = \frac{\lambda_g}{s_y}$ and D is

$$D = \alpha r_{31} t_x + \alpha r_{32} t_y + r_{33}(t_z - a_n). \tag{16}$$

Next we analytically retrieve $\mathbf{U}$ from $\mathbf{Q}$.

5. Analytical Calibration

In this section, we analytically retrieve $\mathbf{U}$ from $\mathbf{Q}$. Given minimal 6 world-image point correspondences Eq. 15 can be solved for $\mathbf{Q}$ in the least squares sense using SVD. Since $\mathbf{Q}$ has 11 entries and $\mathbf{U}$ has 14 calibration variables, solving Eq. (17-27) for $\mathbf{U}$ given $\mathbf{Q}$ is an under-constrained problem. Thus, we propose to use a novel pupil-centric constraint and a technique to separately compute two parameters of CoD, thus making this a well-constrained problem with 12 equations in 12 variables.

Pupil-centric constraint: Assuming (a_x, d) are known in L, Eq. 4 and Eq. 7 can be combined in terms of (a_n, α) to

Table 1. Nonlinear equations relating $\mathbf{Q}$ and $\mathbf{U}$

$\mathtt{Q_1D} = \mathtt{I_0}(\alpha \mathtt{r_{31}s_{11}} + \alpha \mathtt{r_{32}s_{21}} + \mathtt{r_{33}s_{31}}) + \lambda_{\mathtt{gx}}\alpha(\mathtt{r_{22}s_{11}} - \mathtt{r_{21}s_{21}})$	(17)
$\mathtt{Q_2D} = \mathtt{I_0}(\alpha \mathtt{r_{31}s_{12}} + \alpha \mathtt{r_{32}s_{22}} + \mathtt{r_{33}s_{32}}) + \lambda_{\mathtt{gx}}\alpha(\mathtt{r_{22}s_{12}} - \mathtt{r_{21}s_{22}})$	(18)
$\mathtt{Q_3D} = \mathtt{I_0}(\alpha \mathtt{r_{31}s_{13}} + \alpha \mathtt{r_{32}s_{23}} + \mathtt{r_{33}s_{33}}) + \lambda_{\mathtt{gx}}\alpha(\mathtt{r_{22}s_{13}} - \mathtt{r_{21}s_{23}})$	(19)
$\mathtt{Q_4D} = \mathtt{J_0}(\alpha \mathtt{r_{31}s_{11}} + \alpha \mathtt{r_{32}s_{21}} + \mathtt{r_{33}s_{31}}) + \lambda_{\mathtt{gy}}\alpha(\mathtt{r_{11}s_{21}} - \mathtt{r_{12}s_{11}})$	(20)
$\mathtt{Q_5D} = \mathtt{J_0}(\alpha \mathtt{r_{31}s_{12}} + \alpha \mathtt{r_{32}s_{22}} + \mathtt{r_{33}s_{32}}) + \lambda_{\mathtt{gy}}\alpha(\mathtt{r_{11}s_{22}} - \mathtt{r_{12}s_{12}})$	(21)
$\mathtt{Q_6D} = \mathtt{J_0}(\alpha \mathtt{r_{31}s_{13}} + \alpha \mathtt{r_{32}s_{23}} + \mathtt{r_{33}s_{33}}) + \lambda_{\mathtt{gy}}\alpha(\mathtt{r_{11}s_{23}} - \mathtt{r_{12}s_{13}})$	(22)
$\mathtt{Q_7D} = \alpha \mathtt{r_{31}s_{11}} + \alpha \mathtt{r_{32}s_{21}} + \mathtt{r_{33}s_{31}}$	(23)
$\mathtt{Q_8D} = \alpha \mathtt{r_{31}s_{12}} + \alpha \mathtt{r_{32}s_{22}} + \mathtt{r_{33}s_{32}}$	(24)
$\mathtt{Q_9D} = \alpha \mathtt{r_{31}s_{13}} + \alpha \mathtt{r_{32}s_{23}} + \mathtt{r_{33}s_{33}}$	(25)
$\mathtt{Q_{10}D} = \mathtt{I_0D} + \lambda_{\mathtt{gx}}\alpha(\mathtt{r_{22}t_x} - \mathtt{r_{21}t_y})$	(26)
$\mathtt{Q_{11}D} = \mathtt{J_0D} - \lambda_{\mathtt{gy}}\alpha(\mathtt{r_{12}t_x} - \mathtt{r_{11}t_y})$	(27)

get a new constraint given by Eq. 44.

Computing $(\mathtt{I_0}, \mathtt{J_0})$: The 2 degrees of freedom in solving $\mathbf{U}$ given $\mathbf{Q}$ are attributed to the CoD: $(\mathtt{I_0}, \mathtt{J_0})$. First, we assume them as known and solve for remaining 12 parameters in $\mathbf{U}$ (Sec. 5.1). Later, in Sec. 6 we estimate it.

