

Modeling the Effects of Windshield Refraction for Camera Calibration

Frank Verbiest¹, Marc Proesmans¹, and Luc Van Gool^{1,2}

¹ Center for Processing Speech and Images, ESAT-PSI, KU Leuven

² Computer Vision Lab, D-ITET, ETH Zürich

Abstract. In this paper, we study the effects of windshield refraction for autonomous driving applications. These distortion effects are surprisingly large and can not be explained by traditional camera models. Instead of using a generalized camera approach, we propose a novel approach to jointly optimize a traditional camera model, and a mathematical representation of the windshield’s surface. First, using the laws of geometric optics, the refraction is modeled using a local spherical approximation to the windshield’s geometry. Next, a spline-based model is proposed as a refinement to better adapt to deviations from the ideal shape and manufacturing variations. By jointly optimizing refraction and camera parameters, the projection error can be significantly reduced. The proposed models are validated on real windshield observations and custom setups to compare recordings with and without windshield, with accurate laser scan measurements as 3D ground truth.

1 Introduction

Camera calibration is an important topic in computer vision. It is the basis of general image based scene analysis, multi-view 3D reconstruction, visual based robotics, etc. For autonomous driving, a proper calibration of the cameras is important to map the free-space around the car, and estimate distances and relative velocities of the surrounding traffic. Contrary to prototype vehicles with roof mounted equipment, the placement of the cameras in commercial cars has to be both functional and aesthetically pleasing, and that often means behind the car windshield. The effect of refraction by a windshield, illustrated in Figure 1, is significant, and prompts the need to understand the image formation process. Little can be found on the effects of real windshields though, neither on curved surfaces in general for that matter. Hanel *et al.* [7] investigate the influence of a car windshield on depth calculation with stereo camera setups, and show that the difference in base length values can be highly significant. But the paper focuses on observation differences, not the solution.

Over the years, a wealth of camera models and calibration methods have been investigated. Amongst others, Sturm *et al.* [21] provide an extensive overview of camera models, such as pinholes, fisheyes, catadioptric cameras, including a variety of distortion models. For image formation processes involving refraction or reflection, existing work can be roughly subdivided in two main approaches. On

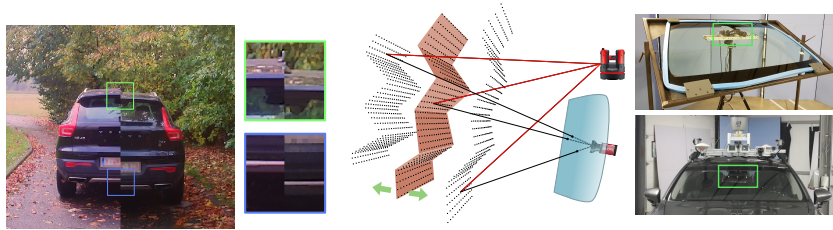


Fig. 1. Windshield refraction. Left: the split image shows the images captured without and with windshield, resp. The effect is visible as a mostly vertical shift. Right: our evaluation setups, an example of a camera behind a separately mounted windshield, and a camera mounted behind the windshield of car.

the one hand there are the generalized camera models that avoid explicit modeling of these processes. Grossberg and Nayer [6] proposed a ray-based camera model where each pixel has an associated 3D ray. To densely calibrate all pixels, they use a structured light-type approach using calibration reference planes with known positions. Many used the concept in complex optical setups, especially in underwater photography to account for refraction by housing or water tanks [22, 5, 13]. Ramalingam and Sturm *et al.* [20, 18, 17] utilize three calibration planes with unknown pose, but the method only applies to known, class specific, ray-distributions. Nishimura *et al.* [14] proposed to use the intersection of the calibration planes and realize a simple linear algorithm that applies to any ray-distribution. Miraldo *et al.* [12] introduced an RBF interpolation technique to densify the ray space when the measurements are sparse, assuming the rays vary smoothly across space.

Next to generalized camera models, there are some approaches that use an explicit formulation of refraction based on Snell’s law. Chari and Sturm [3] analysed the multi-view relationships between cameras when the scene contains a single refractive planar surface separating two different media. Agrawal *et al.* [2] investigated the effects of multi-layered flat refractive geometry, evaluated by means of a water tank. Pedersen *et al.* [16] evaluated the effects of refraction on underwater 3D reconstruction, where the water-housing-air interfaces are considered flat. Kunz and Singh [11] examine refraction of hemispherical pressure housing interfaces in an underwater context, in particular the effects where the camera mount is offset from the ideal center of the housing. Recently, the work by Pavel *et al.* [15] is one of the few to explicitly model refraction through a curved surface. They compare simulated and real checkerboard data viewed through a conic/tubular shaped glass, and estimate the geometry parameters of the surface via model inversion employing an RBF neural network.

Although generalized camera models are a natural choice for solving this particular problem, we consider two issues. First, they do not easily scale to larger settings, especially when dense measurements are needed. The calibration volume in our setting easily covers several meters across to lessen the effects of extrapolation when operating at larger distances. Secondly, by design these

approaches make abstraction of the different internal components of the setup, so they do not attempt to explain the image formation process.

In this paper, we propose an approach where the image formation process is modeled by an explicit formulation of the refraction through a curved surface, together with a traditional camera model (Section 2). First, using the laws of geometric optics, the refraction is modeled as a local spherical approximation to the geometry of the windshield (Section 3). We show that this approximation allows to reduce a non-trivial 3D problem to a mathematical formulation that is easy to solve numerically. By jointly optimizing the refraction and camera parameters, we obtain the surface’s radius, position, and the intrinsics and extrinsics of the camera behind the windshield. Second, to better adapt to deviations from the ideal spherical model and possible manufacturing variations, a spline-based refraction model is proposed as a refinement of the former (Section 4).

