Real Cameras are Radiometric

Ties Harder Felix Seegräber David Nakath Hendrik Sauer Kevin Köser James Pritts
Kiel University, Germany

Abstract

We propose the radiometric camera, a model that lifts any central geometric camera model to a radiometric one. The lifted model keeps the projection of the geometric model and adds a finite aperture, a free-form focus surface and a generic vignetting function. We calibrate it by inverse rendering directly on the relative irradiance of the sensor, from the same images of standard calibration targets used in geometric calibration. Geometric calibration absorbs the bias of its feature detections. The proposed radiometric calibration uses no features, and our radiometric cameras have lower reprojection errors on held-out test sets. The focus surface extends defocus modeling from the thin lens to any central camera, and we use it in a first multi-view radiometric reconstruction, which recovers the in-focus reflectance of a texture from defocused images. The vignetting is estimated jointly and accurately reproduces a flat field measured through a diffuser. Because every standard camera model lifts to a radiometric one, the radiometric camera naturally and directly extends generative scene models beyond multi-view geometry to multi-view radiometric ones, where phenomena such as lighting and attenuation can be accurately estimated.

1. Introduction

Reconstruction pipelines such as COLMAP [58] describe the camera with a few central projection models: Brown–Conrady [10] for perspective lenses, Kannala–Brandt [36] and Double Sphere [64] for fisheyes, and generic spline models [6, 57] where a parametric model falls short. These are the cameras of multiple-view geometry. They model where a scene point is imaged, but not how much light arrives there. All of them assume a pinhole aperture, *i.e*. that a single ray connects a scene point with the sensor. With a pinhole aperture every scene point is in focus and no light is lost toward the image border, so these models cannot represent defocus or vignetting.

In this paper, we make two major contributions, the *radiometric camera* and its *calibration*. The *first* is a model that lifts any central geometric camera model to a radiometric one. The lifted model keeps the projection of the geometric model and replaces the pinhole aperture by a finite aperture with a free-form focus surface and a generic vignetting function. Setting the aperture to a pinhole recovers the geometric model. To our knowledge, it is the first model that generalizes the central camera models in common use with the major radiometric components of a lens, its aperture, focus and vignetting, *and* that is calibrated in all of them jointly from images of a standard calibration target (Tab. 1).

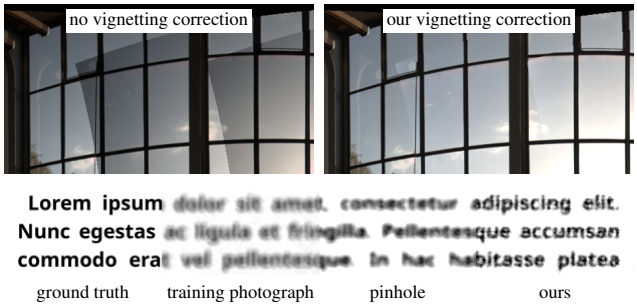


Figure 1. Our radiometric camera model can be applied in down stream tasks like vignetting correction in panorama stichting or texture learning in inverse rendering.

The radiometric camera unlocks more complex multi-view radiometric problems, in which several views jointly constrain radiometric unknowns such as the radiant intensity distribution and the pose of a light source [49, 56, 66, 69] or the attenuation of water [3, 12, 46, 59]. We aim to solve such problems generatively, *i.e*. by a maximum-likelihood estimate of the scene, which requires an image-formation model that predicts the measurements of the sensor from the scene. The camera is the last stage of this model. The irradiance at a sensor point is gathered over the cone of rays that the aperture admits, and where the lens is out of focus, these rays come from different scene points. Earlier radiometric extensions of geometric cameras relate each sensor value to the radiance of a single scene point, or add corrections to the image [29, 37, 54]. The difficulty in extending a geometric camera to radiometry is that the quantities to be added depend on exactly the components that the pinhole aperture removed. The *radiometric* camera

restores them, so that the irradiance becomes the integral over the whole aperture. We address its *calibration* by inverse rendering on the RAW photosites, whose values are proportional to the irradiance that they receive, from the same images of standard calibration targets as a geometric calibration. The constraints are not the incidences of detected features with their projections, but the relative radiant flux of the photosites, which the same target under the same lighting must explain in every view.

However, differentiating a camera with a finite aperture gives rise to boundary terms, from the aperture, the discontinuous edges of the calibration pattern and from visibility, which bias the gradients of standard differentiable renderers [71]. Moreover, the lighting, the reflectance and the shape of the target have to be estimated simultaneously without being confused with the camera's defocus and vignetting. We give a single formulation that resolves all of these difficulties. Neither a parameter-dependent discontinuity nor a parameter-dependent integration domain remains in it, the focus surface is parameterized so that its gradient does not vanish, the lighting varies over the target so that it separates from the vignetting, the gauge freedoms between the camera, the lighting and the target are fixed exactly, the shape of the target is a bending of the board with few parameters, and the linear structure of the model in the lighting and the vignetting gives a closed-form start. We use the calibration to validate the components of the model in three respects. *First*, they benefit the geometric accuracy. We show that the feature detections of widely used calibration tools are biased, and that the radiometric calibration, which uses none, lowers the reprojection error on a held-out high-precision target. *Second*, their estimated parameters represent the physical phenomena. The focus surface is learned as a plane for a wide-angle lens and as a hemisphere for a fisheye, and the vignetting function reproduces a flat field measured through a diffuser. *Third*, the calibrated camera solves a first multi-view radiometric problem: it recovers the unknown in-focus reflectance of a texture from defocused images.

2. Related Work

Lens models. The thin lens [30, 50] gives the perspective camera a finite aperture, so that a scene point off the plane of focus is imaged to a circle of confusion. It is tied to the perspective projection and has no natural counterpart in classic distortion models [10, 14, 36, 64]. The plane of focus and the vignetting that emerge from this model are too simple to capture the actual phenomena. The surface of best focus of a real lens is curved, which is known in optics as field curvature [38]. Tang and Kutulakos [61] derive the blur of a rotationally symmetric lens as an integral over its aperture and show that field curvature acts as a defocus that depends on the position in the image. This is what a focus surface describes, and it justifies our design, which replaces their fixed, symmetric dependence by a free-form one. Their analysis is theoretical, and the aberrations for which the rays of a sensor position no longer meet in a point are not part of our model. Models that trace the rays through the lens elements [39] need the optical design of the lens, which manufacturers rarely publish. The neural lens model of Xian *et al*. [70] replaces the distortion of a pinhole camera by an invertible network and models the vignetting as a gain per pixel of the image, without an aperture. The thin lens is also the only finite aperture of the differentiable renderer Mitsuba [33, 47]. Its cameras are perspective and without lens distortion, and the aperture radius and the focus distance of its thin lens are not differentiable parameters, so neither the projection of a general central model nor its aperture and focus can be calibrated with it. Our calibration keeps every camera parameter differentiable: neither a parameter-dependent discontinuity nor a parameter-dependent integration domain remains in its formulation.

	Proj.	Apert.	Focus	Vign.	Calib.
Thin lens [50], Mitsuba [33]	(✓)	✓	(✓)	(✓)	–
Lens-design ray tracing [39]	(✓)	✓	(✓)	(✓)	–
Tang and Kutulakos [61]	(✓)	✓	(✓)	(✓)	–
Kang and Weiss [35]	(✓)	–	–	(✓)	(✓)
Kim and Pollefeys [37]	–	–	–	✓	(✓)
Neural lens model [70]	✓	–	–	✓	(✓)
Rehder *et al*. [54]	(✓)	–	–	–	(✓)
Hannemose *et al*. [29]	(✓)	–	–	–	(✓)
Lens blur fields [41]	–	(✓)	–	–	(✓)
Radiometric camera (ours)	✓	✓	✓	✓	✓

Table 1. Prior camera models (top) and calibration methods (middle) against the radiometric camera: any central projection, a finite aperture, a free-form focus surface, a generic vignetting function, and the joint calibration of all of them from images. ✓: general; (✓): restricted, *e.g*. to the perspective projection, a plane of focus, the $\cos^4$ falloff or a subset of the parameters; –: not modeled. Section 2 gives the details.

Radiometric calibration and image corrections. Radiometric calibration methods estimate the vignetting, often together with the response function, from the pixel values of overlapping images [22, 37, 74]. In [37, 70] the vignetting is an empirical function of the image position, and each pixel value is related to the radiance of a single scene point. Reconstruction methods likewise add corrections around a geometric camera, such as per-image gains [11], appearance embeddings [44] or blur kernels [42]. These compensate for the unmodeled radiometry in the image, where defocus, which depends on the depth of the scene point, cannot be corrected consistently. The radiometric camera models it inside the camera, as an integral over its aperture.

Calibration from pixel values. Kang and Weiss [35] obtain a rough focal length of a perspective camera from the $\cos^4$ falloff in a single image of a uniform surface. Rehder *et al*. [54] formulate a maximum-likelihood calibration on the image intensities by rendering the target. They assume an evenly lit target and no vignetting, and take a single radiance sample per pixel, to which they attribute that their static calibration is no more precise than one from features; defocus, vignetting and uneven lighting are named as future work. Hannemose *et al*. [29] also render the target densely, with a pinhole camera with Brown–Conrady distortion, but their calibration is photometric and not radiometric: the rendering blurs the texture with a Gaussian whose width is a free parameter of each interest point, and it models neither the lighting nor the intensity scale of the image. A separate line calibrates the blur of a lens as a learned point spread function, apart from the projection of the camera [41]. In the radiometric camera, the defocus follows from the aperture and the focus surface of the same model that projects the scene, for any central projection, and it is calibrated together with the vignetting, the lighting and the target.

3. Radiometric Camera Model

Let $\mathcal{I} \subseteq \mathbb{R}^2$ be the sensor plane. A pixel $\boldsymbol{u}$ with the area $\mathcal{I}_{\boldsymbol{u}} \subseteq \mathcal{I}$ measures the radiant flux [50]

$$\Phi(\boldsymbol{u}) \coloneqq \int_{\mathcal{I}_{\boldsymbol{u}}} \int_{\Omega(\boldsymbol{p})} L(\boldsymbol{x}(\boldsymbol{p},\boldsymbol{\omega}), \boldsymbol{\omega}_o(\boldsymbol{p},\boldsymbol{\omega})) \cos\theta(\boldsymbol{\omega}) \,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{p}, \tag{1}$$

where $\Omega(\boldsymbol{p})$ is the solid angle under which the sensor point $\boldsymbol{p}$ sees the aperture, $\boldsymbol{x}(\boldsymbol{p},\boldsymbol{\omega})$ the scene point visible from $\boldsymbol{p}$ along the direction $\boldsymbol{\omega}$, $L(\boldsymbol{x},\boldsymbol{\omega}_o)$ the radiance that leaves $\boldsymbol{x}$ in the direction $\boldsymbol{\omega}_o$ towards the lens, and θ the angle of incidence on the sensor. The inner integral is the irradiance $E(\boldsymbol{p})$. All quantities are per color channel, and we omit the channel index (Sec. A.1 gives the spectral formulation). The radiance L follows from the lighting and the reflectance of the scene (Eq. (8)). The lens model determines $\Omega(\boldsymbol{p})$, $\boldsymbol{x}(\boldsymbol{p},\boldsymbol{\omega})$ and $\boldsymbol{\omega}_o(\boldsymbol{p},\boldsymbol{\omega})$. With a pinhole aperture, $\Omega(\boldsymbol{p})$ is a single direction, so every scene point is in focus and there is no vignetting (Sec. A.2).

3.1. Projection

Let $\boldsymbol{\pi}^{-1} : \mathcal{I} \to \mathcal{S}^2$ be the back-projection of a central geometric camera model with the center of projection $\boldsymbol{o}$. It maps a sensor point $\boldsymbol{p}$ to the unit direction $\hat{\boldsymbol{d}}(\boldsymbol{p}) \coloneqq \boldsymbol{\pi}^{-1}(\boldsymbol{p})$ of its ray. For a model that is given by its forward projection, we invert the projection numerically per sample. The radiometric camera keeps $\boldsymbol{\pi}^{-1}$ as it is.