Definitions: We assume $\mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}$ (Eq. 11) and $\mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}}$(Eq. 13) are composed of Euler angles: (θ, ϕ, ψ) and $(\rho, \sigma, \tau = 0)$ respectively representing clockwise rotation about $\mathtt{x}, \mathtt{y}, \mathtt{z}$ axis while looking towards the origin. Also $\tau = 0$ implies $\mathtt{r_{21}} = 0$. We also define the 3×3 matrix $\mathtt{RS}_{\mathtt{W}}^{\mathtt{S}}$ relating $\mathtt{C_W}$ and $\mathtt{C_S}$ as:

$$\mathtt{RS}_{\mathtt{W}}^{\mathtt{S}} = \alpha(\mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}}\,\mathtt{A}_{\mathtt{pg}}^{-1}\mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}) \quad (\text{as } \mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}} = \mathtt{R}_{\mathtt{H_1}}^{\mathtt{S}}) \tag{28}$$

whose $(\mathtt{i}, \mathtt{j})^{\mathtt{th}}$ element is

$$\mathtt{rs_{ij}} = \mathtt{r_{i1}s_{1j}} + \mathtt{r_{i2}s_{2j}} + \alpha \mathtt{r_{i3}s_{3j}}. \tag{29}$$

5.1. Analytical Decomposition of Projection Matrix

In this section, we analytically solve the non-linear equations in Tab.1 for 12 calibration parameters. We denote known quantities in bold. Assuming that typical rotation $\mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}}$ between lens and sensor $\leq \frac{\pi}{2}$ along any axis, we deduce that $\mathtt{r_{33}} = \cos(\rho)\cos(\sigma) \geq 0$. Given $(\mathbf{I_0}, \mathbf{J_0})$, we first reduce Eq. (17-25) a set of 5 equations in 5 unknown variables: $(\lambda_{\mathtt{gx}}, \lambda_{\mathtt{gy}}, \mathtt{r_{13}}, \mathtt{r_{23}}, \alpha)$ as follows. Let $\mathtt{D}^2$ from Eq. (23-25) be:

$$\mathtt{D}^2 = \frac{\alpha^2 + (1-\alpha^2)\mathtt{r_{33}^2}}{\mathbf{Q_7^2} + \mathbf{Q_8^2} + \mathbf{Q_9^2}}. \tag{30}$$

Computing $\lambda_{\mathtt{gx}}, \lambda_{\mathtt{gy}}, \mathtt{r_{13}}, \mathtt{r_{23}}, \alpha$:

Equation* 1*: From Eq. (17-19,23-25,28), we get:

$$\mathtt{D}^2(\mathbf{Q_3Q_8} - \mathbf{Q_2Q_9}) = \lambda_{\mathtt{gx}}\alpha \mathtt{r_{33}rs_{21}} \tag{31}$$
$$\mathtt{D}^2(\mathbf{Q_1Q_9} - \mathbf{Q_3Q_7}) = \lambda_{\mathtt{gx}}\alpha \mathtt{r_{33}rs_{22}} \tag{32}$$
$$\mathtt{D}^2(\mathbf{Q_2Q_7} - \mathbf{Q_1Q_8}) = \lambda_{\mathtt{gx}}\alpha \mathtt{r_{33}rs_{23}} \tag{33}$$

Squaring and adding both sides of Eq. (31,32,33):

$$\mathtt{D}^4\mathbf{A} = \lambda_{\mathtt{gx}}^2\alpha^2\mathtt{r_{33}^2}\,(1 + (\alpha^2 - 1)\mathtt{r_{23}^2}) \quad \text{where,} \tag{34}$$

$$\mathbf{A} = (\mathbf{Q_3Q_8} - \mathbf{Q_2Q_9})^2 + (\mathbf{Q_1Q_9} - \mathbf{Q_3Q_7})^2 + (\mathbf{Q_2Q_7} - \mathbf{Q_1Q_8})^2$$

Equation* 2*: From Eq. (20-22,23-25,28), we get:

$$\mathtt{D}^2(\mathbf{Q_5Q_9} - \mathbf{Q_6Q_8}) = \lambda_{\mathtt{gy}}\alpha \mathtt{r_{33}rs_{11}} \tag{35}$$
$$\mathtt{D}^2(\mathbf{Q_6Q_7} - \mathbf{Q_4Q_9}) = \lambda_{\mathtt{gy}}\alpha \mathtt{r_{33}rs_{12}} \tag{36}$$
$$\mathtt{D}^2(\mathbf{Q_4Q_8} - \mathbf{Q_5Q_7}) = \lambda_{\mathtt{gy}}\alpha \mathtt{r_{33}rs_{13}} \tag{37}$$

Squaring and adding both sides of Eq. (35,36, 37):

$$\mathtt{D}^4\mathbf{B} = \lambda_{\mathtt{gy}}^2\alpha^2\mathtt{r_{33}^2}\,(1 + (\alpha^2 - 1)\mathtt{r_{13}^2}) \quad \text{where,} \tag{38}$$

$$\mathbf{B} = (\mathbf{Q_5Q_9} - \mathbf{Q_6Q_8})^2 + (\mathbf{Q_6Q_7} - \mathbf{Q_4Q_9})^2 + (\mathbf{Q_4Q_8} - \mathbf{Q_5Q_7})^2$$

Equation* 3*: From Eq. (17,23), (18, 24), (19,25), we get:

$$(\mathbf{I_0Q_7} - \mathbf{Q_1})^2 + (\mathbf{I_0Q_8} - \mathbf{Q_2})^2 + (\mathbf{I_0Q_9} - \mathbf{Q_3})^2 = \frac{\lambda_{\mathtt{gx}}^2\alpha^2(1 - \mathtt{r_{23}^2})}{\mathtt{D}^2} \tag{39}$$

Equation* 4*: From Eq. (20,23), (21, 24), (22,25), we get:

$$(\mathbf{J_0Q_7} - \mathbf{Q_4})^2 + (\mathbf{J_0Q_8} - \mathbf{Q_5})^2 + (\mathbf{J_0Q_9} - \mathbf{Q_6})^2 = \frac{\lambda_{\mathtt{gy}}^2\alpha^2(1 - \mathtt{r_{13}^2})}{\mathtt{D}^2} \tag{40}$$