The mathematical analysis offers new insights into the relationship between the windshield curvature, the scene depth and the observed distortion in the image plane after projection. We will show that the refraction model compares favorably to the generalized approach. Also, we show that the image formation process is well approximated by a (traditional) pinhole camera model by means of a correction, which is advantageous because of their ease of use, since such models are well documented and widely supported. The proposed methodology is verified with real windshields, both separately mounted, and in real car scenarios. For this study, we use a 3D laser measurement system to create ground truth samples in 3D space and verify the projections in image space (Section 5).

2 Projection Model

Let us assume a basic pinhole camera model [8]. In this model, the image projection $\mathbf{m} = [x \ y]^\top$ of a 3D point $\mathbf{M} = [X \ Y \ Z]^\top$ can be written as,

$$\begin{bmatrix} \mathbf{m} \\ 1 \end{bmatrix} \simeq \mathbf{K} \mathbf{R}^\top (\mathbf{M} - \mathbf{t}) , \quad (1)$$

where \mathbf{K} , \mathbf{R} and \mathbf{t} are the calibration matrix, rotation matrix, and translation vector of the camera, respectively (\simeq denotes equality up to a non-zero scalar multiple). To introduce radial distortion and refraction in the basic pinhole camera model, we rewrite equation (1) as,

$$\begin{bmatrix} \mathbf{m} \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix} \quad (2)$$

$$\text{with } \begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix} \simeq \mathbf{p} \quad \text{and} \quad \mathbf{p} = \mathbf{R}^\top (\mathbf{M} - \mathbf{t}) , \quad (3)$$

where the point \mathbf{p} corresponds to the point \mathbf{M} but converted to the camera coordinate frame and the point $\mathbf{q} = [u \ v]^\top$ is the “normalized” projection before conversion to image coordinates by equation (2). We introduce radial distortion by modifying equation (2) as follows,

$$\begin{bmatrix} \mathbf{m} \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathcal{D}(\mathbf{q}) \\ 1 \end{bmatrix}, \quad (4)$$

with $\mathcal{D}([u \ v]^\top) = d \cdot [u \ v]^\top$ and $d = (1 + \kappa_1 r^2 + \kappa_2 r^4 + \dots)$ and $r^2 = u^2 + v^2$ and $\kappa_1, \kappa_2, \dots$ are parameters of radial distortion. To take into account the effects of refraction, we introduce the function $\mathcal{R}()$ in equation (3) as follows,

$$\begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix} \simeq \mathcal{R}(\mathbf{p}). \quad (5)$$

Considering the projection of point \mathbf{p} , we expect the function $\mathcal{R}(\mathbf{p})$ to return any point on the line of sight after refraction, *i.e.*, the line of sight as perceived by the camera.

3 Spherical Refraction

For a camera mounted directly behind a windshield, we approximate the local geometry of the windshield by a sphere with an appropriate position, radius, thickness and material index. The projection of a point through a refracting medium boils down to finding the ray projected back from the camera, which after refraction passes through that point. This is non-trivial, however, under the spherical assumption it is possible to reduce the 3D problem to a 2D problem that can be solved numerically.

3.1 Plane of Refraction

Consider the spherical geometry shown in Figure 2, where \mathbf{p} is the point that must be projected, \mathbf{c} is the center of the sphere, and \mathbf{t} is the center of projection of the camera. Together, the points \mathbf{p} , \mathbf{c} and \mathbf{t} , in general position, define a unique plane. Since the plane passes through the center of the sphere, it cuts the sphere in two halves. For reasons of symmetry, it then follows that all refraction takes place in this plane, called the *plane of refraction*, thereby reducing the 3D projection problem to a 2D problem. For the special cases where \mathbf{c} coincides with the origin \mathbf{t} or the points \mathbf{p} , \mathbf{c} and \mathbf{t} are collinear, a unique plane does not exist. However, it can be verified that in these cases the refraction has no effect and therefore the pinhole camera can be used as is. Given a plane of refraction, a *refraction coordinate frame* can be defined with its origin in \mathbf{c} , the x-axis pointing from \mathbf{c} to \mathbf{t} , the y-axis in the plane of refraction and perpendicular to the x-axis, and the z-axis perpendicular to the x- and y-axis, as illustrated in Figure 2.

3.2 Planar Ray Refraction

Using Snell's law of refraction, we describe the refraction of rays projected back from the camera. Assume everything is expressed in the refraction coordinate frame and consider the configuration in the plane of refraction, *i.e.*, the xy-plane,

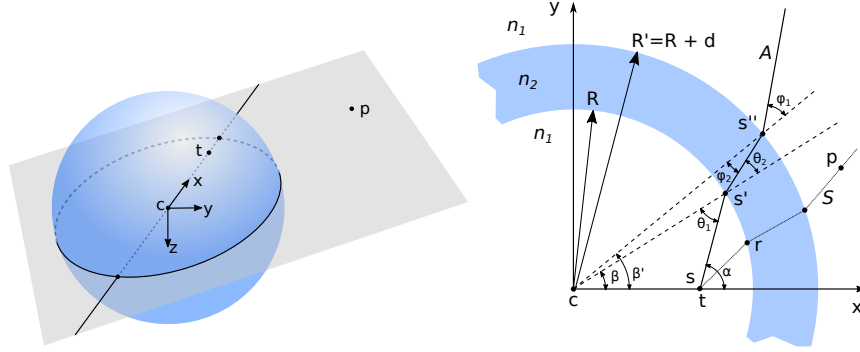


Fig. 2. Spherical geometry: All refraction takes place in the plane defined by the points \mathbf{p} , \mathbf{c} , and \mathbf{t} , *i.e.*, the *plane of refraction*.

as shown in Figure 2, where $\mathbf{p} = [p_1 \ p_2]^\top$ is the point that must be projected, $\mathbf{t} = [t_1 \ 0]^\top$ is the camera projection center, $\mathbf{c} = [0 \ 0]^\top$ is the center of the spheres, R is the radius of the sphere, d is the thickness of the glass, and n_1 and n_2 are the refraction indices of air and glass, respectively. Since we are only interested in the xy -plane, where the z -coordinate is 0, it is more convenient to work with 2D points instead, keeping in mind that the corresponding 3D points are readily obtained by adding a z -coordinate with value 0.