3.2. Aperture and Focus Surface

The thin lens [50] replaces the pinhole by an aperture disc $\mathcal{A}$ of radius r, centered at $\boldsymbol{o}$ and parallel to the sensor. All rays

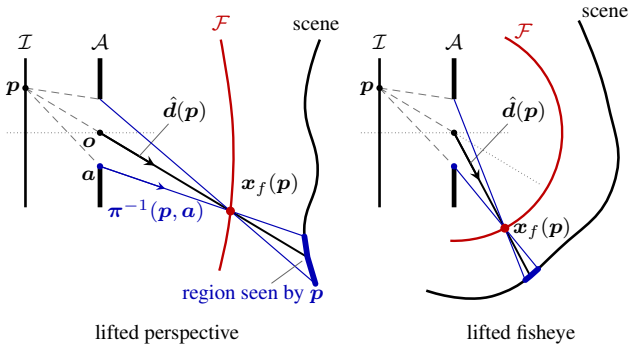


Figure 2. The radiometric camera lifts any central geometric model, here a perspective and a fisheye model, in a cross-section. The geometric model alone sets the direction $\hat{\boldsymbol{d}}(\boldsymbol{p})$ in which the ray of the sensor point $\boldsymbol{p}$ leaves the center $\boldsymbol{o}$ of the aperture $\mathcal{A}$; the model does not trace the light inside the camera (dashed). The lift adds the aperture: the rays through all aperture points $\boldsymbol{a}$ meet in the point of focus $\boldsymbol{x}_f(\boldsymbol{p})$ on the focus surface $\mathcal{F}$, which is close to a plane for the perspective model and surrounds the camera for the fisheye. Where the scene is off $\mathcal{F}$, the cone of these rays meets it in a region (thick blue), from which $\boldsymbol{p}$ collects its light.

from a sensor point $\boldsymbol{p}$ through $\mathcal{A}$ meet in one point $\boldsymbol{x}_f(\boldsymbol{p})$ on the plane of focus. Where the scene does not pass through $\boldsymbol{x}_f(\boldsymbol{p})$, the cone of these rays intersects it in a region and not in a point, and $\boldsymbol{p}$ collects the light of this whole region, which defocuses the image. For a real lens, the points of focus form a surface $\mathcal{F}$ (Fig. 2) that is not a plane in general [30]: it is approximately a plane for a thin-lens-like lens and a sphere for a fisheye [16, 38]. We therefore learn the focus surface. We represent it by the inverse distance $\tau(\boldsymbol{p})$ from $\boldsymbol{o}$ to the point of focus $\boldsymbol{x}_f(\boldsymbol{p}) \in \mathcal{F}$ along the ray of $\boldsymbol{p}$,

$$\boldsymbol{x}_f(\boldsymbol{p}) \coloneqq \boldsymbol{o} + \frac{1}{\tau(\boldsymbol{p})}\,\hat{\boldsymbol{d}}(\boldsymbol{p}), \tag{2}$$

and extend the back-projection to the aperture (Fig. 2). The ray from $\boldsymbol{p}$ through an aperture point $\boldsymbol{a} \in \mathcal{A}$ is directed from $\boldsymbol{a}$ towards $\boldsymbol{x}_f(\boldsymbol{p})$:

$$\boldsymbol{\pi}^{-1}(\boldsymbol{p},\boldsymbol{a}) \coloneqq \frac{\boldsymbol{x}_f(\boldsymbol{p}) - \boldsymbol{a}}{\|\boldsymbol{x}_f(\boldsymbol{p}) - \boldsymbol{a}\|} = \frac{\tau(\boldsymbol{p})(\boldsymbol{o}-\boldsymbol{a}) + \hat{\boldsymbol{d}}(\boldsymbol{p})}{\|\tau(\boldsymbol{p})(\boldsymbol{o}-\boldsymbol{a}) + \hat{\boldsymbol{d}}(\boldsymbol{p})\|}. \tag{3}$$

This is the lift of the geometric model. Setting the aperture $\mathcal{A} = \{\boldsymbol{o}\}$ reduces the lifted model back to the geometric one, since $\boldsymbol{\pi}^{-1}(\boldsymbol{p},\boldsymbol{o}) = \boldsymbol{\pi}^{-1}(\boldsymbol{p})$. Where the latter back-projects to a single ray, $\boldsymbol{\pi}^{-1}(\boldsymbol{p},\mathcal{A})$ back-projects to a cone of rays through $\boldsymbol{x}_f(\boldsymbol{p})$. Similarly, with the perspective back-projection and a plane of focus, the model reduces to the thin lens (Sec. A.3). The inverse distance avoids intersecting the rays with $\mathcal{F}$, and unlike the distance itself, it keeps the gradient from vanishing for a far focus [22, 45] (Sec. A.4). We model τ by a thin plate spline [7] over a grid restricted to the illuminated region of the sensor.

3.3. Vignetting

Written as an integral over the aperture, the flux of the thin lens carries the factor $\cos^4\theta/s'^2$ that expresses its vignetting, with the image distance s' [28] (Sec. A.5). Real lenses differ much from it (Fig. 6). Hence, we replace this factor by a generic vignetting function V. We also normalize by the aperture area πr^2, so that the aperture radius changes the defocus but not the brightness of the image, such that its scale is carried by the lighting (see Sec. 4.2, A.9) alone. To spare us computing a boundary integral (see Sec. 4.1), we substitute the aperture point $\boldsymbol{a} = \boldsymbol{o} + r\boldsymbol{\xi}$ with $\boldsymbol{\xi}$ in the unit disc $\mathcal{D}$ of the aperture plane. The area element $\mathrm{d}\boldsymbol{a} = r^2\,\mathrm{d}\boldsymbol{\xi}$ cancels the r^2 of the normalization, and the radiometric camera measures

$$\Phi(\boldsymbol{u}) = \frac{1}{\pi}\int_{\mathcal{I}_{\boldsymbol{u}}} V(\boldsymbol{p}) \int_{\mathcal{D}} L\big(\boldsymbol{x}(\boldsymbol{p},\boldsymbol{a}), -\boldsymbol{\pi}^{-1}(\boldsymbol{p},\boldsymbol{a})\big)\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{p}, \tag{4}$$

where $\boldsymbol{x}(\boldsymbol{p},\boldsymbol{a})$ is the scene point that the ray from $\boldsymbol{a}$ in the direction $\boldsymbol{\pi}^{-1}(\boldsymbol{p},\boldsymbol{a})$ meets. The radius now appears only in the position of the aperture point, and no longer in the domain of the integral. We assume a vignetting that is radially symmetric about the principal point $\boldsymbol{c} \in \mathcal{I}$ and represent it by a cubic B-spline [51]

$$V(\boldsymbol{p}) \coloneqq \sum_{i=0}^{n} N_{i,3}\Big(\frac{\|\boldsymbol{p}-\boldsymbol{c}\|}{r_{\max} + \|\boldsymbol{p}-\boldsymbol{c}\|}\Big) P_i, \tag{5}$$

with the basis functions $N_{i,3}$, the control points $\boldsymbol{\Theta}_v = \{P_i\}_{i\in\{0,\dots,n\}}$, $n = 15$, and the largest distance $r_{\max}$ from the principal point to an illuminated pixel. Its knots are uniform in the image (Sec. A.6).

4. Calibration by Inverse Rendering

We calibrate the radiometric camera by inverse rendering: we render the images of a calibration target with Eq. (4) and fit them to the captured RAW images. The rendered flux depends on the scene as well as on the camera, so we estimate jointly the poses $\boldsymbol{\Theta}_e$, the projection $\boldsymbol{\Theta}_d$, the focus surface $\boldsymbol{\Theta}_f$, the aperture radius r and the vignetting $\boldsymbol{\Theta}_v$ of the camera, the lighting $\boldsymbol{\Theta}_l$, the material $\boldsymbol{\Theta}_m$ and the shape $\boldsymbol{\Theta}_b$ of the target. We denote all of them by $\bar{\boldsymbol{\Theta}}$. Let $\phi_j(\boldsymbol{u})$ be the value of the RAW photosite $\boldsymbol{u}$ in view j and $\sigma_j(\boldsymbol{u})$ its standard deviation under the Poisson–Gaussian noise model of the sensor. We minimize

$$\sum_j \sum_{\boldsymbol{u}\in\mathcal{S}_j} \Big(\frac{\phi_j(\boldsymbol{u}) - \Phi_j(\boldsymbol{u};\bar{\boldsymbol{\Theta}})}{\sigma_j(\boldsymbol{u})}\Big)^2 + \lambda\big\|\mathrm{D}^3(\hat{\boldsymbol{d}} - \hat{\boldsymbol{d}}^0)\big\|^2 \tag{6}$$

over $\bar{\boldsymbol{\Theta}}$, where $\mathcal{S}_j$ are photosites drawn uniformly from those of view j that see the target. The first term is the negative log-likelihood of the photosites, so that for $\lambda = 0$ the minimizer is the maximum-likelihood estimate of the camera, the poses and the scene. The second term regularizes the central spline model only (Sec. 4.5). The remainder of this section gives what the objective requires: derivatives without boundary terms, a lighting model, the reflectance and the shape of the target, and a initialization value.

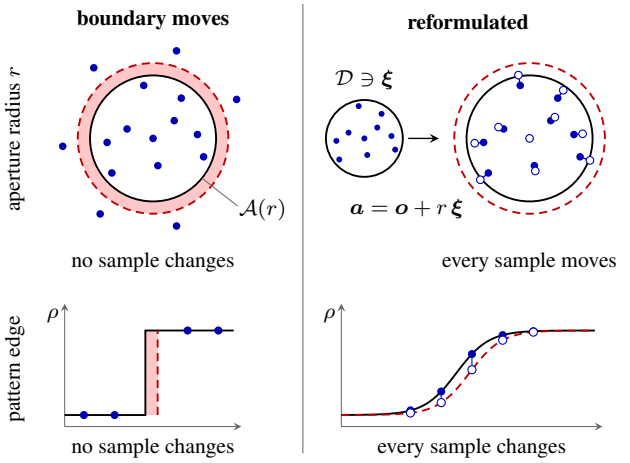


Figure 3. Fixed samples (dots) miss a boundary that moves with a parameter (solid black to dashed red): none lies in the region that it sweeps, so the sampled gradient lacks its contribution (left). We remove both boundaries of the calibration (right). *Top:* the samples of the aperture are placed from the fixed unit disc $\mathcal{D}$ and move with its radius r. *Bottom:* the blur of the print makes the albedo ρ smooth across an edge of the pattern. Hollow dots: after the change.

4.1. Differentiation without Boundary Terms

The derivative of an integral whose domain, or whose integrand's discontinuities, depend on a parameter has a boundary term, which automatic differentiation of a sampled estimate misses (Fig. 3, left). Differentiable renderers estimate such terms by sampling the boundaries [40, 71, 73]. In the calibration, three of them arise: The changing aperture radius, border of the calibration target, and sharp texture edges. We avoid these as follows. The change of aperture radius is addressed by integrating over the fixed unit disc as described above in Sec. 3.3. Second, the hard edges of the calibration pattern are discontinuities of the albedo that move over the aperture and over the pixel with the camera, the pose and the target shape. We model the print as its design convolved with a Gaussian, a 10–70 % edge width of 50 μm [55], so that the albedo is smooth. The boundary terms of a general scene, from silhouettes, shadows and the visibility at every bounce of a ray (Sec. A.7), do not arise in the calibration: we only render the masked target $\mathcal{M} \subseteq \mathcal{I}$, which is considered smoothly illuminated within a constant scene.

4.2. Lighting

We represent the light of the environment by spherical harmonics of degree two [1, 53], $L^{\mathrm{SH}}(\boldsymbol{\omega}; \boldsymbol{\Theta}_l)$, with the coefficients $\boldsymbol{\Theta}_l$ per color channel (Sec. A.8). A calibration target is planar, so every point of it sees the same hemisphere, and a lighting at infinite distance gives the same irradiance at every point of the target: any shading across the target would have to be explained by the vignetting. We therefore place the lighting on a sphere of finite radius r_{SH} about the center $\boldsymbol{b}_0$ of the target, the larger of the diagonal of the target and the largest distance of a camera from $\boldsymbol{b}_0$ at the start. The radiance that arrives at $\boldsymbol{x}$ from the direction $\boldsymbol{\omega}$ is L^{SH} in the direction

$$\hat{\boldsymbol{\omega}}(\boldsymbol{x}, \boldsymbol{\omega}) \coloneqq \frac{\boldsymbol{x} + \gamma\boldsymbol{\omega} - \boldsymbol{b}_0}{\|\boldsymbol{x} + \gamma\boldsymbol{\omega} - \boldsymbol{b}_0\|} \tag{7}$$

from $\boldsymbol{b}_0$ to the point where the ray from $\boldsymbol{x}$ along $\boldsymbol{\omega}$ meets the sphere, with $\gamma > 0$ such that $\|\boldsymbol{x} + \gamma\boldsymbol{\omega} - \boldsymbol{b}_0\| = r_{\mathrm{SH}}$. The lighting then varies over the target and separates from the vignetting.

4.3. Reflectance

Calibration targets usually consist of a pattern printed on background material. To model its appearance properly, we estimate the reflectance distribution of these materials. The radiance that leaves a scene point $\boldsymbol{x}$ in the direction $\boldsymbol{\omega}_o$ is [50]

$$\begin{aligned} &L(\boldsymbol{x}, \boldsymbol{\omega}_o) \\ &\coloneqq \int_{\mathcal{H}^2(\boldsymbol{x})} f(\boldsymbol{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) L(\boldsymbol{x}'(\boldsymbol{x}, \boldsymbol{\omega}_i), -\boldsymbol{\omega}_i) \cos\theta_i \,\mathrm{d}\boldsymbol{\omega}_i, \end{aligned} \tag{8}$$

where $\mathcal{H}^2(\boldsymbol{x})$ is the upper hemisphere over $\boldsymbol{x}$, $\boldsymbol{x}'(\boldsymbol{x}, \boldsymbol{\omega}_i)$ the scene or light point visible from $\boldsymbol{x}$ along $\boldsymbol{\omega}_i$, θ_i the angle between $\boldsymbol{\omega}_i$ and the surface normal, and f the bidirectional scattering distribution function. On the target, the incident radiance is the lighting of Sec. 4.2. For f, we use the Cook–Torrance model [15] with an isotropic GGX microfacet distribution of roughness α [31] and the Fresnel term of a dielectric with the index of refraction η, and a weight $\mu \in [0, 1]$ between its diffuse part $\rho(\boldsymbol{x})/\pi$, with the albedo ρ of the print, and its specular part. We evaluate the integral by Monte Carlo sampling of the visible microfacet normals [19] (Sec. A.8). Ink fills the microfacets of the print, so that inked areas are more specular than blank ones. We therefore blend the roughness and the index of refraction between a value for ink and one for the blank print by the albedo, $\boldsymbol{\Theta}_m = \{\alpha_{\neg\mathrm{ink}}, \alpha_{\mathrm{ink}}, \eta_{\neg\mathrm{ink}}, \eta_{\mathrm{ink}}, \mu\}$.