Equation* 5*: From Eq. (17,20), (18, 21),(19,22):

$$(\mathbf{I_0Q_4} - \mathbf{J_0Q_1})^2 + (\mathbf{I_0Q_5} - \mathbf{J_0Q_2})^2 + (\mathbf{I_0Q_6} - \mathbf{J_0Q_3})^2 = \frac{\alpha^2}{\mathtt{D}^2}[\mathbf{I_0}^2\lambda_{\mathtt{gy}}^2 + \mathbf{J_0}^2\lambda_{\mathtt{gx}}^2 - [\mathbf{I_0}\lambda_{\mathtt{gy}}\mathtt{r_{13}} + \mathbf{J_0}\lambda_{\mathtt{gx}}\mathtt{r_{23}}]^2] \tag{41}$$

Solving $\mathtt{r_{13}^2}, \mathtt{r_{23}^2}, \alpha^2$: Eq. (30,34,38,39,40,41) can be solved in closed form to obtain two sets of solutions to $\mathtt{r_{13}^2}$, $\mathtt{r_{23}^2}$ and α^2. The optimal set is selected based on two criteria: (a) all positive solutions and (b) $0 \leq \mathtt{r_{13}^2}, \mathtt{r_{23}^2} \leq 1$. A set not satisfying any of the criteria is discarded. Thus, $(\mathtt{r_{13}}, \mathtt{r_{23}}, \alpha)$ are known with sign ambiguity which will be resolved later. But, $\mathtt{r_{33}}$ can be determined uniquely as: $\mathtt{r_{33}} = \sqrt{1 - \mathtt{r_{13}^2} - \mathtt{r_{23}^2}}$ and $\mathtt{r_{33}} \geq 0$.

Computing $\mathbf{D^2}$: Using Eq. 30 as $\mathbf{r_{33}^2}, \alpha^2$ are known.

Computing $\lambda_{\mathtt{gx}}, \lambda_{\mathtt{gy}}$: Given $\mathbf{r_{13}^2}, \alpha^2, \mathbf{D^2}$ and $\lambda_{\mathtt{gx}} > 0$ (sensor is behind the lens), Eq. 39 can be used to compute $\lambda_{\mathtt{gx}}$ uniquely. Similarly, using Eq. 40, $\lambda_{\mathtt{gy}}$ can be determined.

Determining $\mathtt{sign}(\alpha)$: From α^2, the magnitude of α is determined. The $\mathtt{sign}(\alpha)$ can be determined by verifying the following two conditions derived from known values of $(\mathtt{a_x}, \mathtt{d})$ and the constraint that $\mathtt{F} > 0$ for a converging lens. Let $\kappa = \mathtt{d} - \mathtt{a_x}$. Then,

- **Condition 1**: If $\kappa < 0$, then $\mathtt{sign}(\alpha) = \text{"+"}$, as from Fig. 4(a), $\mathtt{F} > 0$ only when $\alpha \in [0, 1)$.
- **Condition 2**: If $\kappa > 0$, then $\mathtt{sign}(\alpha) = -\mathtt{sign}(\mathtt{a_n})$, where $\mathtt{sign}(\mathtt{a_n})$ can be obtained from lens data-sheet.

Proof: Using Eq. 4 and Eq. 7, we get:

$$\frac{1}{\mathtt{a_n}} + \frac{1}{\kappa} = \frac{1}{\mathtt{F}}; \quad \mathtt{F} = \frac{\mathtt{a_n}}{1 - \alpha} \tag{42}$$

They can be solved for $(\mathtt{a_n}, \mathtt{F})$ in terms of κ and α as:

$$\mathtt{F} = \kappa\left(\frac{-\alpha}{1 - \alpha}\right) \tag{43}$$

$$\mathtt{a_n} = -\kappa * \alpha \quad \text{(pupil-centric constraint)} \tag{44}$$

Eq. 43 is a rectangular hyperbola in $(\mathtt{F}, \alpha)$ with asymptotes $\mathtt{F} = \kappa$ and $\alpha = 1$ as shown in Fig. 4(a,b) for $\kappa < 0$ and $\kappa > 0$ respectively. From Fig. 4(a), if $\kappa < 0$, then $\mathtt{F} > 0$ only when $\alpha \in [0, 1)$. Hence, **Condition 1** follows.

Similarly, from Fig. 4(b) if $\kappa > 0$, then $\mathtt{F} > 0$ only when $\alpha \notin [1, 0)$, which implies α can take both "+" and "−" sign. But, from Eq. 44, since $\kappa > 0$, we have that $\mathtt{sign}(\alpha)$ is opposite to $\mathtt{sign}(\mathtt{a_n})$. Here, we assume that $\mathtt{sign}(\mathtt{a_n})$ is

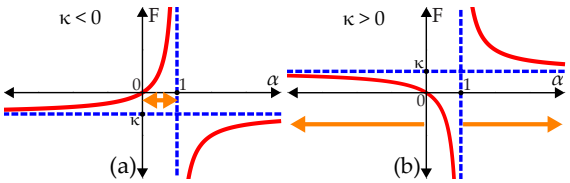

Figure 4. Hyperbolic relationship of $(\alpha, \mathtt{F})$ given (a) $\kappa < 0$ (b) $\kappa > 0$.

known from the lens data-sheet. Hence, **Condition 2** follows. Note that we only need $\mathtt{sign}(\mathtt{a_n})$ as prior knowledge which is a much weaker constraint than knowing $\mathtt{a_n}$, which in-fact we can now compute from 44. $\mathtt{F}$ can also be computed from Eq. 43.