We describe the refraction of ray A in Figure 2. Let (\mathbf{s}, \mathbf{k}) , $(\mathbf{s}', \mathbf{k}')$ and $(\mathbf{s}'', \mathbf{k}'')$ be the starting points and direction vectors for each of the successive ray segments, respectively. The first ray segment originates in the camera projection center \mathbf{t} , where it makes an angle α with the x -axis, so $\mathbf{s} = \mathbf{t}$ and $\mathbf{k} = [\cos \alpha \ \sin \alpha]^\top$. Then, it intersects the inner sphere and is refracted. Points on the line associated with the first ray segment are given by, $\mathbf{m} = \mathbf{s} + \lambda \mathbf{k}$, where λ parametrizes the line. The intersections with the inner sphere are obtained by solving the quadratic equation, $\|\mathbf{s} + \lambda \mathbf{k}\|^2 = R^2$ for λ . For a starting point \mathbf{s} strictly inside the sphere, there is a positive $\lambda_+ > 0$ and a negative $\lambda_- < 0$ solution. The solution λ_+ is picked since it corresponds to an intersection that is on the forward travel path of the ray. The starting point of the second ray segment is given by this intersection, so $\mathbf{s}' = [s'_1 \ s'_2]^\top = \mathbf{s} + \lambda_+ \mathbf{k}$. Since $\theta_1 = \alpha - \beta$ with $\beta = \text{atan2}(s'_2, s'_1)$, the angle θ_2 can be obtained from Snell's law of refraction $n_1 \sin \theta_1 = n_2 \sin \theta_2$. The direction vector of the second ray segment is then given by $\mathbf{k}' = [\cos(\beta + \theta_2) \ \sin(\beta + \theta_2)]^\top$. For the ray segment exiting the outer sphere, the starting point \mathbf{s}'' and direction vector \mathbf{k}'' are obtained using similar steps.

3.3 Identifying the Solution

Finding the back-projected ray that after refraction passes through a point \mathbf{p} is equivalent to the problem of projecting that point \mathbf{p} . Consider again the plane of refraction as illustrated in Figure 2. We observe that the choice of angle α fully determines the path of ray A , so the search for a “solution” ray boils down to

the search for an angle. Also, the line corresponding to the ray segment $(\mathbf{s}'', \mathbf{k}'')$ exiting the outer sphere, divides the refraction plane in two sides. When this line passes through the point \mathbf{p} , the solution is found, otherwise, the point is on one side or the other. Because the line has a well defined direction vector \mathbf{k}'' associated with it, both sides are unambiguously defined. By scanning angles, the solution is identified as the angle for which the point changes sides. In practice, it is more convenient to work with a signed distance function, where the sign of the returned value determines which side of the line a point is on and the absolute value gives the point-to-line distance. The zero-crossing of this function corresponds to the solution. Let $\mathbf{k}'' = [k_1'' k_2'']^\top$, the distance function is defined as the projection of the vector $(\mathbf{p} - \mathbf{s}'')$ onto the direction perpendicular to \mathbf{k}'' ,

$$f(\alpha) = (\mathbf{p} - \mathbf{s}'')^\top \begin{bmatrix} -k_2'' \\ k_1'' \end{bmatrix}. \quad (6)$$

The *Regula-Falsi* method [23] is used for finding the root of $f(\alpha)$. This iterative method requires that we specify an interval for which the solution is known to lie in. Let $\mathbf{p} = [p_1 p_2]^\top$. For a point \mathbf{p} in the upper half-plane, *i.e.*, $p_2 > 0$, it is clear that a unique solution can be found in the interval $(0, \pi)$ and that $f(0)f(\pi) < 0$. Similarly, for a point \mathbf{p} in the lower half-plane, *i.e.*, $p_2 < 0$, the search interval is set to $(\pi, 2\pi)$. For a point \mathbf{p} on the x-axis, *i.e.*, $p_2 = 0$, the ray is not refracted, so the solutions for $p_1 > 0$ and $p_1 < 0$ are given by $\alpha = 0$ and $\alpha = \pi$, respectively.

In Figure 2, the solution is given by ray S , which intersects the inner sphere at point \mathbf{r} . Since the point \mathbf{r} is on line of sight as perceived by the camera, it can serve as the definition for the function $\mathcal{R}()$ in equation (5), where it should be noted that the function $\mathcal{R}()$ is expected to return points expressed in the camera coordinate frame. Automatic differentiation techniques (*e.g.*, Ceres [1]) are used to evaluate the derivatives for the above procedure.

3.4 Qualitative Assessment

This section intends to illustrate the effects of refraction on the image formation process when a windshield is placed in front of the camera. Consider a traditional pinhole camera and a back-projected ray for this camera. The points $\mathbf{M} = [X Y Z]^\top$ along this back-projected ray all project to a point $\mathbf{m} = [x y]^\top$. The offsets Δx and Δy are defined as the horizontal and vertical offsets/corrections to the projection \mathbf{m} as a result of introducing the spherical refraction model. We investigate these offsets for points along the back-projected ray as a function of depth Z , the radius of the sphere, and the position of \mathbf{m} in the image.