4.4. Target Shape

A printed target is not exactly planar, and its print is not exactly its design. We chart the board by $\boldsymbol{b} \in \mathcal{B}$, in units of its width, and model its shape by a height field $h : \mathcal{B} \to \mathbb{R}$ along the board normal $\boldsymbol{n}_{\mathcal{B}}$. In a view with the pose $(\mathrm{R}, \boldsymbol{t}) \in \boldsymbol{\Theta}_e$, the chart point $\boldsymbol{b}$ is the scene point

$$\boldsymbol{s}(\boldsymbol{b}) \coloneqq \mathrm{R}\big(\boldsymbol{b}_{\mathcal{B}} + h(\boldsymbol{b})\, \boldsymbol{n}_{\mathcal{B}}\big) + \boldsymbol{t}, \tag{9}$$

where $\boldsymbol{b}_{\mathcal{B}}$ is the point of $\boldsymbol{b}$ on the plane of the board. The height field is a thin plate spline [7] through the heights $\boldsymbol{h} \in \mathbb{R}^{25}$ at a 5×5 grid of nodes. A thin plate spline is the sum of an affine function and a bending, and affine heights, an offset and two tilts, describe a plane, which is already represented by the pose. We therefore restrict the heights to the pure bendings, $\boldsymbol{h} = \mathrm{B}\boldsymbol{y}$, where the columns of $\mathrm{B} \in \mathbb{R}^{25\times 22}$ span the nodal heights whose spline has no affine part, and optimize $\boldsymbol{\Theta}_b \coloneqq \boldsymbol{y}$. The shape then has no gauge freedom with the poses, without a penalty or a fixed node (Sec. A.9), and only 22 parameters. For the spline camera, we also estimate the in-plane error of the print by an aspect and a skew, at which the albedo is read.

The scene point of Eq. (4) is the intersection of the ray from $\boldsymbol{a}$ in the direction $\boldsymbol{d} \coloneqq \boldsymbol{\pi}^{-1}(\boldsymbol{p}, \boldsymbol{a})$ with this surface. We solve for its chart point: with an orthonormal pair $\boldsymbol{e}_1, \boldsymbol{e}_2$ perpendicular to $\boldsymbol{d}$,

$$\boldsymbol{g}(\boldsymbol{b}) \coloneqq \begin{pmatrix} \boldsymbol{e}_1^\top \\ \boldsymbol{e}_2^\top \end{pmatrix} \big(\boldsymbol{s}(\boldsymbol{b}) - \boldsymbol{a}\big) = \boldsymbol{0}, \tag{10}$$

a root of two equations in two unknowns that we find by Levenberg–Marquardt iterations from the intersection with the plane of the board. It gives $\boldsymbol{x}(\boldsymbol{p}, \boldsymbol{a}) = \boldsymbol{s}(\boldsymbol{b})$ together with the chart point $\boldsymbol{b}$ at which the albedo is read. The root is differentiated w.r.t. the camera, the pose and the heights by the implicit function theorem, $\partial\boldsymbol{b}/\partial\theta = -(\partial\boldsymbol{g}/\partial\boldsymbol{b})^{-1}\, \partial\boldsymbol{g}/\partial\theta$, and not through its iterations. The effect of the shape on the reflectance is small, so the reflectance uses the normal of the plane.

4.5. Initialization and Optimization

The projection and the poses start from a geometric calibration (Sec. A.12), the target as a plane, the focus surface as the plane parallel to the sensor at the mean distance of the target, and the aperture radius as $f/(2N)$ with the focal length f and the f-number N of the lens. The flux is linear in the vignetting and in the lighting and affine in μ (Sec. A.10), so they start from one pass of linear least-squares solves: the lighting with $V \equiv 1$, then the vignetting, then μ. We minimize Eq. (6) with the Levenberg–Marquardt method [43], with $\lambda = 0$ for the parametric camera models. The nodes of the central spline at the border of the lens coverage are hardly constrained by the photosites. For this model only, the second term regularizes the change of the nodal directions $\hat{\boldsymbol{d}} \subseteq \bar{\boldsymbol{\Theta}}$ from their start $\hat{\boldsymbol{d}}^0$ with a P-spline penalty [21], where D^3 takes the third finite differences, the discrete change of curvature, along both axes of the grid and $\lambda = 0.1$ relative to the data term. It vanishes for

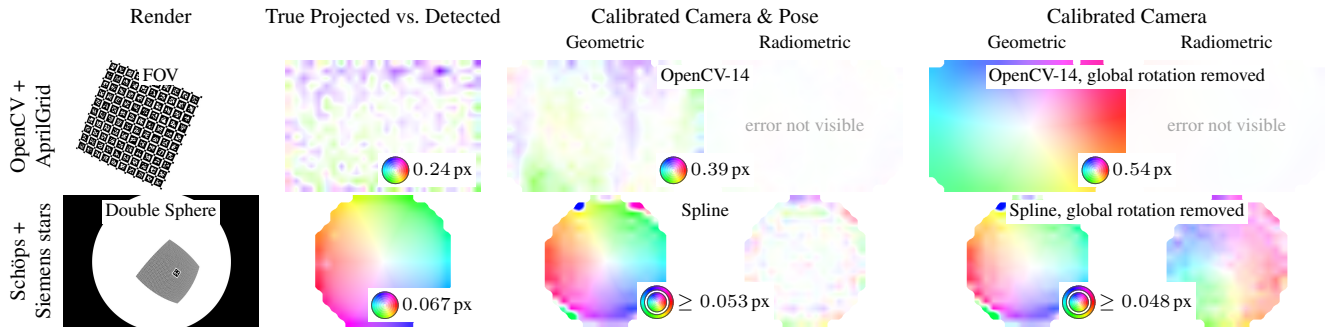


Figure 4. The bias of a detector (column 2) is absorbed by a geometrically calibrated camera (columns 3 and 4). With the calibrated poses, the bias is pushed into the camera model (column 3); with the true poses (column 4), it shows that the poses compensate for part of it. Our radiometric refinement does not rely on a detector and removes most of the bias in both evaluations.

a locally quadratic change, so the penalty leaves a smooth change of the camera free and suppresses its oscillation, and the border nodes follow the interior. No other component is regularized. The parameters stay in their physical ranges by their parameterization: we optimize $\log r$, $\log \alpha$, $\log(\eta - 1)$ and the logit of μ, the poses on $SE(3)$ [60], and the inverse focus distances are kept non-negative. The gauge freedoms are removed by the parameterization (Sec. A.9).

5. Experiments

We show that geometric calibration is biased by its feature detections (Sec. 5.1) and that the radiometric camera improves the geometric accuracy (Sec. 5.2), compare the learned vignetting with a measured flat field (Sec. 5.3) and validate the focus model (Sec. 5.4). All captures are RAW images of a Sony ILME-FX3 (4256×2848 pixels) and are listed in Sec. A.11 together with the runtime for each experiment; all errors are in pixels of this image.

5.1. Detector Bias

Feature detectors introduce a measurement bias that degrades the regressed camera model. We rendered 10,000 images each of the AprilGrid [26, 48] with an FOV camera model [18], which has a mild distortion, and of the Siemens-star target of Schöps *et al*. [57] with a Double Sphere model [64], which has a strong fisheye distortion, in evenly lit calibration scenes. The second column of Fig. 4 shows the displacement between the true projection of each feature and its detection, for the AprilGrid detector of OpenCV [9] (top) and the Siemens-star detector of Schöps *et al*. (bottom). Both detectors exhibit a bias that grows toward the image border. The third column projects the features with the camera and the poses of a calibration. In the geometric calibration, OpenCV-14 for the AprilGrid and the central spline for the Siemens stars, the detection bias is absorbed by both the camera and the poses, whereas the radiometric refinement largely removes it. The fourth column substitutes the true poses for the estimated ones, which isolates the camera model. This column is the relevant one: the poses are discarded after the calibration, and a downstream task depends on the camera alone. The geometrically calibrated camera has absorbed a strong bias from the feature detections, also for the highly flexible spline of Schöps *et al*., and our radiometric refinement removes most of it.

5.2. Geometric Accuracy

We lift four commonly used central parametric camera models, Brown–Conrady [10], the 14-parameter model of OpenCV [9], Kannala–Brandt [36] and Double Sphere [64], and the central generic spline model of Schöps *et al*. [57] to the radiometric camera. We calibrate them on three lenses chosen for their diverse fields of view and distortions (Tab. A1): a Tamron 20 mm wide-angle lens with barrel distortion (94° diagonal field of view), a Sigma 14 mm rectilinear wide-angle lens (114°) and a Sigma 8 mm circular fisheye (180°). We captured two standard calibration targets, the Siemens-star target of Schöps *et al*. [57] and an AprilGrid [26], each $1\,\mathrm{m} \times 1\,\mathrm{m}$ and printed directly on Alu-Dibond, at $f/22$ (the fisheye's Siemens stars at $f/32$) and ISO 100, to minimize defocus and thus give optimal conditions to the geometric camera models, which do not model defocus. Each model is started from the geometric calibration of its native tool where the tool supports the model and the target, and from our own estimator otherwise (Sec. A.12). The radiometric refinement samples 8,192 photosites per image for the parametric models and 16,384 for the spline.

The test set consists of backlit images of the Thorlabs R3L3S7N1 high-precision distortion target [62], captured by rotating the camera about a pivot, which reduces the poses to one rotation per view. Every dot is evaluated with a pose that it did not help to fit (Sec. A.13).

Table 2 compares the reprojection error of each model at its geometric start and after the radiometric refinement on the Thorlabs test set (Tab. A2). We call a change significant

Model	Tamron 20 mm		Sigma 14 mm		Sigma 8 mm fisheye	
	AprilGrid	Siemens stars	AprilGrid	Siemens stars	AprilGrid	Siemens stars
Median reprojection error [px]						
Brown–Conrady	0.119 → **0.067**	0.288 → **0.135**	0.499 → **0.377**	0.687 → **0.359**	–	–
OpenCV-14	0.107 → **0.060**	0.308 → **0.131**	0.476 → **0.359**	0.600 → **0.384**	1.98 → **1.70**	0.123 → **0.079**
Kannala–Brandt	0.117 → **0.106**	0.183 → 0.176	0.456 → **0.351**	0.567 → **0.329**	2.00 → **1.52**	0.113 → **0.063**
Double Sphere	0.261 → *0.286*	0.219 → *0.268*	1.42 → 1.34	0.734 → **0.461**	2.09 → **1.49**	0.109 → **0.067**
Central spline	0.040 → *0.048*	0.161 → **0.136**	0.479 → **0.331**	0.504 → **0.366**	1.99 → **1.45**	0.121 → **0.100**
95th percentile reprojection error [px]						
Brown–Conrady	0.356 → **0.173**	0.855 → **0.321**	2.07 → **1.78**	2.73 → **1.72**	–	–
OpenCV-14	0.373 → **0.155**	0.969 → **0.337**	2.16 → **1.74**	2.87 → **1.93**	9.15 → **5.68**	0.818 → **0.440**
Kannala–Brandt	0.446 → **0.285**	0.730 → **0.515**	1.99 → **1.79**	2.37 → **1.75**	6.80 → **5.43**	0.590 → **0.254**
Double Sphere	1.70 → **1.16**	0.717 → **0.625**	5.13 → **4.42**	2.42 → **1.94**	7.41 → **5.45**	0.591 → **0.289**
Central spline	0.131 → **0.119**	0.436 → **0.331**	2.45 → **1.76**	2.35 → **1.97**	7.58 → **5.51**	0.996 → **0.538**

Table 2. Reprojection error on the held-out high-precision target, geometric start → radiometric refinement, for each camera model, lens and calibration target. The refined value is bold where it is significantly lower than the start and italic where it is significantly higher (95 % bootstrap interval of the difference, Tab. A4); the mean and the RMS are in Tab. A5. –: Brown–Conrady cannot represent the 180° field of view of the fisheye.

where the 95 % bootstrap interval of the paired difference excludes zero (Sec. A.13). The radiometric refinement lowers the median reprojection error significantly in 82 % of the 28 combinations of model, lens and target (23, by 9 to 57 %), raises it significantly in 11 % (three, all on the Tamron 20 mm), and lowers the 95th percentile significantly in all of them, by 9 to 65 %. On the Sigma 14 mm and on the fisheye, no model is significantly worse in any statistic. Figure 5 follows the two refinements of the central spline on the Sigma 14 mm over their iterations. In both, the error on the held-out target falls throughout, while the error with respect to the detections that the geometric calibration was fitted to rises after the first four iterations (from 0.59 to 0.71 px on the Siemens stars and from 0.56 to 0.68 px on the AprilGrid). The test error decreases while the training error increases, which shows the bias: the refinement moves the camera away from the biased detections that its start was fitted to. We attribute part of this bias to the freedom of the geometric calibration: The feature wise parametrization of the target in [57] allows to absorb the bias, while our lower parameter parametrization (Sec. 4.4) cannot.