Computing $3^{\mathtt{rd}}$ row of $\mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}$: From Eq.(17-24) and uniquely determined $\lambda_{\mathbf{gx}}, \lambda_{\mathbf{gy}}, \mathbf{r_{33}}, \mathbf{D^2}$, we get

$$\mathtt{s}_{31} = \frac{\mathbf{D^2}((\mathbf{Q_2Q_6}-\mathbf{Q_3Q_5})+\mathbf{I_0}(\mathbf{Q_5Q_9}-\mathbf{Q_6Q_8})+\mathbf{J_0}(\mathbf{Q_3Q_8}-\mathbf{Q_2Q_9}))}{\lambda_{\mathbf{gx}}\lambda_{\mathbf{gy}}\alpha^2\mathbf{r_{33}}}$$
$$\mathtt{s}_{32} = \frac{\mathbf{D^2}((\mathbf{Q_3Q_4}-\mathbf{Q_1Q_6})+\mathbf{I_0}(\mathbf{Q_6Q_7}-\mathbf{Q_4Q_9})+\mathbf{J_0}(\mathbf{Q_1Q_9}-\mathbf{Q_3Q_7}))}{\lambda_{\mathbf{gx}}\lambda_{\mathbf{gy}}\alpha^2\mathbf{r_{33}}}$$
$$\mathtt{s}_{33} = \frac{\mathbf{D^2}((\mathbf{Q_1Q_5}-\mathbf{Q_2Q_4})+\mathbf{I_0}(\mathbf{Q_4Q_8}-\mathbf{Q_5Q_7})+\mathbf{J_0}(\mathbf{Q_2Q_7}-\mathbf{Q_1Q_8}))}{\lambda_{\mathbf{gx}}\lambda_{\mathbf{gy}}\alpha^2\mathbf{r_{33}}}$$

Computing $1^{\mathtt{st}}$ and $2^{\mathtt{nd}}$ row of $\mathtt{RS}_{\mathtt{W}}^{\mathtt{S}}$: The $1^{\mathtt{st}}$ row of $\mathtt{RS}_{\mathtt{W}}^{\mathtt{S}}$: $(\mathtt{rs}_{11}, \mathtt{rs}_{12}, \mathtt{rs}_{13})$ can be determined from Eq. (35-37) as $(\mathbf{D^2}, \mathbf{r_{33}}, \lambda_{\mathbf{gy}}, \alpha)$ are known. Similarly, $2^{\mathtt{nd}}$ row of $\mathtt{RS}_{\mathtt{W}}^{\mathtt{S}}$: $(\mathtt{rs}_{21}, \mathtt{rs}_{22}, \mathtt{rs}_{23})$ can be determined from Eq. (31-33).

Computing $\mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}}, \mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}$: Since $\mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}}$ is parameterized as $(\rho, \sigma, \tau = 0)$, we have that $\mathtt{r}_{21} = \cos(\sigma)\sin(\tau) = 0$. Also, since $\mathtt{r}_{13} = \cos(\rho)\sin(\sigma)$ and $\mathtt{r}_{23} = -\sin(\rho)$, $\mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}}$ can be uniquely determined if $(\mathtt{r}_{13}, \mathtt{r}_{23})$ are known uniquely. But up till now, we have determined $(\mathtt{r}_{13}, \mathtt{r}_{23})$ with sign ambiguity leading to 4 possible solutions of $\mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}}$. To solve this ambiguity, lets assume that correct signs of $(\mathtt{r}_{13}, \mathtt{r}_{23})$ are known. Then $\mathtt{r}_{22} = \cos(\rho)$, $\mathtt{r}_{23} = -\sin(\rho)$, $\mathtt{r}_{21} = 0$ are known. Given $2^{\mathtt{nd}}$ row of $\mathtt{RS}_{\mathtt{W}}^{\mathtt{S}}$ and known $3^{\mathtt{rd}}$ row of $\mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}$, Eq. 29 can be used to obtain the following constraint,

$$\mathbf{rs_{21}} = \mathbf{r_{22}}\mathtt{s}_{21} + \alpha\mathbf{r_{23}}\mathtt{s}_{31} \tag{45}$$

which can be uniquely solved for $\mathtt{s}_{21}$. Forming similar linear equations for $\mathbf{rs_{22}}$ and $\mathbf{rs_{23}}$ in Eq. 29, $\mathtt{s}_{22}$ and $\mathtt{s}_{23}$ can be determined uniquely. Thus, the $2^{\mathtt{nd}}$ row of $\mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}$ is determined. By taking cross-product of $2^{\mathtt{nd}}$ and $3^{\mathtt{rd}}$ row of $\mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}$ (computed uniquely earlier), the $1^{\mathtt{st}}$ row and thus $\mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}$ can be determined. Thus, for all 4 signed solutions of $(\mathtt{r}_{13}, \mathtt{r}_{23})$, we get 4 solutions to $\mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}}$ and $\mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}$ leading to 4 solutions to $\mathtt{RS}_{\mathtt{W}}^{\mathtt{S}}$ in Eq. 46, where α is known uniquely. But as shown before, the $1^{\mathtt{st}}$ and $2^{\mathtt{nd}}$ row of $\mathtt{RS}_{\mathtt{W}}^{\mathtt{S}}$ have already been uniquely computed by a different constraint. Thus, by comparing the first two rows of known unique solution and predicted 4 solutions of $\mathtt{RS}_{\mathtt{W}}^{\mathtt{S}}$, we find optimal $\mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}}$ and $\mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}$ as the one with minimum Frobenius norm.