Consider a prototypical configuration for a camera mounted behind the windshield. The image has width $w = 1920$ and height $h = 1200$. The camera has focal length $f = 2200$ (in pixels), principal point in the middle of the image, and no radial distortion. The sphere has radius $R = 3$ (in m), thickness $d = 0.005$ (in m), and the refraction indices for air and glass are $n_1 = 1$ and $n_2 = 1.5$, respectively. The sphere position, in camera coordinates, is given by $\mathbf{c} = [0 \ (R - l) \sin \alpha \ -(R - l) \cos \alpha]^\top$, where $l = 0.05$ (in m) is the distance of

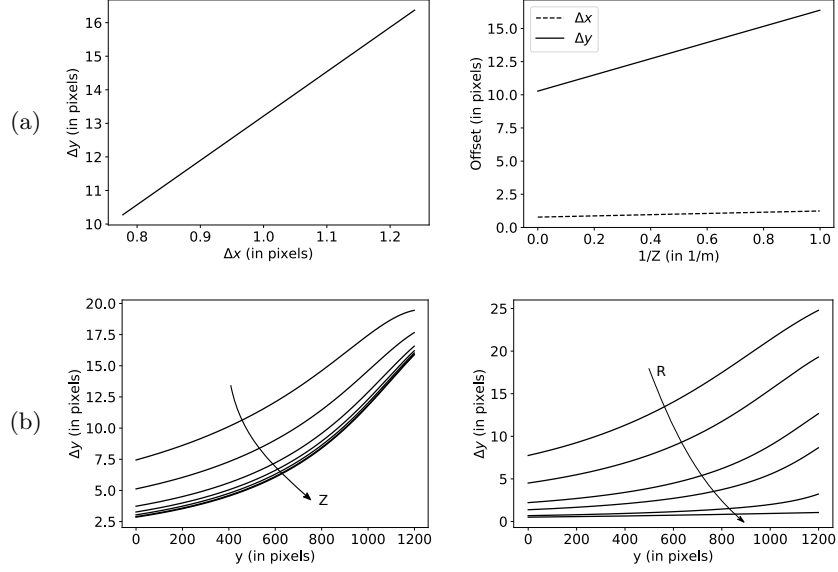


Fig. 3. Offset analysis. (a) Left: vertical offsets Δy plotted against horizontal offsets Δx for different depths. Right: horizontal offsets Δx and vertical offsets Δy plotted against the inverse depth $\frac{1}{Z}$. (b) Left: vertical offset Δy plotted against vertical positions y for different depths $Z \in \{1, 2, 5, 10, 20, 50, 100\}$ (in m), and $R = 3$ (in m). Right: vertical offset Δy for different radii $R \in \{1, 2, 5, 10, 50, +\infty\}$ (in m), and $Z = 10$ (in m).

the camera center to the glass, and $\alpha = 70$ (in degrees) is the angle the vector \mathbf{c} makes with the negative z -axis, *i.e.*, the orientation of the windshield. We arbitrarily select the point $\mathbf{m} = [\frac{3}{4}w, \frac{3}{4}h]^\top$ for analysis. In Figure 3a, the vertical offsets Δy are plotted against the horizontal offset Δx , for different depths Z . We observe that they form a line and that the offsets in the vertical direction are much larger. The fact that they form a line is unsurprising since, for a single image point, the back-projection and the different projections all take place in the same plane of refraction, whose intersection with the image plane presents as a line. Alternatively, globally the pinhole camera cannot model the effects of refraction, but locally in the image it can be a good approximation. So, the "offset" lines can be seen as epipolar lines resulting from the epipolar geometry formed by the such a local camera and the back-projection camera. Also, it is expected that as the depth is increased the resulting offsets along the epipolar line show an approximately inverse proportional relationship, which is confirmed in Figure 3a, where the offsets Δx and Δy are plotted against the inverse depth Z . However, this only holds for larger distances (as in our application). To illustrate the effect of refraction along the vertical direction, for different depths $Z \in \{1, 2, 5, 10, 20, 50, 100\}$ (in m), we plot the vertical offsets Δy against the vertical positions y of the points $\mathbf{m} = [\frac{1}{2}w, y]^\top$ for $y \in [0, h]$. The results are

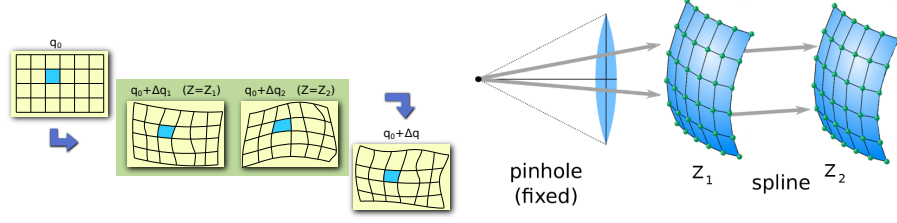


Fig. 4. The correction $\Delta \mathbf{q}(\mathbf{q}_0, Z)$ to \mathbf{q}_0 due to refraction is decomposed in two simpler 2D only corrections terms $\Delta \mathbf{q}_1(\mathbf{q}_0)$ and $\Delta \mathbf{q}_2(\mathbf{q}_0)$ corresponding to two different depths. Right: illustration of the ray-based refraction using splines.

shown in Figure 3b. The offsets increase substantially from top to bottom, up to ~ 20 pixels. The closer the rays are making a right angle with the surface of the sphere (*i.e.*, closer to the x-axis in Figure 2), the lower the effect of refraction, which for this configuration happens to be the case for points near the top of the image. Also, we observe that the highly non-linear offset curves converge for increasing depth. Similarly, the effect of the radius is shown in Figure 3b for $R \in \{1, 2, 5, 10, 50, +\infty\}$ (in m) and $Z = 10$ (in m). The important conclusions here is that the effect of refraction is substantial for even moderate amounts of curvature, *e.g.*, for $R = 5$ (in m) the vertical offset can go up to ~ 10 pixels. For a flat surface, *i.e.*, $R = \infty$ (in m), the effect is small and constant.