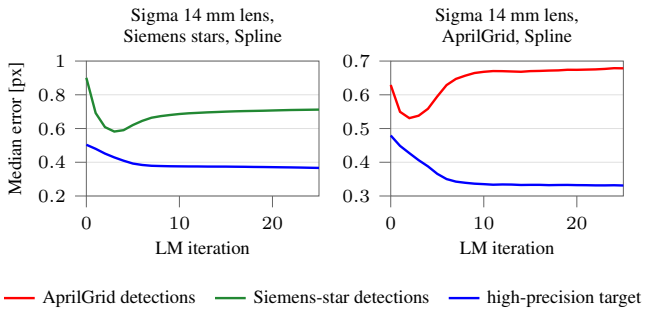


Figure 5. Median reprojection error during the radiometric refinement of the central spline on the Sigma 14 mm, with respect to the detections that the geometric calibration was fitted to and on the held-out high-precision target. After the first iterations the curves diverge in both panels: the error on the detections rises while the error on the held-out target falls, which shows that the geometric calibration had absorbed the bias of its detector.

5.3. Vignetting

We use the three lenses of Sec. 5.2 and the spline model with Siemens-star captures at a large aperture ($f/4$; $f/2.8$ for the Tamron) for stronger vignetting, and one flat-field image per lens (Tab. A2): the lens, with its hood on, covered by a white acrylic plate and lit by diffuse daylight through a light diffuser. Figure 6 compares, along the direction of the sensor diagonal through the principal point, the learned vignetting, a B-spline (Sec. 3.3), with the flat field, read off the green photosites, and with the $\cos^4$ falloff of the thin lens. All profiles are normalized at the principal point. The learned vignetting follows the flat field for all three lenses:

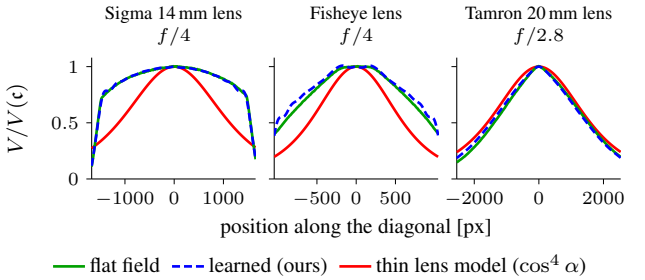


Figure 6. Normalized vignetting along the direction of the sensor diagonal through the principal point: the $\cos^4$ falloff of the thin lens and our learned vignetting against the measured flat field.

its RMS deviation along this line, in units of the value at the principal point, is 0.009 for the 14 mm lens, 0.042 for the

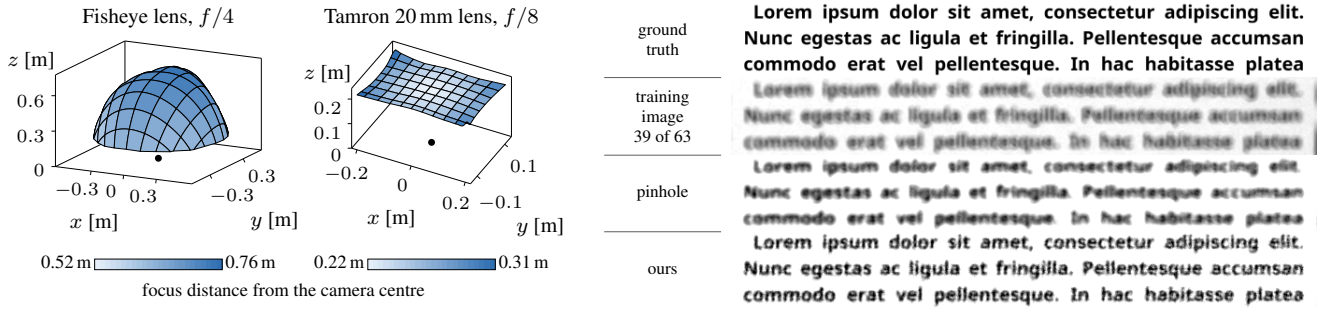


Figure 7. The focus model. *Left:* the learned focus surfaces, a hemisphere for the fisheye at $f/4$ and a plane for the Tamron 20 mm at $f/8$. *Right:* a texture estimated from defocused images. The radiometric camera recovers sharp text, whereas the pinhole camera has to put the blur of the images into the texture.

fisheye and 0.025 for the Tamron, against 0.244, 0.199 and 0.059 for the $\cos^4$ falloff.

5.4. Defocus

We validate the focus model in three ways. First, Fig. 7 (left) shows the focus surfaces learned with the spline model for the Tamron 20 mm at $f/8$ and for the fisheye at $f/4$: a plane for the wide-angle lens and a hemisphere for the fisheye, as expected for these lens types [16, 38].

Second, we solve a first multi-view radiometric problem with the calibrated camera: we estimate the unknown reflectance of a surface from many defocused views of it by inverse rendering. With a pinhole aperture, the estimated texture has to explain the defocus of the images and is blurred. We calibrate the Sigma 14 mm at $f/2.8$ on Siemens stars, once with the aperture and the focus surface and once as a pinhole, and estimate the texture of a board with printed text from 63 defocused images with each camera. Figure 7 (right) compares both textures with the printed design: the radiometric camera recovers sharp text where the pinhole camera returns the blur of the images. Since the calibration works with relative irradiance, a texture is recovered only up to a scale and an offset, which we fit once per texture before comparing it with the design. With the design ranging from 0 to 1, the RMS difference to it is 0.177 for the radiometric camera and 0.235 for the pinhole, and it is lower for the radiometric camera on 99 % of the page.

Third, we ablate the focus model on two lenses with visible defocus, the Tamron 20 mm at $f/8$ with the spline model and a 24–70 mm zoom lens at 48 mm and $f/10$ with Brown–Conrady, by refining the same geometric start once with the aperture and the focus surface and once as a pinhole. Table 3 shows that modeling the aperture and the focus surface lowers the median reprojection error on the Thorlabs target by 25 % and 16 % compared with the pinhole refinement, both significantly; on the Tamron, the pinhole refinement is even worse than the geometric start.

Refinement	**Median [px]**	**95th perc. [px]**
Tamron 20 mm, $f/8$, central spline		
Geometric start	0.160	0.499
Pinhole	*0.199*	*0.534*
Aperture and focus	**0.149**	**0.410**
− pinhole	-0.051 $[-0.054, -0.047]$	-0.124 $[-0.139, -0.103]$
Zoom at 48 mm, $f/10$, Brown–Conrady		
Geometric start	0.068	0.194
Pinhole	**0.050**	**0.150**
Aperture and focus	**0.042**	**0.117**
− pinhole	-0.008 $[-0.013, -0.003]$	-0.032 $[-0.047, -0.017]$

Table 3. Defocus ablation on the held-out target: the refinement as a pinhole and with the aperture and the focus surface, and their difference with its 95 % bootstrap interval. Bold and italic as in Tab. 2.

6. Conclusion

We presented the radiometric camera, which lifts any central geometric camera model to a radiometric one with a finite aperture, a free-form focus surface and a generic vignetting function. We also presented its calibration from the RAW photosites of ordinary calibration images, formulated without parameter-dependent discontinuities or integration domains and with every gauge freedom fixed exactly. To our knowledge, it is the first model to unite aperture, focus and vignetting with any central projection and to calibrate all of them jointly from the images of a geometric calibration (Tab. 1). The calibrated cameras have a lower reprojection error on a held-out high-precision target than their geometric starts for most models and lenses, their vignetting reproduces a measured flat field, and their focus surface is learned as a plane for a wide-angle lens and as a hemisphere for a fisheye. In a first multi-view radiometric reconstruction, they recover the in-focus reflectance of a texture from defocused images. The estimation of light sources and of the attenuation of water is future work. We will release the code, the data and the configurations of every experiment.

References

[1] Peter-Pike Sloan , Jan Kautz , and John Snyder. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *ACM SIGGRAPH*, pages 527–536. ACM, 2002. 5

[2] Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6723–6732, 2018. 1

[3] Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[4] Derya Akkaynak, Tali Treibitz, Tom Shlesinger, Yossi Loya, Raz Tamir, and David Iluz. What is the space of attenuation coefficients in underwater computer vision? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4931–4940, 2017. 1

[5] Ronen Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:218–233, 2003. 5

[6] Johannes Beck and Christoph Stiller. Generalized b-spline camera model. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 2137–2142, 2018. 1

[7] F.L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989. 3, 5

[8] Ronald N. Bracewell. The fourier transform and its applications. 1966. 1

[9] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 6, 2, 7

[10] Dean P. Brown. Decentering distortion of lenses. 1966. 1, 2, 6

[11] Matthew Brown and David G. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007. 2

[12] Mitch Bryson, Matthew Johnson-Roberson, Oscar Pizarro, and Stefan B. Williams. True color correction of autonomous underwater vehicle imagery. *Journal of Field Robotics*, 33 (6):853–874, 2016. 1

[13] James Casey. On the derivation of jump conditions in continuum mechanics. *International Journal of Structural Changes in Solids – Mechanics and Applications*, (2). Dedicated to Professor K. R. Rajagopal with Warmth and Admiration. 5

[14] A. E. Conrady. Decentred lens-systems. *Monthly Notices of the Royal Astronomical Society*, 79:384–390, 1919. 2

[15] R. L. Cook and K. E. Torrance. A reflectance model for computer graphics. *ACM Trans. Graph.*, 1(1):7–24, 1982. 5, 6

[16] Donald G. Dansereau, Glenn Schuster, Joseph Ford, and Gordon Wetzstein. A wide-field-of-view monocentric light field camera. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3757–3766, 2017. 3, 8

[17] Maryam Mohammadzadeh Darrodi, Graham Finlayson, Teresa Goodman, and Michal Mackiewicz. Reference data set for camera spectral sensitivity estimation. *J. Opt. Soc. Am. A*, 32(3):381–391, 2015. 1

[18] Frédéric Devernay and Olivier Faugeras. Straight lines have to be straight. *Machine Vision and Applications*, 13(1):14–24, 2001. 6

[19] Jonathan Dupuy and Anis Benyoub. Sampling visible ggx normals with spherical caps, 2023. 5, 6

[20] Bradley Efron. Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–185, 1987. 9

[21] Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121, 1996. 5

[22] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):611–625, 2018. 2, 3, 4

[23] Graham D. Finlayson, Roshanak Zakizadeh, and Arjan Gijsenij. The reproduction angular error for evaluating the performance of illuminant estimation algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(7):1482–1488, 2017. 1

[24] Harley Flanders. Differentiation under the integral sign. *The American Mathematical Monthly*, 80(6):615–627, 1973. 5

[25] Wolfgang Förstner and Eberhard Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, 1987. 7

[26] Paul Furgale, Joern Rehder, and Roland Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1280–1286, 2013. 6, 7

[27] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco J. Madrid-Cuevas, and Manuel J. Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280–2292, 2014. 7

[28] Thomas Goossens, Bert Geelen, Andy Lambrechts, and Chris Van Hoof. Vignetted-aperture correction for spectral cameras with integrated thin-film fabry–perot filters. *Applied Optics*, 58(7):1789, 2019. 4

[29] Morten Hannemose, Jakob Wilm, and Jeppe R. Frisvad. Superaccurate camera calibration via inverse rendering. In *Modeling Aspects in Optical Metrology VII*, page 40. SPIE, 2019. 1, 2, 3

[30] Eugene Hecht. *Optik*. De Gruyter Oldenbourg, Berlin, Boston, 2023. 2, 3

[31] Eric Heitz. Understanding the masking-shadowing function in microfacet-based brdfs. *Journal of Computer Graphics Techniques (JCGT)*, 3(2):48–107, 2014. 5, 6

[32] Harro Heuser. *Lehrbuch der Analysis*. Vieweg+Teubner Verlag, Wiesbaden, 13 edition, 2004. 1

[33] Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. Mitsuba 3 renderer. https://mitsuba-renderer.org, 2022. Version 3.1.1. 2

[34] Jun Jiang, Dengyu Liu, Jinwei Gu, and Sabine Süsstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pages 168–179, 2013. 1

[35] Sing Bing Kang and Richard Weiss. Can we calibrate a camera using an image of a flat, textureless Lambertian surface? In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 640–653, 2000. 2, 3

[36] Juho Kannala and Sami Brandt. A generic camera calibration method for fish-eye lenses. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 1 - Volume 01*, page 10–13, USA, 2004. IEEE Computer Society. 1, 2, 6

[37] Seon Joo Kim and Marc Pollefeys. Robust radiometric calibration and vignetting correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:562–576, 2008. 1, 2

[38] Rudolf Kingslake and R. Barry Johnson. *Lens Design Fundamentals*. Academic Press, Burlington, MA, 2 edition, 2010. 2, 3, 8

[39] Craig Kolb, Don Mitchell, and Pat Hanrahan. A realistic camera model for computer graphics. In *Proceedings of SIGGRAPH*, pages 317–324, 1995. 2

[40] Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. Differentiable monte carlo ray tracing through edge sampling. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 37 (6):222:1–222:11, 2018. 4, 5

[41] Esther Y. H. Lin, Zhecheng Wang, Rebecca Lin, Daniel Miau, Florian Kainz, Jiawen Chen, Xuaner Zhang, David B. Lindell, and Kiriakos N. Kutulakos. Learning lens blur fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2, 3

[42] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V. Sander. Deblur-NeRF: Neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

[43] Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. 5

[44] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2

[45] J. Montiel, Javier Civera, and Andrew Davison. Unified inverse depth parametrization for monocular slam. 2006. 3, 4