$$\underbrace{\begin{bmatrix} \mathtt{r}_{11} & \mathtt{r}_{12} & \mathtt{r}_{13} \\ 0 & \mathtt{r}_{22} & \mathtt{r}_{23} \\ \mathtt{r}_{31} & \mathtt{r}_{32} & \mathtt{r}_{33} \end{bmatrix}}_{\mathtt{R}_{\mathtt{E_n}}^{\mathtt{S}}(\rho,\sigma,0)} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \alpha \end{bmatrix}}_{\alpha\mathbf{A}_{\mathbf{pg}}{}^{-1}} \underbrace{\begin{bmatrix} \mathtt{s}_{11} & \mathtt{s}_{12} & \mathtt{s}_{13} \\ \mathtt{s}_{21} & \mathtt{s}_{22} & \mathtt{s}_{23} \\ \mathtt{s}_{31} & \mathtt{s}_{32} & \mathtt{s}_{33} \end{bmatrix}}_{\mathtt{S}_{\mathtt{W}}^{\mathtt{H_1}}(\theta,\phi,\psi)} = \underbrace{\begin{bmatrix} \mathtt{rs}_{11} & \mathtt{rs}_{12} & \mathtt{rs}_{13} \\ \mathtt{rs}_{21} & \mathtt{rs}_{22} & \mathtt{rs}_{23} \\ \mathtt{rs}_{31} & \mathtt{rs}_{32} & \mathtt{rs}_{33} \end{bmatrix}}_{\mathtt{RS}_{\mathtt{W}}^{\mathtt{S}}} \tag{46}$$

Computing $\mathtt{D}$: Since $\mathbf{R}_{\mathbf{E_n}}^{\mathbf{S}}, \mathbf{S}_{\mathbf{W}}^{\mathbf{H_1}}$ are known uniquely Eq. 23 can be used to determine $\mathbf{D}$.

Computing $\mathbf{T}_{\mathbf{W}}^{\mathbf{H_1}}$: Eq. (26,27) can be solved to get $(\mathtt{t_x}, \mathtt{t_y})$ uniquely. $\mathtt{t_z}$ can then be determined from Eq. 16.

6. Finding Center of Radial Distortion (CoD)

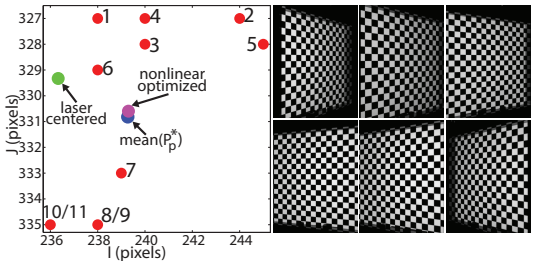

Figure 5. (best viewed in color) Left: blue dot $(239.3, 330.8)$ is analytical estimate mean($\mathtt{P}_{\mathtt{p}}^{*}$), green dot $(236.3, 329.3)$ is laser calibrated and magenta dot $(239.3, 330.6)$ is non-linear result initialized by laser center. It is close to analytical estimate. Right: real calibration images (Sec. 9).

In this section, we determine the center of radial distortion (CoD) or $\mathtt{P}_{\mathtt{p}} = (\mathtt{I}_0, \mathtt{J}_0)$. An accurate estimate of $\mathtt{P}_{\mathtt{p}}$ is critical as the origin of $\mathtt{C_S}$ (Sec. 4) is assumed to lie here. In order to compute $\mathtt{P}_{\mathtt{p}}$, we leverage the distinguishing property of CoD, that the distorted and undistorted points on a frontal sensor should be radially aligned. Thus, we define a cost function in Eq. 47, where different points in the image are hypothesized as valid candidates for $\mathtt{P}_{\mathtt{p}}$ and $\mathbf{U}$ is obtained via analytical calibration of Sec. 5. Then, $\mathbf{U}$ is used to project each world point $\mathtt{P}_{\mathtt{C_W}}^{\mathtt{i}}$ onto a *frontal* sensor to get undistorted sensor point $\mathtt{P}_{\mathtt{C_S}}^{\mathtt{i}}$ as well as de-warp the distorted observed image point $\mathtt{P}_{\mathtt{C_I}}^{\mathtt{i}}$ onto the *frontal* sensor as $\mathtt{P}_{\mathtt{C_S}}^{\mathtt{i}'}$. For optimal $\mathtt{P}_{\mathtt{p}}^{*}$, the two predicted points $\mathtt{P}_{\mathtt{C_S}}^{\mathtt{i}}$ and $\mathtt{P}_{\mathtt{C_S}}^{\mathtt{i}'}$ should be radially aligned about $\mathtt{P}_{\mathtt{p}}^{*}$. The discrepancy in this expected behavior can be used to define an error function as:

$$\mathtt{P}_{\mathtt{p}}^{*} = \arg\min_{\mathtt{P_p}} \sum_{\mathtt{i}=1}^{\mathtt{N}} \cos^{-1}\left[\frac{(\mathtt{P}_{\mathtt{C_S}}^{\mathtt{i}}-\mathtt{P_p})^{\mathtt{T}}(\mathtt{P}_{\mathtt{C_S}}^{\mathtt{i}'}-\mathtt{P_p})}{\|(\mathtt{P}_{\mathtt{C_S}}^{\mathtt{i}}-\mathtt{P_p})\|\|(\mathtt{P}_{\mathtt{C_S}}^{\mathtt{i}'}-\mathtt{P_p})\|}\right] \tag{47}$$