4 Spline Refraction

The previous section described the joint optimization of both a pinhole camera and spherical refraction model. In this section, we deal with the fact that the local windshield geometry may not be perfectly spherical, and that possible manufacturing variations are difficult to model. We propose a more generic refraction model as a refinement of the former (the pinhole camera parameters are kept fixed). This correction is decomposed in simpler parts that are each represented by spline-based functions (see Figure 4).

4.1 Decomposition

Consider the normalized projection \mathbf{q}_0 if there had been no refraction, which, using equation (3), can be expressed as,

$$\begin{bmatrix} \mathbf{q}_0 \\ 1 \end{bmatrix} \simeq \mathbf{p} . \quad (7)$$

The introduction of a correction $\Delta \mathbf{q}$ to the projection \mathbf{q}_0 allows to define the refraction function $\mathcal{R}()$ from equation (5) as follows,

$$\mathcal{R}(\mathbf{p}) = \begin{bmatrix} \mathbf{q}_0 + \Delta \mathbf{q} \\ 1 \end{bmatrix} , \quad (8)$$

where the correction $\Delta \mathbf{q}$ is a displacement in the normalized image plane due to refraction. The correction $\Delta \mathbf{q}$ is a function of the point \mathbf{p} , or equivalently, a function of the projection \mathbf{q}_0 and the depth Z .

Based on observations made for the spherical model (see Section 3.4), we assume that, for a fixed point \mathbf{q}_0 , the corrections for different depths form a line parametrized by inverse depth. Two corrections are needed to form a line. Assume the function $\Delta \mathbf{q}(\mathbf{q}_0, Z)$ is available, then by evaluation for two different depths Z_1 and Z_2 , we obtain the corrections $\Delta \mathbf{q}_1(\mathbf{q}_0) = \Delta \mathbf{q}(\mathbf{q}_0, Z_1)$ and $\Delta \mathbf{q}_2(\mathbf{q}_0) = \Delta \mathbf{q}(\mathbf{q}_0, Z_2)$. Then based on the above assumptions, the correction $\Delta \mathbf{q}(\mathbf{q}_0, Z)$ is approximated as,

$$\Delta \mathbf{q}(\mathbf{q}_0, Z) \approx \frac{\Delta \mathbf{q}_1(\mathbf{q}_0) \left(\frac{1}{Z_2} - \frac{1}{Z} \right) + \Delta \mathbf{q}_2(\mathbf{q}_0) \left(\frac{1}{Z} - \frac{1}{Z_1} \right)}{\left(\frac{1}{Z_2} - \frac{1}{Z_1} \right)} \quad (9)$$

where $\Delta \mathbf{q}_1(\mathbf{q}_0)$ and $\Delta \mathbf{q}_2(\mathbf{q}_0)$ are defined over the normalized image only. Essentially, the original correction term is decomposed in two simpler 2D only corrections terms, see Figure 4. Next, the corrections $\Delta \mathbf{q}_1(\mathbf{q}_0)$ and $\Delta \mathbf{q}_2(\mathbf{q}_0)$ are modeled using spline-based functions $\mathbf{f}_1(\mathbf{q}_0; \boldsymbol{\theta}_1)$ and $\mathbf{f}_2(\mathbf{q}_0; \boldsymbol{\theta}_2)$, respectively, where the parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the subject of optimization during a calibration procedure. The final spline-based correction term is then given by,

$$\Delta \mathbf{q}(\mathbf{q}_0, Z) \approx \frac{\mathbf{f}_1(\mathbf{q}_0; \boldsymbol{\theta}_1) \left(\frac{1}{Z_2} - \frac{1}{Z} \right) + \mathbf{f}_2(\mathbf{q}_0; \boldsymbol{\theta}_2) \left(\frac{1}{Z} - \frac{1}{Z_1} \right)}{\left(\frac{1}{Z_2} - \frac{1}{Z_1} \right)} \quad (10)$$

We arbitrarily select $Z_1 = 1$ and $Z_2 = +\infty$.

4.2 Cubic Spline Representation

Each of the spline-based functions $\mathbf{f}_1(\mathbf{q})$ and $\mathbf{f}_2(\mathbf{q})$ from the previous section has two scalar components (a horizontal and vertical component). Let the function $s(\mathbf{q})$, or $s(u, v)$, denote such a component, where $\mathbf{q} = [u v]^\top$ is the normalized projection. The function $s(u, v)$ consists of a 2D grid of connected bi-cubic spline patches, where neighbouring patches share function values and derivatives at their corners to ensure continuity across patches.

A single bi-cubic patch function $p(u, v)$ is written as,

$$p(u, v) = [u^3 u^2 u 1] \mathbf{C} [v^3 v^2 v 1]^\top, \quad (11)$$

where the coefficients of the 4×4 matrix \mathbf{C} are completely determined by the function values s and the derivatives $\frac{\partial s}{\partial u}$, $\frac{\partial s}{\partial v}$, and $\frac{\partial^2 s}{\partial u \partial v}$ stored at the four corners of the patch. Indeed, matching the function $p(u, v)$ with these values yields 16 linear equations in the 16 coefficients. The function values and derivatives that are shared across patches are used as parameters in an optimization. Because the above spline representation has many parameters, regularization is required to prevent over-fitting. Smoothness is enforced using thin-plate constraints on each patch $p(u, v)$, $E = \iint_{\Omega} \left(\left(\frac{\partial^2 p}{\partial u^2} \right)^2 + 2 \left(\frac{\partial^2 p}{\partial u \partial v} \right)^2 + \left(\frac{\partial^2 p}{\partial v^2} \right)^2 \right) du dv$, where Ω is the domain of

the patch. It is easy to confirm that this can be written as, $E = \mathbf{z}^\top \mathbf{A} \mathbf{z}$, where \mathbf{z} is a 16-vector that holds the function values and derivatives of the four corners of the patch, and \mathbf{A} is a fixed 16×16 symmetric positive semi-definite matrix. The cost for the function $s(u, v)$ is the sum over of all patches.