[46] David Nakath, Mengkun She, Yifan Song, and Kevin Köser. In-situ joint light and medium estimation for underwater color restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 3731–3740, 2021. 1

[47] Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. Mitsuba 2: A retargetable forward and inverse renderer. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 38(6), 2019. 2, 5

[48] Edwin Olson. AprilTag: A robust and flexible visual fiducial system. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, pages 3400–3407, 2011. 6

[49] Jaesik Park, Sudipta N. Sinha, Yasuyuki Matsushita, Yu-Wing Tai, and In So Kweon. Calibrating a non-isotropic near point light source using a plane. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 1

[50] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. MIT Press, 2023. 2, 3, 5, 1, 4, 6

[51] Les Piegl and Wayne Tiller. *The NURBS Book*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2 edition, 1997. Published 14 November 1996; eBook published 06 December 2012. 4, 3

[52] Ravi Ramamoorthi. *Modeling Illumination Variation with Spherical Harmonics*. 2006. 5

[53] Ravi Ramamoorthi and Pat Hanrahan. An efficient representation for irradiance environment maps. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, page 497–500, New York, NY, USA, 2001. Association for Computing Machinery. 5

[54] Joern Rehder, Janosch Nikolic, Thomas Schneider, and Roland Siegwart. A direct formulation for camera calibration. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 6479–6486, 2017. 1, 2, 3

[55] Sandra Rosalen and Johannes Backhaus. Comparing quality attributes of coated cardboards for inkjet printing by using different methods. *Journal of Print and Media Technology Research*, 8(4):211–226, 2019. Professional communication. 4, 5

[56] Hiroaki Santo, Michael Waechter, Masaki Samejima, Yusuke Sugano, and Yasuyuki Matsushita. Light structure from pin motion: Simple and accurate point light calibration for physics-based modeling. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 1

[57] Thomas Schoeps, Viktor Larsson, Marc Pollefeys, and Torsten Sattler. Why having 10,000 parameters in your camera model is better than twelve, 2019. 1, 6, 7, 3

[58] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113, 2016. 1

[59] Katherine A. Skinner, Eduardo Iscar, and Matthew Johnson-Roberson. Automatic color correction for 3d reconstruction of underwater scenes. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 5140–5147, 2017. 1

[60] Joan Solà, Jeremie Deray, and Dinesh Atchuthan. A micro lie theory for state estimation in robotics, 2021. 6

[61] Huixuan Tang and Kiriakos N. Kutulakos. What does an aberrated photo tell us about the lens and the scene? In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–10, 2013. 2

[62] Thorlabs. R3L3S7N1 negative grid distortion target, 90 mm x 90 mm, 5000 μm spacing. https://www.thorlabs.

com/thorproduct.cfm?partnumber=R3L3S7N1. Manufacturer's drawing R3L3S7N1, 15 Nov 2024. 6, 7

[63] Mikko E Toivonen and Arto Klami. Practical camera sensor spectral response and uncertainty estimation. *Journal of Imaging*, 6(8):79, 2020. 1

[64] Vladyslav Usenko, Nikolaus Demmel, and Daniel Cremers. The double sphere camera model. In *2018 International Conference on 3D Vision (3DV)*, page 552–560. IEEE, 2018. 1, 2, 6

[65] Eric Veach. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, Stanford, CA, USA, 1997. Ph.D. dissertation, Department of Computer Science. 1

[66] Marco Visentini-Scarzanella and Hiroshi Kawasaki. Simultaneous camera, light position and radiant intensity distribution calibration. In *Pacific-Rim Symposium on Image and Video Technology (PSIVT)*, 2015. 1

[67] Bruce Walter, Stephen Marschner, Hongsong Li, and Kenneth Torrance. Microfacet models for refraction through rough surfaces. pages 195–206, 2007. 5

[68] Greg Ward and Elena Eydelberg-Vileshin. Picture perfect rgb rendering using spectral prefiltering and sharp color primaries. In *Proceedings of the 13th Eurographics Workshop on Rendering*, page 117–124, Goslar, DEU, 2002. Eurographics Association. 1

[69] Martin Weber and Roberto Cipolla. A practical method for estimation of point light-sources. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 49.1–49.10, 2001. 1

[70] Wenqi Xian, Aljaž Božič, Noah Snavely, and Christoph Lassner. Neural lens modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2

[71] Cheng Zhang, Bailey Miller, Kai Yan, Ioannis Gkioulekas, and Shuang Zhao. Path-space differentiable rendering. *ACM Trans. Graph.*, 39(4):143:1–143:19, 2020. 2, 4, 5

[72] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000. 7

[73] Ziyi Zhang, Nicolas Roussel, and Wenzel Jakob. Projective sampling for differentiable rendering of geometry. *ACM Trans. Graph.*, 42(6), 2023. 4, 5

[74] Yuanjie Zheng, S. Lin, C. Kambhamettu, Jingyi Yu, and Sing Bing Kang. Single-image vignetting correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:2243–2256, 2009. 2

Real Cameras are Radiometric

Supplementary Material

A.1. Spectral Formulation of the Image Formation

The camera sensor has a certain spectral sensitivity $S(\lambda)$ [2, 4, 17, 50, 63] and measures the spectral radiant flux as [34]

$$S(\lambda)\Phi(\boldsymbol{u},\lambda) = \int_{\mathcal{I}_{\boldsymbol{u}}} S(\lambda)E(\boldsymbol{p},\lambda)\,\mathrm{d}\boldsymbol{p}.$$

This spectral formulation requires knowledge of the camera's spectral sensitivity $S(\lambda)$ which is difficult to recover in practice because it requires specialized equipment [4, 17, 63]. Tri-color (RGB) renderers avoid this by working directly with per-channel quantities [23, 68]. Let Λ_c be the set of wavelengths associated with the channel $c \in \{r, g, b\}$. The per-channel radiant flux measured by the sensor in pixel $\boldsymbol{u}$ becomes

$$\begin{aligned}\Phi_c(\boldsymbol{u}) &\coloneqq \int_{\mathcal{I}_{\boldsymbol{u}}} \underbrace{\int_{\Lambda_c} S_c(\lambda)E(\boldsymbol{p},\lambda)\,\mathrm{d}\lambda}_{=:E_c(\boldsymbol{p})}\,\mathrm{d}\boldsymbol{p}\\ &= \int_{\mathcal{I}_{\boldsymbol{u}}}\int_{\Omega(\boldsymbol{p})} \underbrace{\int_{\Lambda_c} S_c(\lambda)L(\boldsymbol{x}(\boldsymbol{p},\boldsymbol{\omega}),\boldsymbol{\omega}_o(\boldsymbol{p},\boldsymbol{\omega}),\lambda)\,\mathrm{d}\lambda}_{=:L_c(\boldsymbol{x}(\boldsymbol{p},\boldsymbol{\omega}),\boldsymbol{\omega}_o(\boldsymbol{p},\boldsymbol{\omega}))}\\ &\qquad\cdot\cos\theta(\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{p},\end{aligned}$$

where Fubini's theorem [32] is used to swap the order of the integrals. In the main text and in the following, all quantities are per color channel and the channel index c is omitted.

A.2. Geometric Camera Models and the Pinhole Aperture

The pinhole model can be constructed from the thin lens model by making the aperture radius infinitesimal so that for each sensor position only a single ray passes through the pinhole $\boldsymbol{o} \in \mathcal{A}$ [50]. Using the Dirac delta distribution δ, which satisfies the sifting property [8, 65]

$$\int_{\Omega(\boldsymbol{p})} g(\boldsymbol{\omega})\delta(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\,\mathrm{d}\boldsymbol{\omega} = g(\boldsymbol{\omega}_0) \tag{A1}$$

to select the radiance along the single ray direction $\boldsymbol{\omega}_0 \coloneqq \boldsymbol{o}-\boldsymbol{p}$ through the pinhole [65], the radiant flux Eq. (1) simplifies to

$$\begin{aligned}&\int_{\mathcal{I}_{\boldsymbol{u}}}\int_{\Omega(\boldsymbol{p})} L(\boldsymbol{x}(\boldsymbol{p},\boldsymbol{\omega}),\boldsymbol{\omega}_o(\boldsymbol{p},\boldsymbol{\omega}))\cos\theta(\boldsymbol{\omega})\delta(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{p}\\ &\overset{(A1)}{=} \int_{\mathcal{I}_{\boldsymbol{u}}} L(\boldsymbol{x}(\boldsymbol{p},\boldsymbol{\omega}_0),\boldsymbol{\omega}_o(\boldsymbol{p},\boldsymbol{\omega}_0))\cos\theta(\boldsymbol{\omega}_0)\,\mathrm{d}\boldsymbol{p} = \Phi(\boldsymbol{u}).\end{aligned} \tag{A2}$$

Eq. (A2) shows that the pinhole model is a pure geometric model and physically incorrect: a vignetting falloff is missing and every scene point is in perfect focus.

The geometric camera models that the radiometric camera lifts are all models of a pinhole aperture. The radiometric camera uses them through their back-projection $\boldsymbol{\pi}^{-1}$ alone (Eq. (3)), so it places no condition on how a model is parameterized. The remainder of this section defines the five models of Sec. 5.2, derives their back-projections, gives the geometric meaning of their parameters $\boldsymbol{\Theta}_d$, and separates them into *global* and *local* models. The distinction decides how a model is regularized and how the vignetting, which is radially symmetric about the principal point (Sec. 3.3), is attached to it.

Model	Parameters $\boldsymbol{\Theta}_d$	Scope	Principal point
Brown–Conrady	$f_x, f_y, \boldsymbol{c}$, 5 coeff.	global	parameter $\boldsymbol{c}$
OpenCV-14	$f_x, f_y, \boldsymbol{c}$, 14 coeff.	global	parameter $\boldsymbol{c}$
Kannala–Brandt	$f_x, f_y, \boldsymbol{c}$, 4 coeff.	global	parameter $\boldsymbol{c}$
Double Sphere	$f_x, f_y, \boldsymbol{c}, \xi, \alpha$	global	parameter $\boldsymbol{c}$
Central spline	$\boldsymbol{d}_{ij}$ per node	local	estimated, Eq. (A16)

A.2.1. Global Models

Let the z-axis of the camera frame be the optical axis. A parametric model maps a direction $\boldsymbol{d} = (X, Y, Z)^\top$ of this frame to a point (x_d, y_d) of the normalized image plane and from there to the sensor point

$$\boldsymbol{p} = \big(f_x x_d + c_x,\ f_y y_d + c_y\big)^\top, \tag{A3}$$

with the focal lengths f_x, f_y in pixels and the principal point $\boldsymbol{c} = (c_x, c_y)^\top$. The models differ in the map $\boldsymbol{d} \mapsto (x_d, y_d)$, which we write with the angle θ between $\boldsymbol{d}$ and the optical axis and the azimuth φ about it,

$$\theta = \operatorname{atan2}\big(\sqrt{X^2+Y^2}, Z\big), \qquad \varphi = \operatorname{atan2}(Y, X). \tag{A4}$$

All four models are defined by this forward projection. The back-projection $\boldsymbol{\pi}^{-1}$ inverts it: Eq. (A3) gives (x_d, y_d) from $\boldsymbol{p}$, and the inverse of the map $\boldsymbol{d} \mapsto (x_d, y_d)$, derived below for each model, gives the direction. We call these models *global*: each of their few parameters changes the back-projection of every sensor point.

Brown–Conrady. The model starts from the perspective projection $(x, y) = (X/Z, Y/Z)$, whose radius is $\rho = \sqrt{x^2+y^2} = \tan\theta$. A lens that is rotationally symmetric about its optical axis displaces an image point along its radius, and by symmetry the displaced radius is an odd function of ρ, so that $\rho_d = \rho\, R(\rho)$ with a factor R that is even

in ρ. Brown [10] truncates its series after three terms and adds the displacement that Conrady [14] derived for lens elements whose centers are not on a common axis, which is not radial:

$$\begin{aligned} x_d &= x\,R(\rho) + 2p_1xy + p_2(\rho^2 + 2x^2), \\ y_d &= y\,R(\rho) + p_1(\rho^2 + 2y^2) + 2p_2xy, \\ R(\rho) &= 1 + k_1\rho^2 + k_2\rho^4 + k_3\rho^6. \end{aligned} \tag{A5}$$

The radial coefficients k_1, k_2, k_3 describe barrel distortion where $R < 1$ and pincushion distortion where $R > 1$. The tangential coefficients p_1, p_2 are the components of the decentering along the two image axes. Equation (A5) has no closed-form inverse. We solve it for (x, y) iteratively from $(x, y) = (x_d, y_d)$, which has a unique solution where the distortion is monotonic, *i.e.* where $\rho\,R(\rho)$ increases with ρ, and the back-projection is $\boldsymbol{\pi}^{-1}(\boldsymbol{p}) = (x, y, 1)^\top / \|(x, y, 1)\|$. Since $\rho = \tan\theta$ diverges for $\theta \to 90°$, the model cannot represent a field of view of $180°$.