which minimizes the angle between the ray $\overrightarrow{\mathtt{P_p}\mathtt{P}_{\mathtt{C_S}}^{\mathtt{i}}}$ and $\overrightarrow{\mathtt{P_p}\mathtt{P}_{\mathtt{C_S}}^{\mathtt{i}'}}$. $\mathtt{P_p}$ is iteratively selected from a square window around the image center (half of image height and width) and the cost (Eq. 47) is computed at each of these locations. Fig. 5, shows $\mathtt{P}_{\mathtt{p}}^{*}$ obtained using our approach (red dots) on 11 input calibration images from a real dataset (Sec. 9). The mean of these points (blue dot) is chosen as the final optimal $\mathtt{P_p}$. $\mathtt{P_p}$ obtained from a laser-centering method [16] (green dot) and after a nonlinear optimization process (magenta dot) initialized by the laser-centering method is also shown. From Fig. 5, mean $\mathtt{P}_{\mathtt{p}}^{*}$ (blue dot) computed from our method is quite close to the nonlinearly optimized result.

7. Non-Linear Refinement

The analytical solution in Sec. 5 is used to initialize the non-linear optimization of $\mathbf{U}$ taking radial distortion into account. This step is optimized over 13 parameters in $\mathbf{U}$ excluding $\mathtt{F}$, as otherwise Eq. 4 will be required to constrain $(\mathtt{a_n}, \mathtt{F})$. We fix $\mathtt{F}$ from our analytical solution. An undistorted image point $\mathtt{X_u}$=$(\mathtt{x_u}, \mathtt{y_u})$ radially distorts to $\mathtt{X_d}$=$(\mathtt{x_d}, \mathtt{y_d})$ in $\mathtt{C_S}$ as $\mathtt{X_d} = \mathtt{X_u} + \mathtt{g}(\mathtt{X_u}, \mathtt{k}_1, \mathtt{k}_2)$ where $(\mathtt{k}_1, \mathtt{k}_2)$ are radial distortion parameters and

$$\mathtt{g}(\mathtt{X_u}, \mathtt{k}_1, \mathtt{k}_2) = [\mathtt{x_u}(\mathtt{k}_1\mathtt{r}_{\mathtt{u}}^2 + \mathtt{k}_2\mathtt{r}_{\mathtt{u}}^4) \quad \mathtt{y_u}(\mathtt{k}_1\mathtt{r}_{\mathtt{u}}^2 + \mathtt{k}_2\mathtt{r}_{\mathtt{u}}^4)] \tag{48}$$

with $r_u = \sqrt{x_u^2 + y_u^2}$. The iterative non-linear refinement of $(\mathbf{U}, k_1, k_2)$ is initialized using analytical $\mathbf{U}$ form Sec. 4 and $k_1 = k_2 = 0$. Given initial calibration parameters, i^{th} world point $P^i_{C_w}$ is projected onto a frontal sensor to get X^i_u and then distorted to get X^i_d. This point is then projected back to a non-frontal sensor using $(R^S_{E_n})^{-1}$ and pixel coordinates $P^{i'}_{C_I}$ are obtained. The optimal $\mathbf{U}^*$ is obtained for N observations using Levenberg-Marquardt optimization [8]:

$$\mathbf{U}^* = \arg\min_{\mathbf{U}} \sum_{i=1}^{N} \| P^i_{C_I} - P^{i'}_{C_I}(\mathbf{U}, k_1, k_2) \|_2^2 \qquad (49)$$

8. Experiments with Synthetic Data

The analytical calibration proposed in Sec. 5.1 assumes that no image noise or distortion and exact projection of world to image points, but both are present in real images. Therefore, its necessary to evaluate the robustness of our technique to these factors as follows. **Data generation:** A non-frontal camera is simulated with $\mathbf{U}$: $(\rho, \sigma, \tau) = (2.0, 4.0, 0.0)$ deg; $(\lambda_{px}, \lambda_{py}) = (8.4, 8.4)$ mm; $(I_0, J_0) = (240, 320)$ pixels; $(a_n, a_x, d, F) = (10.3, 28.8, 0.6, 16.2)$ mm; (k_1, k_2)=$(0.0022, -0.000013)$; (θ, ϕ, ψ)=$(0.1, 43.3, 0.0)$ deg; (t_x, t_y, t_z)=$(-65.0, -41.0, 102.2)$ mm. The synthetic 3D world points P_{C_w} are simulated and image points P_{C_I} are obtained (Sec. 4). Gaussian noise with 0 mean and std. deviation $\mu = \{.01, .02, .03, .04, .05, .1, .2, \cdots, 1\}$ pixels is added to the synthesized points. Then, at each of these noise levels, analytical calibration estimates are obtained (Sec. 5.1) for all calibration parameters except (I_0, J_0) which are assumed to be known (equal to $(240, 320)$ in this case). This experiment is repeated for 100 trials. Fig. 6 shows the results for various calibration parameters. The x-axis denotes the variation of noise and y-axis plots the mean of the difference between ground truth and estimates. The vertical bars represent the std. deviation of the estimates. As can be observed, for lower noise levels and the amount of distortion we applied, the analytical estimates are close to ground truth values with low std. dev. e.g. for λ_{px}, at 0.01 pixels noise, the std. deviation of the estimate is 0.08 pixels. But, for 1 pixel noise, the std. dev. is 16.6 pixels. Our corner detection on real data (Sec. 9) has std. dev. : ≈ 0.011 pixels.