5 Experiments

5.1 Acquisition

The acquisition setup used to obtain accurate real-world 2D-3D correspondences for camera calibration is shown in Figure 1. Here, a 3D laser measurement device (Leica 3D Disto) scans the scene using a laser dot, where for each position of the laser an image is taken. The 2D image features are obtained by sub-pixel measurement of the observed laser ‘blobs’ in the images. The scene consists of movable panels to provide sufficient depth variation. Typically, around 500 points are measured for each of the 3 to 4 panel positions.

An important argument in considering this approach is that it allows to get rid of any uncertainty on the 3D measurements in our evaluation and thus any possible interference/interchange between camera parameters and feature point locations is avoided. It is useful to note that our measurement approach also relates to methods using structured light or chess board panels with known postures [22, 5, 13, 6, 12], but given our working volumes, the laser is less bulky.

5.2 Calibration Procedure

Once the 2D-3D correspondences are known, we can proceed with the calibration process. As a first step, we ignore refraction and determine the intrinsics and extrinsics of the pinhole camera behind the windshield through the use of the minimal solver [10] inside a RANSAC [4] loop. The initial camera is then refined by minimizing the reprojection error using non-linear least squares [1] with a robust loss function (M-estimators [9]). The scale parameters for the robust loss function is set to the *Median Absolute Deviation* (MAD) [19] of the residuals.

Next, the parameters of the spherical refraction model are initialized to represent a typical configuration (*e.g.*, see assessment above) relative to the previously determined camera. Initially, this relative configuration is kept fixed, while the parameters of the camera are optimized by robustly minimizing the reprojection error. This brings the camera closer to its expected solution. Finally, by jointly optimizing the parameters of the camera and the spherical refraction model, the calibration is obtained. Note, in practice we fix the thickness of the windshield ($d = 5.3 \text{ mm}$) and the refraction indices of glass and air ($n_{air} = 1$ and $n_{glass} = 1.5$), as they can be measured beforehand.

Finally, we keep the camera parameters fixed and optimize the parameters of a spline-based refraction model, again, by robustly minimizing the reprojection error, and the parameters initialized to 0. We typically use 4×4 spline patches. To prevent over-fitting, a thin-plate regularization term is used, so that the

	CamA						CamA'				
	P	WS-P	WS-BC	WS-SPH	WS-SPL		P	WS-P	WS-BC	WS-SPH	WS-SPL
f	1841.2	1852.1	1865.5	1842.3	/		2173.8	2193.3	2223.8	2174.2	/
c_x	940.9	939.6	1190.1	941.3	/		968.3	965.8	1433.5	961.9	/
c_y	708.6	706.2	675.0	710.7	/		604.8	593.6	564.1	600.2	/
σ_{MAD}	0.137	1.44	0.312	0.199	0.095		0.135	2.47	0.57	0.264	0.11

Table 1. Comparisons on camera intrinsics (P=pinhole, WS=windshield, BC=Brown-Conrady, SPH=spherical refraction, SPL=spline refraction)

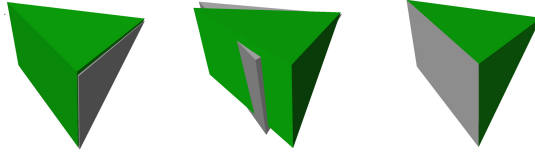


Fig. 5. 3D view on the extrinsics. From left to right: WS-P, WS-BC, WS-SPH (grey = ground truth, green = calculated)

cost function becomes $E = E_{reproj.} + \lambda E_{thinplate}$, where λ controls the amount of regularization. The λ is determined automatically by splitting the 2D-3D correspondences in a training set and a test set. The training set is used during optimization, and the resulting parameters are verified against the test set to find the optimal λ , *i.e.*, minimal cost on the test set.

5.3 Comparison with and without Windshield

In this paragraph, we provide a number of numerical results on the proposed approach, using a variety of visualizations. For the illustrations, we chose a few prototypical example cameras, for the separate windshield setup (CamA) and in the real automotive context (CamB). Yet, we've observed that the results are very similar for various setups and windshields that we have experimented with so far. In total about 6 different cameras with various focal lengths have been used on 4 different windshields.

Table 1 and Figure 5 display some interesting observations when determining the camera intrinsics and extrinsics as viewed behind a windshield. The table lists two 2Mpix PointGrey cameras CamA and CamA' with different focal lengths. The ground truth in the resp. first columns, are the measurements based on observations without windshield. The next three columns indicate the intrinsics when minimizing the pinhole alone (step 1 in 5.2), a Brown-Conrady camera model, and the spherical refraction method (step 2 in 5.2). The parameters of the Brown-Conrady model have been defined to cover both radial and tangential distortion of lens systems, however in this context, these additional parameters tend to interchange information between focal length, distortion parameters and principal point to try to "explain" the distortion. The re-projection error drops (σ_{MAD}), however, the numbers clearly deviate from the actual cameras. Using

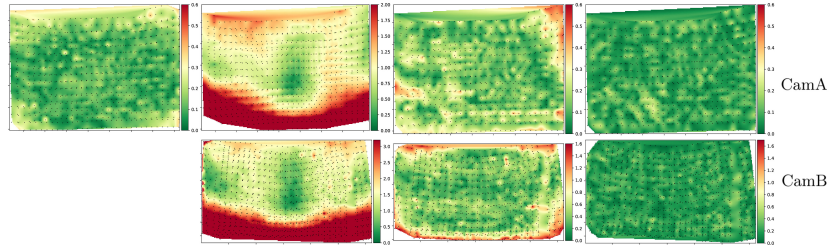


Fig. 6. Projection error maps, Ground truth comparison from left to right: (1) without WS (2) with WS, no refraction (3) with WS, spherical refraction (4) with WS. Top row: separate windshield setup. Bottom row: automotive setup.

the spherical refraction model though, the parameters come very close to the original. Using the spline refraction model, the re-projection error drops even more (cfr. sequel), but the camera parameters are not affected during this stage (Section 5.2).