OpenCV-14. The 14-parameter model of OpenCV [9] extends Eq. (A5) by three effects. First, the radial factor becomes a rational function,

$$R(\rho) = \frac{1 + k_1\rho^2 + k_2\rho^4 + k_3\rho^6}{1 + k_4\rho^2 + k_5\rho^4 + k_6\rho^6}, \tag{A6}$$

which tends to a constant for large ρ, where a polynomial diverges, and therefore follows a wide-angle lens further toward the image border. Second, thin-prism terms describe lens elements that are tilted against the optical axis, which displace the image by even powers of the radius along a fixed direction:

$$\begin{aligned} x' &= x\,R(\rho) + 2p_1xy + p_2(\rho^2 + 2x^2) + s_1\rho^2 + s_2\rho^4, \\ y' &= y\,R(\rho) + p_1(\rho^2 + 2y^2) + 2p_2xy + s_3\rho^2 + s_4\rho^4. \end{aligned} \tag{A7}$$

Third, the sensor may be tilted against the plane perpendicular to the optical axis. With the rotation $\mathrm{T} = \mathrm{R}_y(\tau_y)\,\mathrm{R}_x(\tau_x)$ of the sensor by the angles τ_x, τ_y about the image axes, with the entries T_{ij}, the point is mapped by the homography

$$\lambda \begin{pmatrix} x_d \\ y_d \\ 1 \end{pmatrix} = \begin{pmatrix} T_{33} & 0 & -T_{13} \\ 0 & T_{33} & -T_{23} \\ 0 & 0 & 1 \end{pmatrix} \mathrm{T} \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}, \tag{A8}$$

which rotates the image plane into the plane of the tilted sensor and projects it back along the optical axis, so that the principal point stays where the optical axis meets the sensor. The 14 parameters are $k_1, \ldots, k_6, p_1, p_2, s_1, \ldots, s_4$ and τ_x, τ_y, and $k_4 = \cdots = \tau_y = 0$ is the Brown–Conrady model. For the back-projection, Eq. (A8) is inverted in closed form, as a homography, and Eq. (A7) is then solved for (x, y) iteratively as above. The model distorts the perspective projection as well and shares its limit of $180°$.

Kannala–Brandt. A fisheye lens is not designed to image a direction at the radius $\tan\theta$ of the perspective projection, but at a radius that stays finite at $\theta = 90°$, *e.g.* θ (equidistant), $2\tan\frac{\theta}{2}$ (stereographic) or $2\sin\frac{\theta}{2}$ (equisolid angle). All of these, and the perspective projection itself, are odd functions of θ. Kannala and Brandt [36] therefore model a rotationally symmetric lens by an odd polynomial in θ, whose linear coefficient is absorbed by the focal length:

$$\begin{aligned} (x_d, y_d) &= r(\theta)\,(\cos\varphi, \sin\varphi), \\ r(\theta) &= \theta + k_1\theta^3 + k_2\theta^5 + k_3\theta^7 + k_4\theta^9, \end{aligned} \tag{A9}$$

with θ and φ of Eq. (A4). The first term is the equidistant projection, and $k_1, \ldots, k_4$ describe the deviation of the lens from it. The model has no decentering terms. For the back-projection, the azimuth is that of the image point, $\varphi = \operatorname{atan2}(y_d, x_d)$, and the angle is the root of $r(\theta) = \sqrt{x_d^2 + y_d^2}$, which we find iteratively from $\theta = \sqrt{x_d^2 + y_d^2}$ and which is unique where r is monotonic. Then

$$\boldsymbol{\pi}^{-1}(\boldsymbol{p}) = (\sin\theta\cos\varphi,\ \sin\theta\sin\varphi,\ \cos\theta)^\top. \tag{A10}$$

Since θ is not restricted to less than $90°$, the model represents fields of view of $180°$ and more.

Double Sphere. The model of Usenko *et al.* [64] is constructed geometrically. The direction is projected onto a unit sphere about the center of projection, which gives $\boldsymbol{d}/d_1$ with $d_1 = \|\boldsymbol{d}\|$. This point is expressed in the frame of a second unit sphere whose center is displaced by ξ along the optical axis, where it is $(X, Y, Z + \xi d_1)^\top / d_1$, and is projected onto the second sphere by dividing by its norm d_2/d_1,

$$d_2 = \sqrt{X^2 + Y^2 + (\xi d_1 + Z)^2}. \tag{A11}$$

Finally, the point $(X, Y, Z + \xi d_1)^\top / d_2$ on the second sphere is projected perspectively from a center that is displaced by $\frac{\alpha}{1-\alpha}$ behind the center of this sphere, which gives

$$(x_d, y_d) = \frac{(X, Y)}{\alpha d_2 + (1 - \alpha)(\xi d_1 + Z)}. \tag{A12}$$

The parameter ξ is thus the distance between the two spheres and $\alpha \in [0, 1]$ the position of the final center of projection, and $\xi = 0$, $\alpha = 0$ is the perspective projection. Both describe a rotationally symmetric lens, and the model has no decentering terms. Unlike the three models above, it has a closed-form inverse. With $\rho^2 = x_d^2 + y_d^2$,

$$\begin{aligned} m_z &= \frac{1 - \alpha^2\rho^2}{\alpha\sqrt{1 - (2\alpha - 1)\rho^2} + 1 - \alpha}, \\ \boldsymbol{\pi}^{-1}(\boldsymbol{p}) &= \frac{m_z\xi + \sqrt{m_z^2 + (1 - \xi^2)\rho^2}}{m_z^2 + \rho^2} \begin{pmatrix} x_d \\ y_d \\ m_z \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \xi \end{pmatrix}, \end{aligned} \tag{A13}$$

which follows from inverting the three steps in reverse order: the first line intersects the ray of the shifted perspective projection with the second sphere, and the second line intersects the ray from the center of the second sphere through this point with the first sphere. The inverse is a unit vector, and it is defined where $(2\alpha - 1)\rho^2 \leq 1$, which bounds the image radius for $\alpha > \frac{1}{2}$. The model represents fields of view beyond $180°$.

A.2.2. The Local Model

The central generic model of Schöps *et al*. [57] has no projection function with physical parameters. It stores the back-projection itself: directions at the nodes of a regular lattice over the sensor, interpolated by a bicubic B-spline [51] and normalized. With the control directions $\boldsymbol{d}_{ij} \in \mathbb{R}^3$ at the nodes (i, j) and the cubic B-spline basis functions $N_{i,3}$, $N_{j,3}$ over the lattice,

$$\hat{\boldsymbol{d}}(\boldsymbol{p}) = \frac{\tilde{\boldsymbol{d}}(\boldsymbol{p})}{\|\tilde{\boldsymbol{d}}(\boldsymbol{p})\|}, \quad \tilde{\boldsymbol{d}}(\boldsymbol{p}) = \sum_{i,j} N_{i,3}(p_x)\, N_{j,3}(p_y)\, \boldsymbol{d}_{ij}, \tag{A14}$$

and $\boldsymbol{\Theta}_d = \{\boldsymbol{d}_{ij}\}$. On the Sigma 14 mm, for example, the lattice has 35×25 nodes at a spacing of about 160 pixels. We call this model *local*: a cubic basis function is nonzero over four cells of the lattice, so $\hat{\boldsymbol{d}}(\boldsymbol{p})$ depends only on the 4×4 control directions around $\boldsymbol{p}$, and a control direction changes the back-projection only in the 4×4 cells around its node. The model is given by its back-projection and not, as the global models are, by its forward projection, which has no closed form. Where a projection is needed, as for the reprojection errors of Sec. 5.2, we invert Eq. (A14) numerically by Levenberg–Marquardt iterations; a second spline regressed to the inverse is an alternative, and it is local as well.

The locality has three consequences for the calibration. *First*, the model can follow a distortion that no global model describes, but a node is determined only by the photosites in its own support. Nodes at the border of the lens coverage, which few photosites see, are therefore not determined by the data, whereas the parameters of a global model are determined by the whole image. This is why the central spline, and no other model, is regularized (Sec. 4.5). *Second*, the model has no distinguished camera frame: rotating every control direction by a rotation R and every pose by its inverse changes no ray in the scene. A global model does not have this gauge freedom, since its frame is fixed by its optical axis (Sec. A.9). *Third*, and for the same reason, the model has no principal point. The intrinsics of Eq. (A3) do not enter Eq. (A14), which is defined over the sensor directly, so the point about which the vignetting is symmetric has to be estimated (Sec. A.2.3).

A.2.3. The Center of the Vignetting

The vignetting V of Sec. 3.3 is radially symmetric about a point $\boldsymbol{c}$ of the sensor. The vignetting of a lens is symmetric about its optical axis, so $\boldsymbol{c}$ has to be the sensor point that images the axis, which in general is not the center of the sensor.

Global models. The principal point of Eq. (A3) is this point by definition: each of the four models projects the direction of the optical axis, $\boldsymbol{d} = (0, 0, 1)^\top$, to $(x_d, y_d) = (0, 0)$ and thus to $\boldsymbol{p} = \boldsymbol{c}$. We center the vignetting at the principal point of the geometric start.

The local model. For the central spline, we estimate the optical axis as the axis about which the camera is closest to radially symmetric. Let

$$\boldsymbol{\pi}_{\mathrm{sym}}(\boldsymbol{d}) \coloneqq \boldsymbol{c}' + r(\theta) \begin{pmatrix} f_x \cos\varphi \\ f_y \sin\varphi \end{pmatrix}, \quad r(\theta) = \theta + \sum_{k=1}^{4} \kappa_k \theta^{2k+1}, \tag{A15}$$

be the projection of a radially symmetric camera, with θ and φ of Eq. (A4): the model of Eq. (A9) with its own focal lengths f_x, f_y, principal point $\boldsymbol{c}'$ and coefficients $\boldsymbol{\kappa}$. We regress this camera and a rotation $\mathrm{R} \in SO(3)$ to the rays of the spline,

$$\min_{\mathrm{R},\, \boldsymbol{c}',\, f_x,\, f_y,\, \boldsymbol{\kappa}} \sum_{\boldsymbol{p} \in \mathcal{R}} \left\| \boldsymbol{p} - \boldsymbol{\pi}_{\mathrm{sym}}\left(\mathrm{R}\, \hat{\boldsymbol{d}}(\boldsymbol{p})\right) \right\|^2, \tag{A16}$$

where $\mathcal{R}$ are the pixels in which the geometric start is defined. We minimize it by Levenberg–Marquardt iterations, first with the coefficient κ_1 alone and then with one more coefficient at a time, each from the previous minimum, which keeps the higher orders from folding the profile at the image border. The symmetric camera is not used as a camera model. Its role is to locate the axis: at the minimum, $\mathrm{R}^\top (0, 0, 1)^\top$ is the axis about which the rays of the spline are closest to symmetric, the optical axis of the lens, and $\boldsymbol{c}'$ is the sensor point that images it. A lens that is not symmetric about any axis leaves a residual in Eq. (A16), but the axis remains defined as its minimizer.

We then turn the camera frame onto this axis. Every control direction is rotated, $\boldsymbol{d}_{ij} \leftarrow \mathrm{R}\, \boldsymbol{d}_{ij}$, which rotates every ray of Eq. (A14) by R, since the normalization commutes with a rotation, and every pose $(\mathrm{R}_k, \boldsymbol{t}_k)$ from the world to the camera becomes $(\mathrm{R}\mathrm{R}_k, \mathrm{R}\boldsymbol{t}_k)$. This uses the gauge freedom of the local model and changes no ray in the scene. Afterwards the z-axis of the camera frame is the optical axis, and $\boldsymbol{c} \coloneqq \boldsymbol{c}'$ is the principal point of the spline camera, with the same meaning as that of a global model. The regression is done once, on the geometric start. In our experiments it turns the frame of the spline starts by $0.2°$ to $1.6°$, and the principal point lies 8 to 25 pixels from the center of the sensor.

During the refinement. For every model, $\boldsymbol{c}$ is held at this start value and is not a parameter of the refinement, and the gauge $V(\boldsymbol{c}) = 1$ between the vignetting and the lighting is fixed there (Sec. A.9). The rotation gauge of the spline camera is fixed with respect to the same axis: the held tangent coordinates of Sec. A.9 are those of the node whose direction is closest to it.

A.3. Thin Lens Equivalence

Let $\boldsymbol{\pi}^{-1}$ be the perspective back-projection, $\hat{\boldsymbol{d}}(\boldsymbol{p}) = \frac{\boldsymbol{o}-\boldsymbol{p}}{\|\boldsymbol{o}-\boldsymbol{p}\|}$, and let the focus surface be the plane at the object distance s from $\boldsymbol{o}$, perpendicular to the optical axis. The ray of $\boldsymbol{p}$ meets this plane at the distance $t(\boldsymbol{p}) = \frac{s}{\cos\theta(\boldsymbol{p})}$ from $\boldsymbol{o}$, where $\theta(\boldsymbol{p})$ is the angle between $\hat{\boldsymbol{d}}(\boldsymbol{p})$ and the optical axis, so $\tau(\boldsymbol{p}) = \frac{\cos\theta(\boldsymbol{p})}{s}$ and $\boldsymbol{x}_f(\boldsymbol{p}) = \boldsymbol{o} + \frac{s}{\cos\theta(\boldsymbol{p})}\hat{\boldsymbol{d}}(\boldsymbol{p})$. In the thin lens model, all rays from $\boldsymbol{p}$ through the aperture meet in the point in which the ray through the center $\boldsymbol{o}$ of the lens, which is not refracted, meets the plane of focus [50]. This point is $\boldsymbol{x}_f(\boldsymbol{p})$, so that the ray through an aperture point $\boldsymbol{a}$ has the direction $\frac{\boldsymbol{x}_f(\boldsymbol{p})-\boldsymbol{a}}{\|\boldsymbol{x}_f(\boldsymbol{p})-\boldsymbol{a}\|} = \boldsymbol{\pi}^{-1}(\boldsymbol{p},\boldsymbol{a})$ of Eq. (3): without distortion and with a plane of focus, the radiometric camera is the thin lens.