9. Experiments with Real Images

We calibrate a non-frontal camera using a checkerboard pattern and four calibration methods, namely: (A) generalized pupil-centric model initialized by our analytical solution and (B) basic initialization (C) Gennery [7], (D) decentering distortion model [8]; and based on re-projection error (Tab. 2) and undistortion accuracy (Fig. 7) show that method (A) proposed in this paper outperforms all other techniques. **Camera Setup**: We use a AVT Marlin F-033C camera fitted with 1/2 inch Sony CCD sensor and C-mount Schneider Cinegeon 1.4/8mm Compact lens. The sensor is intentionally tilted with respect to the lens by $\approx 5^\circ$ (Fig. 7(b)).

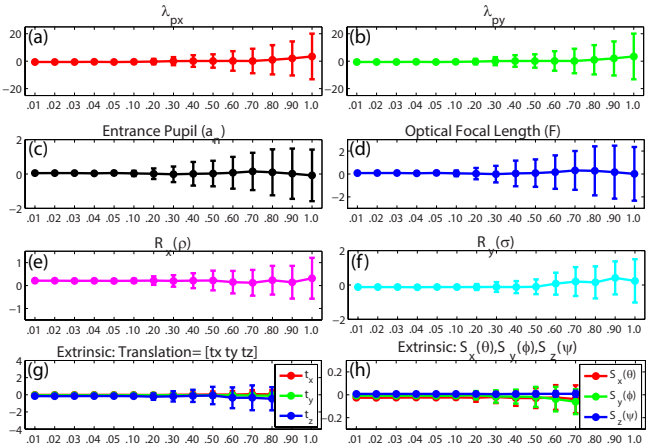

Figure 6. Absolute error, standard deviation vs. noise level (pixels) for various calibration parameters in $\mathbf{U}$. (best viewed in color)

The captured images have a resolution of 640×480 pixels. **Checkerboard(CB)**: A custom made precise glass checkerboard (Fig. 7(c)) with 20×20 squares, each having dimension 5×5mm is used to model known world points. The checkerboard corner location accuracy on the glass surface is $\pm .001$mm. **Data Acquisition**: A 2.5D dataset is collected by moving the CB along its surface normal (Fig. 7(d)) by $(0, 0.5, \cdots, 4.5)$mm (10 depths) and at each position an image is captured. The complete 2.5D dataset is collected thrice and the images averaged to handle random errors in CB positioning. The camera is then moved to a different location and another 2.5D data is collected. Likewise, 11 2.5D data sets are collected (See Fig. 5 (right) sample images from this dataset). The CB corners are detected using MATLAB Bouguet's toolbox [4]. The corner detection accuracy is ≈ 0.011 pixels along both image directions and is computed using the technique from [11].

Table 2. Calibration results on real data.

Calibration Method		A		B		C [7]		D [8]
		Initial	Final	Initial	Final	Initial	Final	Final
Scale	$\frac{s_y}{s_x}$	1.0002	0.9981	1.00	0.9989	1.00	1.0049	1.0049
	λ_p	8.742	8.738	8.198	8.128	8.198	8.154	8.288
Center of radial distortion (CoD)	I_0	239.27	238.90	240.0	237.51	239.27	229.48	226.86
	J_0	330.82	330.78	320.0	330.11	330.82	335.47	332.62
Radial distortion	k_1	-	− 0.002	-	− 0.002	-	− 0.004	− 0.0021
	k_2	-	.00002	-	.000005	-	.00016	.000012
Decentering distortion	p_1	-	-	-	-	-	-	− .000022
	p_2	-	-	-	-	-	-	− .00027
Non-frontal tilt angles	ρ	− 0.075	0.113	0.0	0.246	− 0.075	− 0.118	-
	σ	3.874	5.298	0.0	3.805	3.874	1.008	-
Entrance Pupil	a_n	6.691	6.817	6.5	6.437	-	-	-
Optical Focal Length F		8.503	-	8.2	-	-	-	-
Reprojection Error		-	**0.0504**	-	0.0959	-	0.3807	0.0534

Table 3. Std. deviation of final estimates from methods A,B,C,D.

Calibration Method	$\lambda_{px} = \frac{\lambda_p}{s_x}$	$\lambda_{py} = \frac{\lambda_p}{s_y}$	I_0	J_0	ρ	σ	a_n
A	**0.007**	**0.004**	**0.014**	**0.012**	**0.002**	**0.004**	**0.001**
B	0.017	0.013	0.027	0.024	0.005	0.008	0.003
C	0.027	0.008	0.08	0.078	0.006	0.007	–
D	0.031	0.028	0.028	0.03	–	–	–

Reprojection Error: Tab. 2 shows the calibration results of intrinsic parameters in $\mathbf{U}$ and the re-projection error using methods A, B, C and D from the camera and the 2.5D data captured above. Column A (left) shows the results from analytical method proposed in Sec. 5. Since, there are 11