Figure 5 provides a few 3D impressions of the results. The ground truth cone is in gray, it is the camera without windshield. The green cones are the ones when fitting the measurements behind windshield with resp. the pinhole alone, the Brown-Conrady model, and the spherical refraction method. The visualization confirms the observations above, the Brown-Conrady model tends to overcompensate, the spherical method provides the best fit.

Figure 6 shows the re-projections plots for the different stages of the calibration process described in Section 5.2. They show the error between the actual observed image coordinates of the 3D (laser) measurements, and the projected image coordinates. The top row shows the results when using the separately mounted windshield (CamA). The left image shows the error plot of the camera, the ground truth calculation without windshield, the 3 consecutive images show the error plots when fitting the data as observed behind the windshield, with resp. the pinhole alone, with the spherical refraction model, and finally with the spline refraction model. As indicated before and expected, the non-linear vertical shift can not be explained by the pinhole alone. When introducing the spherical refraction, the error already reduces dramatically. And finally, when introducing the spline approach, the average error rate σ_{MAD} drops from a pixel error 1.44 down to 0.199 and further to < 0.1 pixel. These numbers are also indicated in the Table 1 (CamA). The bottom row of Figure 6 shows the results for a real automotive setup. Since the camera was mounted inside the car, one can not compare to the situation without windshield, but the resemblance of the plots with the experiments above is striking. For this particular case CamB, σ_{MAD} drops from close to 2.1 pixel error, over 0.54, arriving at slightly around 0.23 pixel. A useful footnote in this context is that the lower resolution of these cameras (1.3Mpix), and the bayer effects have more impact on the subpixel accuracy of the measurements, and consequently the average error.

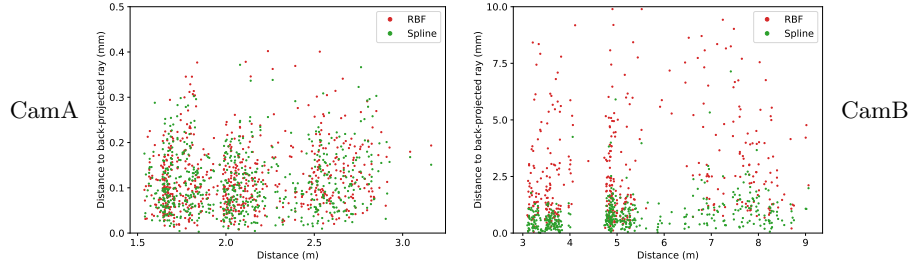


Fig. 7. Comparison with generalized calibration: distance between the original measurements, and the back-projected ray assigned to the measurements.

5.4 Comparison with Generalized Calibration

Generalized camera models are natural candidates to solve problem involving refraction. Actually, our method can be seen as a class-specific ray-pixel mapping, whereas in earlier work the ray distribution is governed by specific smoothness constraints, ours is governed by the refraction of the spherical surface we introduced. We will show that the mathematical formulation of the joint camera-refraction model imposes better constraints on the ray distribution. and our approach compares favorably to the generalized approach.

Given the fact that our laser measurement space is sparse, we used the RBF-based ray interpolation method [12] as a proper representative. In Figure 7, we considered two cameras CamA and CamB, also used in Figure 6. We subdivided the measurements (for each camera about 1300 samples) in a training (800) and test set (500), where the training set is used for the calibration process, and the test set for evaluation. For the generalized method, we used the interpolation function $\phi(r) = \sqrt{\gamma^2 + r^2}$. We executed several runs where each time 50 control centers are randomly chosen, and we looped over the shape parameter γ to find the optimal solution. For our approach, we use the ray representation of the spline refraction model: the forward projection (see Section 4) of a plane at a distance Z to the camera yields a 3D point for each 2D point in the image plane. The rays are defined by two such planes at different distances.

Our evaluation criterion is as follows. For each 2D-3D measurement pair (\mathbf{m}, \mathbf{M}) , we define the error as the distance between the original measurement \mathbf{M} , and the back-projected ray assigned to the 2D point \mathbf{m} . Figure 7 plots these errors against the distance of the original measurements. The RMSE for both methods for the higher resolution camera CamA are 0.147817 mm (SPL) and 0.159189 mm (RBF). Those for the automotive camera CamB are 1.31364 mm (SPL) and 6.19127 mm (RBF). As can be observed by the figure and the numbers, the errors are in favor of the spline refraction approach and are especially noticeable for the automotive camera. A possible reason is that the larger noise levels for these cameras are better handled by an explicit noise model for the projection – such as our Maximum Likelihood Estimation (MLE) formulation

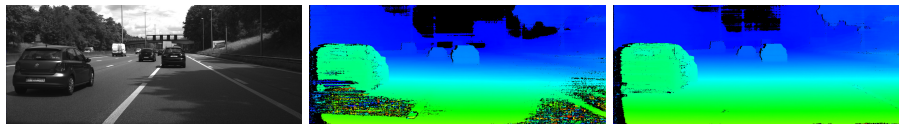


Fig. 8. Left: image of a stereo pair of type CamB, middle: stereo disparity when using (traditional) camera pinhole calibration and rectification. right: stereo disparity when using rectification after spline based refraction modeling.

above (Section 5.2) – as opposed to the set of algebraic equations (set to be 0) used by [12] that may well fit the noise itself and distort the ray distribution.