A.4. Inverse Focus Distance

The lifted back-projection Eq. (3) depends on the focus surface through the distance $t(\boldsymbol{p}) = 1/\tau(\boldsymbol{p})$ from $\boldsymbol{o}$ to the point of focus. We show that its derivative w.r.t. the distance t vanishes for a far focus, whereas its derivative w.r.t. the inverse distance τ does not, which is why we parameterize the focus surface by τ [22, 45]. Let $\boldsymbol{o} = \boldsymbol{0}$ without loss of generality. With the distance,

$$\frac{\partial}{\partial t}\frac{t(\boldsymbol{p})\hat{\boldsymbol{d}}(\boldsymbol{p}) - \boldsymbol{a}}{\|t(\boldsymbol{p})\hat{\boldsymbol{d}}(\boldsymbol{p}) - \boldsymbol{a}\|}$$
$$= \frac{\hat{\boldsymbol{d}}(\boldsymbol{p})\|t(\boldsymbol{p})\hat{\boldsymbol{d}}(\boldsymbol{p}) - \boldsymbol{a}\| - (t(\boldsymbol{p})\hat{\boldsymbol{d}}(\boldsymbol{p}) - \boldsymbol{a})\frac{t(\boldsymbol{p})\|\hat{\boldsymbol{d}}(\boldsymbol{p})\|^2 - \hat{\boldsymbol{d}}(\boldsymbol{p})\cdot\boldsymbol{a}}{\|t(\boldsymbol{p})\hat{\boldsymbol{d}}(\boldsymbol{p}) - \boldsymbol{a}\|}}{\|t(\boldsymbol{p})\hat{\boldsymbol{d}}(\boldsymbol{p}) - \boldsymbol{a}\|^2}$$

$$\lim_{t\to\infty}\frac{\partial}{\partial t}\frac{t(\boldsymbol{p})\hat{\boldsymbol{d}}(\boldsymbol{p}) - \boldsymbol{a}}{\|t(\boldsymbol{p})\hat{\boldsymbol{d}}(\boldsymbol{p}) - \boldsymbol{a}\|}$$
$$= \lim_{t\to\infty}\frac{\hat{\boldsymbol{d}}(\boldsymbol{p})\|\hat{\boldsymbol{d}}(\boldsymbol{p}) - \frac{\boldsymbol{a}}{t(\boldsymbol{p})}\| - (\hat{\boldsymbol{d}}(\boldsymbol{p}) - \frac{\boldsymbol{a}}{t(\boldsymbol{p})})\frac{\|\hat{\boldsymbol{d}}(\boldsymbol{p})\|^2 - \frac{\hat{\boldsymbol{d}}(\boldsymbol{p})\cdot\boldsymbol{a}}{t(\boldsymbol{p})}}{\|\hat{\boldsymbol{d}}(\boldsymbol{p}) - \frac{\boldsymbol{a}}{t(\boldsymbol{p})}\|}}{\|\hat{\boldsymbol{d}}(\boldsymbol{p}) - \frac{\boldsymbol{a}}{t(\boldsymbol{p})}\|^2}$$
$$= 0.$$

A focus surface that is far away thus receives no gradient. With the inverse distance $\tau = \frac{1}{t}$,

$$\frac{\partial}{\partial\tau}\frac{-\tau(\boldsymbol{p})\boldsymbol{a} + \hat{\boldsymbol{d}}(\boldsymbol{p})}{\|-\tau(\boldsymbol{p})\boldsymbol{a} + \hat{\boldsymbol{d}}(\boldsymbol{p})\|}$$
$$= \frac{-\boldsymbol{a}\|\hat{\boldsymbol{d}}(\boldsymbol{p}) - \tau(\boldsymbol{p})\boldsymbol{a}\| + (\hat{\boldsymbol{d}}(\boldsymbol{p}) - \tau(\boldsymbol{p})\boldsymbol{a})\frac{\hat{\boldsymbol{d}}(\boldsymbol{p})\cdot\boldsymbol{a} - \tau(\boldsymbol{p})\|\boldsymbol{a}\|^2}{\|\hat{\boldsymbol{d}}(\boldsymbol{p}) - \tau(\boldsymbol{p})\boldsymbol{a}\|}}{\|\hat{\boldsymbol{d}}(\boldsymbol{p}) - \tau(\boldsymbol{p})\boldsymbol{a}\|^2}$$

$$\lim_{\tau\to 0}\frac{-\boldsymbol{a}\|\hat{\boldsymbol{d}}(\boldsymbol{p}) - \tau(\boldsymbol{p})\boldsymbol{a}\| + (\hat{\boldsymbol{d}}(\boldsymbol{p}) - \tau(\boldsymbol{p})\boldsymbol{a})\frac{\hat{\boldsymbol{d}}(\boldsymbol{p})\cdot\boldsymbol{a} - \tau(\boldsymbol{p})\|\boldsymbol{a}\|^2}{\|\hat{\boldsymbol{d}}(\boldsymbol{p}) - \tau(\boldsymbol{p})\boldsymbol{a}\|}}{\|\hat{\boldsymbol{d}}(\boldsymbol{p}) - \tau(\boldsymbol{p})\boldsymbol{a}\|^2}$$
$$= \frac{-\boldsymbol{a}\|\hat{\boldsymbol{d}}(\boldsymbol{p})\| + \hat{\boldsymbol{d}}(\boldsymbol{p})\frac{\hat{\boldsymbol{d}}(\boldsymbol{p})\cdot\boldsymbol{a}}{\|\hat{\boldsymbol{d}}(\boldsymbol{p})\|}}{\|\hat{\boldsymbol{d}}(\boldsymbol{p})\|^2}$$

and since $\|\hat{\boldsymbol{d}}(\boldsymbol{p})\| = 1$, this limit is $-\big(\boldsymbol{a} - (\boldsymbol{a}\cdot\hat{\boldsymbol{d}}(\boldsymbol{p}))\,\hat{\boldsymbol{d}}(\boldsymbol{p})\big)$, the negative of the component of $\boldsymbol{a}$ perpendicular to the ray. It is nonzero for every aperture point off the ray, so the focus surface keeps a gradient however far it is.

A.5. Vignetting of the Thin Lens

The radiant flux Eq. (1) integrates over the solid angle $\Omega(\boldsymbol{p})$ under which the sensor point $\boldsymbol{p}$ sees the aperture. For a thin lens with the image distance s', changing the integration domain to the aperture area $\mathcal{A}$ by the transformation $\boldsymbol{\omega} = (\boldsymbol{a}-\boldsymbol{p})/\|\boldsymbol{a}-\boldsymbol{p}\|$ [50] gives

$$\Phi(\boldsymbol{u}) = \int_{\mathcal{I}_{\boldsymbol{u}}}\int_{\mathcal{A}}\frac{\cos^4\theta(\boldsymbol{a}-\boldsymbol{p})}{s'^2} L\big(\boldsymbol{x}(\boldsymbol{p},\boldsymbol{a}), -\boldsymbol{\pi}^{-1}(\boldsymbol{p},\boldsymbol{a})\big)\,\mathrm{d}\boldsymbol{a}\,\mathrm{d}\boldsymbol{p}, \tag{A17}$$

where $\theta(\boldsymbol{a}-\boldsymbol{p})$ is the angle between $\boldsymbol{a}-\boldsymbol{p}$ and the optical axis. This separates the falloff toward the sensor border from the solid angle: the factor $\cos^4\theta/s'^2$ is a well-known property of the thin lens and describes its vignetting [28]. Figure 6 shows that the vignetting of real lenses differs much from it. In the radiometric camera, we therefore replace this factor by the generic vignetting function $V(\boldsymbol{p})$, normalized by the aperture area πr^2:

$$\Phi(\boldsymbol{u}) \approx \int_{\mathcal{I}_{\boldsymbol{u}}}\frac{V(\boldsymbol{p})}{\pi r^2}\int_{\mathcal{A}} L\big(\boldsymbol{x}(\boldsymbol{p},\boldsymbol{a}), -\boldsymbol{\pi}^{-1}(\boldsymbol{p},\boldsymbol{a})\big) \cdot \cancel{\frac{\cos^4\theta(\boldsymbol{a}-\boldsymbol{p})}{s'^2}}\,\mathrm{d}\boldsymbol{a}\,\mathrm{d}\boldsymbol{p}, \tag{A18}$$

where the back-projection of the thin lens is replaced by the lifted back-projection $\boldsymbol{\pi}^{-1}(\boldsymbol{p},\boldsymbol{a})$ of Eq. (3), which is Eq. (4) after the substitution $\boldsymbol{a} = \boldsymbol{o} + r\boldsymbol{\xi}$ that maps the aperture to the unit disc $\mathcal{D}$, with $\mathrm{d}\boldsymbol{a} = r^2\,\mathrm{d}\boldsymbol{\xi}$.

A.6. Knots of the Vignetting Spline

It is $\|\boldsymbol{p}-\boldsymbol{c}\| \le r_{\max}$ and therefore

$$\frac{\|\boldsymbol{p}-\boldsymbol{c}\|}{r_{\max}+\|\boldsymbol{p}-\boldsymbol{c}\|} \in [0, \frac{1}{2}].$$

Let p_k, p_{k+1} two uniform consecutive sensor positions, then $\frac{\|\boldsymbol{p}_{k+1}-\boldsymbol{c}\|}{r_{\max}+\|\boldsymbol{p}_{k+1}-\boldsymbol{c}\|}$ and $\frac{\|\boldsymbol{p}_{k}-\boldsymbol{c}\|}{r_{\max}+\|\boldsymbol{p}_{k}-\boldsymbol{c}\|}$ are the two corresponding consecutive samples chosen in $[0, \frac{1}{2}]$.

A.7. Boundary Terms of the Radiant Flux

Section 4.1 removes the two boundary terms that arise in the calibration. Here we list the boundary terms of the radiant flux in a general scene and say how each is treated.

The problem. Differentiable ray tracing learns scene parameters by minimizing the error between rendered and captured images [40, 47]. It has to differentiate integrals whose domain, or whose integrand's discontinuities, depend on the parameters [40, 71, 73]. By the multidimensional Leibniz integral rule, such a derivative is the sum of the integral of the derivative, the interior integral, and a boundary integral over the moving boundaries [13, 24, 71]. Automatic differentiation of a sampled estimate computes the interior integral only [71, 73].

The boundary terms of a general scene. Differentiating the radiant flux Φ (Eq. (4)) w.r.t. the parameters $\bar{\boldsymbol{\Theta}}$ differentiates the integrals over the pixel $\mathcal{I}_{\boldsymbol{u}}$ and the aperture $\mathcal{A}(r)$ and, for each bounce of a ray, the integral over the hemisphere $\mathcal{H}^2(\boldsymbol{x})$ in the radiance L (Eq. (8)). With

$$\hat{E}(\boldsymbol{p},\boldsymbol{a}) \coloneqq L\big(\boldsymbol{x}(\boldsymbol{p},\boldsymbol{a}), -\boldsymbol{\pi}^{-1}(\boldsymbol{p},\boldsymbol{a})\big),$$

we identify the following sources of boundary terms.
I. $\boldsymbol{a} \mapsto \hat{E}(\boldsymbol{p},\boldsymbol{a})$ has parameter-dependent discontinuities over $\mathcal{A}(r)$, from surface silhouettes, sharp shadows and a discontinuous surface texture, *e.g.* the pattern of a calibration target.
II. $\boldsymbol{p} \mapsto \hat{E}(\boldsymbol{p},\boldsymbol{a})$ has the same discontinuities over $\mathcal{I}_{\boldsymbol{u}}$.
III. The aperture $\mathcal{A}(r)$ depends on the radius $r \in \bar{\boldsymbol{\Theta}}$.
IV. By the chain rule, every bounce requires $\frac{\partial}{\partial \boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\omega}_o)$, the derivative of the integral over $\mathcal{H}^2(\boldsymbol{x})$ in Eq. (8). Its integrand has discontinuities from the silhouettes, the shadows and the discontinuous textures seen from $\boldsymbol{x}$, which move with $\boldsymbol{x}$, and its domain $\mathcal{H}^2(\boldsymbol{x})$ depends on $\boldsymbol{x}$. Light sources with a boundary, such as area lights, add a further discontinuity of the incident radiance. This repeats at every further bounce.

The two terms of the calibration. Two of these sources remain in a calibration scene, and we remove both by a reformulation (Sec. 4.1, Fig. 3). For **III**, we transform the aperture to the fixed unit disc $\mathcal{D} \subseteq \mathbb{R}^2$ with $\boldsymbol{a} = \boldsymbol{o} + r\boldsymbol{\xi}$, $\boldsymbol{\xi} \in \mathcal{D}$, which gives Eq. (4),

$$\Phi(\boldsymbol{u}) = \frac{1}{\pi} \int_{\mathcal{I}_{\boldsymbol{u}}} V(\boldsymbol{p}) \int_{\mathcal{D}} \hat{E}(\boldsymbol{p}, \boldsymbol{o} + r\boldsymbol{\xi}) \,\mathrm{d}\boldsymbol{\xi} \,\mathrm{d}\boldsymbol{p},$$

whose domain does not depend on r. For the edges of the pattern in **I** and **II**, we use that a print is never perfectly sharp. We model an edge of the print by the convolution of the print design with a Gaussian of the standard deviation $\sigma_w = w/1.805951$, where $w = 50\,\mu\mathrm{m}$ is the 10–70 % edge width of a realistic print [55]. The albedo is then the cumulative Gaussian of the signed distance to the edges of the design, which we compute from the print file, and is smooth.