2.5D datasets captured from different camera orientations, we have 11 analytical solutions for intrinsic parameters. We select the one which has closest (a_n, F) values with those in lens data-sheet. Also, we select points close to our computed CoD for this step where distortion is expected to be least. Column A (right) shows final refined parameters. In Column B, we assume the same imaging model as Column A, but instead use initial parameters from the lens data-sheet and assume tilt is 0. This is the *default* initialization. In Column C, we present the results from implementation of Gennery [7]. First, their imaging model assumes angle of incidence and exit at the entrance pupil is same which is not accurate [3]. Second, they optimize over decentering distortion which is redundant [13] and third, their initialization is default. In order to do a basic test of their calibration method, we avoided optimization over decentering distortion and initialized the sensor tilt and the CoD with our analytical initialization which is more accurate than using default initialization (Column C (left)). For other parameters we used default initialization. Column C (right) shows the results after non-linear refinement. The obtained re-projection error is 0.38 pixels which is highest implying the calibration from method C is inaccurate. Lastly, Column D shows results of applying decentering [8] calibration. Comparing the re-projection error from last row of Tab. 2 and std. deviation of estimated parameters from Tab.3, we observe that proposed method (A) achieves the lowest re-projection error of 0.0504 with lowest parameter variance across all methods. The sensor tilt is computed as $\approx 5.3°$ and used for omnifocus imaging (Fig. 8(a)) and depth from focus (Fig. 8(b)).

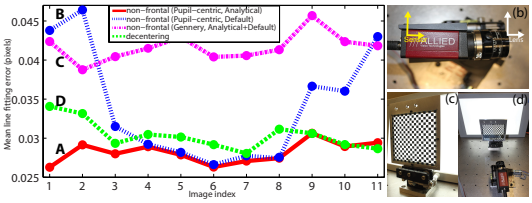

Figure 7. (a) Straight line fitting error to undistorted images using calibration parameters from technique A,B,C,D. (b,c,d) A non-frontal camera and the calibration setup. (best viewed in color)

Undistortion Error: Since accurate camera calibration would lead to accurate image undistortion using the obtained calibration parameters, we use this metric to compare various imaging model and calibration techniques. Specifically, we take the calibration results of technique A, B, C and D and undistort 11 CB images. Then, we detect all the corners in the undistorted image using MATLAB Bouguet's toolbox [4] and fit vertical and horizontal lines to detected corner points using orthogonal line fitting regression and compute the mean line fitting error as a metric for straightness of undistorted CB lines. Fig. 7 shows the results for all 11 images (x-axis) with methods A, B, C, D and their refined estimates from Tab. 2. As we see, the error is least (red solid line) for majority of the images undistorted by the method A proposed in this paper. The error (magenta dotted line) is maximum for the technique (C) [7]. The second best performance is by decentering based approach (green dotted line). **Application:** The sensor tilt is used for omnifocus imaging [10] and scene depth estimation [9] (Fig. 8(a-e)).

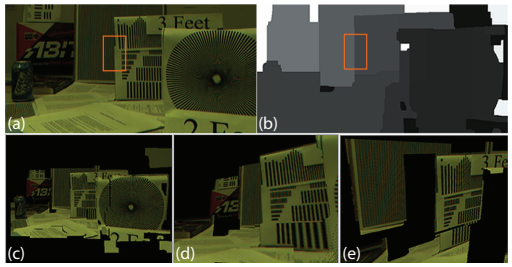

Figure 8. (a) Omnifocus Image [10]. (b) Depth from focus [9]. (c,d,e) 3D textured renderings of scene. (best viewed in color)

10. Conclusion

We have generalized pupil-centric imaging to arbitrary lens-sensor tilt and derived a mapping between pupil-centric and thin-lens imaging. We have proposed an analytical calibration technique assuming known center of radial distortion and a radial alignment metric based on analytical estimates to obtain all non-frontal calibration parameters.

11. Acknowledgments

This work was supported by US Office of Naval Research grant N00014-12-1-0259.

References

[1] Y. I. Abdel-Aziz and H. M. Karara. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In *Proceedings of the Symposium on Close-Range photogrammetry*, volume 1, 1971. 2, 4

[2] M. Aggarwal and N. Ahuja. Estimating sensor orientation in cameras. *ICPR*, 2000. 2

[3] M. Aggarwal and N. Ahuja. A pupil-centric model of image formation. *IJCV*, 2002. 2, 8

[4] J.-Y. Bouguet. Camera calibration toolbox for matlab. Website, 2000. http://www.vision.caltech.edu/bouguetj/calib_doc/. 4, 7, 8

[5] D. Brown. Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3):444–462, May 1966. 1

[6] S. Ganapathy. Decomposition of transformation matrices for robot vision. In *ICRA*, 1984. 2

[7] D. Gennery. Generalized camera calibration including fish-eye lenses. *IJCV*, 2006. 1, 2, 7, 8

[8] J. Heikkila and O. Silven. A four-step camera calibration procedure with implicit image correction. In *CVPR*, 1997. 1, 2, 4, 7, 8

[9] A. Krishnan and N. Ahuja. Range estimation from focus using a non-frontal imaging camera. *IJCV*, 1996. 1, 2, 8

[10] A. Kumar and N. Ahuja. A generative focus measure with application to omnifocus imaging. In *ICCP*, 2013. 1, 2, 8

[11] T. Melen and J. G. Balchen. Modeling and calibration of video cameras. In *SPIE*, 1994. 2, 7

[12] D. Scaramuzza, A. Martinelli, and R. Siegwart. A toolbox for easily calibrating omnidirectional cameras. In *IROS*, 2006. 2

[13] G. P. Stein. Internal camera calibration using rotation and geometric shapes. Technical report, MIT, 1993. 1, 2, 8

[14] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IJRA*, 1987. 1, 2

[15] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *PAMI*, 1992. 1, 2

[16] R. Willson and S. Shafer. What is the center of the image? In *CVPR*, 1993. 6

[17] Z. Zhang. A flexible new technique for camera calibration. *PAMI*, 2000. 1, 2