5.5 Windshield Image Undistortion and Rectification

One would not easily associate the concept of image undistortion with complex image formation processes involving refraction. Yet, following the discussion in Section 4, we have shown that the effects of windshield refraction can be approximated by a image plane distortion on the pinhole camera model, and furthermore, we can settle for a given Z since the dependency on depth decreases rapidly for large Z 's (Figure 3b). This can be advantageous when there is a need to use camera models in traditional computer vision pipelines. We evaluated this approximation on an automotive CamB type stereo setup behind the windshield. Figure 8 shows the effect of rectification on the stereo disparity. When only using the pinhole model calibration, there will be misalignments between the images which disturbs the disparity calculation. When using our refractive undistortion, the improvement is substantial.

6 Conclusion

We have proposed two models of refraction to deal with the distortion effects due to a windshield placed in front of a camera. The spherical refraction model shows that with using even a simple geometry it is possible to significantly reduce the re-projection error, as well as, accurately determine the camera behind the windshield as demonstrated by ground-truth experiments. This strengthens our belief that the approach is valid. The spline-based refraction is more generic in approach and reduces the error even further, and can deal with variations otherwise difficult to model. A potential comment is that our calibration is confined to a specialized lab setting. So, as an indication for future research, we will investigate the integration of our spherical model in a structure-from-motion based calibration approach. As a final note, we invite the reader to view our observations also in the context of other applications that involve refraction through curved surfaces.

Acknowledgements This work was supported by the TRACE project with Toyota Motors Europe (TME).

References

1. Agarwal, S., Mierle, K., Others: Ceres solver. <http://ceres-solver.org>
2. Agrawal, A., Ramalingam, S., Taguchi, Y., Chari, V.: A theory of multilayer flat refractive geometry. 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3346–3353. IEEE (2012) (2012)
3. Chari, V., Sturm, P.: Multiple-view geometry of the refractive plane. 20th British Machine Vision Conference. pp. 1–11. in BMVA (2009)
4. 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)
5. Gregson, J., Krimerman, M., Hullin, M.B., Heidrich, W.: Stochastic tomography and its applications in 3d imaging of mixing fluids. *ACM Trans. Graph. (Proc. SIGGRAPH 2012)* **31**(4), 52:1–52:10 (2012)
6. Grossberg, M.D., Nayar, S.K.: The raxel imaging model and ray-based calibration. *International Journal on Computer Vision* **61**(2), 119–137 (Feb 2005)
7. Hanel, A., Hoegner, L., Stilla, U.: Towards the influence of a car windshield on depth calculation with a stereo camera system. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLI-B5* pp. 461–468 (2016)
8. Hartley, R.I., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edn. (2004)
9. Huber, P.J.: *Robust Statistics*, pp. 1248–1251. Springer Berlin Heidelberg, editors Lovric, Miodrag, Berlin, Heidelberg (2011)
10. Kukeleva, Z., Bujnak, M., Pajdla, T.: Real-time solution to the absolute pose problem with unknown radial distortion and focal length. In: *Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV'13)*. pp. 2816–2823. Washington, DC, USA (2013)
11. Kunz, C., Singh, H.: Hemispherical refraction and camera calibration in underwater vision. *OCEANS2008* pp. 1–7 (2008)
12. Miraldo, P., Araujo, H.: Calibration of smooth camera models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9 (2013)
13. Narasimhan, S., Nayar, S., Sun, B., Koppal, S.: Structured light in scattering media. In *Proc. ICCV*, pages 420–427 (2005)
14. Nishimura, M., Nobuhara, S., Matsuyama, T., Shimizu, S., Fujii, K.: A linear generalized camera calibration from three intersecting reference planes. In *Proc. ICCV* pp. 2354–2362 (2015)
15. Pavel, S., Sandor, C., Csato, L.: Distortion estimation through explicit modeling of the refractive surface. *Artificial neural networks and machine learning - ICANN* pp. 17–28 (2019)
16. Pedersen, M., Bengtson, S.H., Gade, R., Madsen, N., Moeslund, T.B.: Camera calibration for underwater 3d reconstruction based on ray tracing using snell's law. *CVPR Workshops (CVPRW)* pp. 1410–1417 (2018)
17. Ramalingam, S., Lodha, S., Sturm, P.: A generic structure-from-motion framework. *Computer Vision and Image Understanding*, vol. 103, pp. 218–228, September 2006. (2006)
18. Ramalingam, S., Sturm, P., Lodha, S.: Towards complete generic camera calibration. *IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA*, pp. 1093–1098 (2005)

19. Rousseeuw, P.J., Croux, C.: Alternatives to the median absolute deviation. *Journal of the American Statistical Association* 88 **88**(424), 1273–1283 (1993)
20. Sturm, P., Ramalingam, S.: A generic concept for camera calibration. In *Proc. ECCV*, pages 1–13 (2004)
21. Sturm, P., Ramalingam, S., Tardif, J.P., Gasparini, S., Barreto, J.: Camera models and fundamental concepts used in geometric computer vision. *Foundations and Trends® in Computer Graphics and Vision* **6**(1–2), 1–183 (2011)
22. Trifonov, B., Bradley, D., Heidrich, W.: Tomographic reconstruction of transparent objects. *Eurographics Conference on Rendering Techniques* (2006)
23. Wikipedia: False position method — Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/False_position_method (2018)