The terms that do not arise. The remaining sources are absent from a calibration scene. The silhouettes in **I** and **II** are not seen: we render only the pixels of a mask $\mathcal{M} \subseteq \mathcal{I}$, with a padding, whose rays all meet the target. For **IV**, the scene contains the target alone, and the light scattered by its environment is represented by a smooth lighting (Sec. 4.2), so that no silhouette, shadow or boundary of a light is seen from $\boldsymbol{x}$. The dependence of the hemisphere $\mathcal{H}^2(\boldsymbol{x})$ on $\boldsymbol{x}$ and the silhouette of the target seen from a point of the target contribute nothing, since the cosine in Eq. (8) vanishes at the horizon, $\cos(\pi/2) = 0$. A general scene requires these terms, which differentiable renderers estimate by sampling the boundaries [40, 71, 73]; we leave them to future work.

A.8. Lighting and Reflectance

Lighting. The radiance of the spherical-harmonic lighting of degree $l_{\max}$ seen from the direction $\boldsymbol{\omega}$ is

$$L^{\mathrm{SH}}(\boldsymbol{\omega}; \boldsymbol{\Theta}_l) \coloneqq \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} a_l^m \, y_l^m(\theta(\boldsymbol{\omega}), \phi(\boldsymbol{\omega})),$$

with the basis functions y_l^m, the coefficients $\boldsymbol{\Theta}_l = \{a_l^m\} \subseteq \mathbb{R}$ per color channel and the spherical coordinates $(\theta(\boldsymbol{\omega}), \phi(\boldsymbol{\omega}))$ of $\boldsymbol{\omega}$ [1, 5, 52, 53]. We use $l_{\max} = 2$ [53]. Section 4.2 places it on a sphere of finite radius.

Reflectance. We compute Eq. (8) by Monte Carlo estimation [50]. N microfacet normals [67] $\boldsymbol{n}_k$ are sampled from the GGX distribution of visible normals, $\boldsymbol{n}_k \sim D^{\mathrm{v}}(\cdot, \boldsymbol{\omega}_o; \alpha)$, by sampling a spherical cap [19], and give the directions $\boldsymbol{\omega}_{i,k}$ by reflection, with $\mathrm{PDF}(\boldsymbol{n}_k, \boldsymbol{\omega}_o; \alpha) = D^{\mathrm{v}}(\boldsymbol{n}_k, \boldsymbol{\omega}_o; \alpha)/(4(\boldsymbol{\omega}_o \cdot \boldsymbol{n}_k))$, so that

$$L(\boldsymbol{x}, \boldsymbol{\omega}_o; \boldsymbol{\Theta}_r) \approx \frac{1}{N} \sum_{k=1}^{N} \frac{f_k^{\mathrm{CT}} \, L_k \cos\theta_{i,k}}{\mathrm{PDF}(\boldsymbol{n}_k, \boldsymbol{\omega}_o; \alpha)},$$

with $L_k := L(\boldsymbol{x}'(\boldsymbol{x}, \boldsymbol{\omega}_{i,k}), -\boldsymbol{\omega}_{i,k}; \boldsymbol{\Theta}_r)$ and the Cook–Torrance reflectance [15, 19]

$$f_k^{\mathrm{CT}} := f^{\mathrm{CT}}(\boldsymbol{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_{i,k}; \boldsymbol{\Theta}_m) = \mu \frac{\rho(\boldsymbol{x})}{\pi} + (1-\mu) \frac{F(\boldsymbol{\omega}_o, \boldsymbol{\omega}_h; \eta)\, D(\boldsymbol{\omega}_h; \alpha)\, G_2(\boldsymbol{\omega}_o, \boldsymbol{\omega}_{i,k})}{4 \cos\theta_o \cos\theta_{i,k}}.$$

Here $\boldsymbol{\omega}_h$ is the half-vector between $\boldsymbol{\omega}_o$ and $\boldsymbol{\omega}_{i,k}$, θ_o and $\theta_{i,k}$ are the angles between the directions and the surface normal, F is the exact unpolarized Fresnel term of a dielectric with the index of refraction η per color channel [50], D the isotropic GGX distribution, $\alpha_x = \alpha_y = \alpha$ [31], and G_2 the height-correlated masking and shadowing function [19, 31]. $\boldsymbol{\Theta}_r := \boldsymbol{\Theta}_l \cup \boldsymbol{\Theta}_m$ are the parameters of the lighting and the material. The roughness and the index of refraction are blended between their values for ink and for the blank print by the albedo $\rho(\boldsymbol{x})$.

A.9. Gauge Freedoms

The parameters $\bar{\boldsymbol{\Theta}}$ contain three families of transformations that leave every rendered photosite value Φ unchanged. We fix each of them exactly by the parameterization, *i.e.* by removing the corresponding direction from the unknowns, and not by a penalty or a prior, so that no weight has to be chosen and the remaining parameters keep their meaning.

Vignetting and lighting. Φ is the product of the vignetting V and a radiance that is linear in the lighting $\boldsymbol{\Theta}_l$. Scaling V by $\sigma > 0$ and the lighting by $1/\sigma$ therefore leaves Φ unchanged. We hold $V = 1$ at the centre of the vignetting, the principal point: the control points are scaled once so that the centre value is one, and the direction of the control points that changes the centre value is eliminated from the unknowns, so that every update keeps it exactly. The overall scale is then carried by the lighting alone.

Rotation of the spline camera and the poses. A parametric model fixes its camera frame by its optical axis. The directions of the generic spline model, in contrast, can all be rotated by the same rotation, and the poses of all views by its inverse, without changing any ray in the scene. We hold three coordinates of the directions that determine this rotation: both tangent coordinates of the node whose direction is closest to the optical axis, and, at the node farthest from it in angle, the tangent coordinate most aligned with the roll about that direction. These three coordinates are a nonsingular function of the rotation, so holding them removes the gauge and no degree of freedom of the lens. The poses remain free.

Board warping and the poses. A height field over the board that is affine in the board coordinates, *i.e.* an offset along the normal and two tilts, is a rigid motion of the board and is already described by the pose of each view. The nodal heights of the thin-plate spline are therefore restricted to those whose interpolant has no affine term, which leaves pure bending. Likewise, the skew of the pixel grid is held for every model, since it is interchangeable with the in-plane shear of the printed target.

Lens	Type	Geom.	Vign.	Defocus
Tamron 20 mm F/2.8 Di III	wide angle	✓	✓	✓
Sigma 14 mm F2.8 EX HSM	rectilinear	✓	✓	✓
Sigma 8 mm F3.5 EX DG	circular fisheye	✓	✓	✓
24–70 mm zoom at 48 mm	normal			✓

Table A1. The lenses of the data set and the experiments they take part in: geometric accuracy (Sec. 5.2), vignetting (Sec. 5.3) and defocus (Sec. 5.4). The camera is a Sony ILME-FX3 throughout.

A.10. Linear Parameters of the Radiant Flux

The start of the calibration (Sec. 4.5) uses that the radiant flux Eq. (4) is linear in two of its parameter blocks and affine in a third.

Vignetting. The vignetting $V(\boldsymbol{p}) = \sum_i N_{i,3}(\cdot)\, P_i$ is linear in its control points $\boldsymbol{\Theta}_v = \{P_i\}$ and multiplies the integral over the aperture, so Φ is linear in $\boldsymbol{\Theta}_v$.

Lighting. The lighting L^{SH} is linear in its coefficients $\boldsymbol{\Theta}_l = \{a_l^m\}$ (Sec. A.8), the radiance that leaves the target (Eq. (8)) is linear in the incident radiance, and the integrals over the hemisphere, the aperture and the pixel are linear, so Φ is linear in $\boldsymbol{\Theta}_l$.

Diffuse weight. The Cook–Torrance reflectance is $f = \mu f_{\mathrm{d}} + (1-\mu) f_{\mathrm{s}}$ with its diffuse part f_{d} and its specular part f_{s}, so Φ is affine in μ.

With the other parameters held, each of the three blocks is therefore the solution of a linear least-squares problem in the residuals of Eq. (6), which we solve once per block at the start: the lighting with $V \equiv 1$, then the vignetting, then μ. The flux is bilinear in the vignetting and the lighting, which is the scale gauge of Sec. A.9.

A.11. Data Set

Table A2 lists every capture of the data set (Sec. 5), grouped by experiment. All images are Sony ILME-FX3 RAW files (4256×2848 decoded pixels) at ISO 100. The board distance is not measured: it is the median perpendicular distance from the camera centre to the target plane over the estimated poses of a set. *Captured* counts the images of a set and *used* those in the experiment's configuration: a training image is left out when the geometric pipeline that

starts our refinement does not localize the board in it, and a test image when its dots are not registered. Each Thorlabs test set is captured together with its training set, in the same session and with the same focus and aperture.

The runtime of a calibration is given in Sec. A.14.

A.12. Geometric Starts

Every radiometric refinement starts from a geometric calibration of the same model on the same images. All of them are calibrations from correspondences between detected image points and the known points of the target, which differ in the detector and in the tool that estimates the camera.

AprilGrid. The corners are detected by OpenCV's ArUco detector [27] with sub-pixel refinement. Brown–Conrady and OpenCV-14 are calibrated by OpenCV's calibration [9, 72], Kannala–Brandt by its fisheye calibration, Double Sphere by Kalibr [26], and the central spline by the calibration of Schöps *et al*. [57] on these corners.

Siemens stars. The features are detected by the detector of Schöps *et al*. [57], whose calibration gives the central spline. OpenCV and Kalibr cannot detect this target. Brown–Conrady and OpenCV-14 are calibrated by OpenCV's calibration on the correspondences of Schöps's detector, and Kannala–Brandt by its fisheye calibration on them or, on some sets, by our own estimator. Double Sphere, which OpenCV does not provide, is always calibrated by our own estimator: a robust pose of the target per image, followed by a joint refinement of the camera and all poses that minimizes the reprojection error.

A.13. Thorlabs Evaluation

The geometric accuracy of every camera model in Sec. 5.2 is measured on images of a high-precision distortion target that none of the models was calibrated on. Section A.13.1 defines the target and Sec. A.13.2 the evaluation protocol.

A.13.1. The Thorlabs Distortion Target

The test target is the Thorlabs R3L3S7N1 negative grid distortion target [62] (Fig. A1): 17×17 clear dots of $0.5\,\mathrm{mm}$ diameter at $5\,\mathrm{mm}$ spacing in a low-reflectivity chrome coating on a $90\,\mathrm{mm} \times 90\,\mathrm{mm}$ soda-lime glass plate, with the dot diameter, the spacing and the corner-to-corner distance within $\pm 5\,\mu\mathrm{m}$ (Tab. A3). Its dimensions, from the manufacturer's drawing, define the target points $\boldsymbol{X}$; no part of the target is estimated. A tolerance of $\pm 5\,\mu\mathrm{m}$ is far below the size of one pixel's footprint on the target at our capture distances, so the target geometry can be treated as exact.

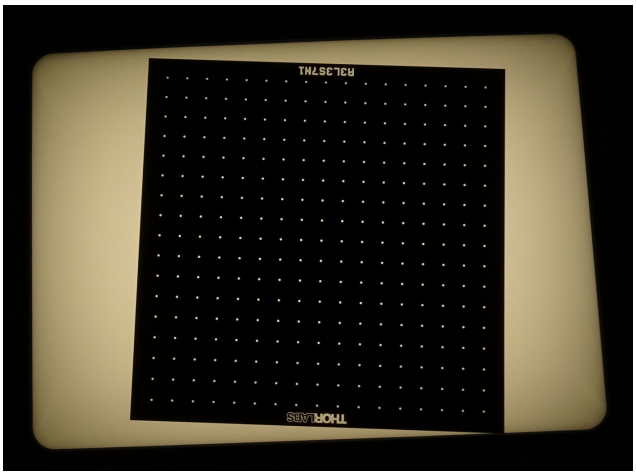


Figure A1. The Thorlabs R3L3S7N1 target on its LED backlight, from our capture.

A.13.2. Evaluation Protocol

Capture. The target is backlit by an LED panel in a dark room, so the dots are the only light in the image. The target is kept fixed, and the camera, mounted on a ball head on a pivot arm, is rotated between the images so that the dots cover the whole image, the rim included. The distance to the target is in the same range as for the Siemens-star and AprilGrid captures of the same lens, and the aperture is the same ($f/22$; $f/32$ for the fisheye at $30\,\mathrm{cm}$), so the defocus is comparable.

Detection. We detect the dots with a Förstner blob detector on the full-resolution green image, in which each red and blue photosite is interpolated bilinearly from its green neighbours (the image every geometric detector of our pipeline uses, the Siemens-star front end included), [25] and take the blob centroid as the image of the dot centre. The imaged dots cover a median of 12 to 34 px^2 per set (diameters of 4 to 7 px, full-resolution pixels) across the image set, so we assume that identifying the blob centroid with the dot centre introduces negligible bias. The imaged dots cover a median of 12 to 34 px^2 per set (diameters of 4 to 7 px), so identifying the blob centroid with the dot centre introduces negligible bias.

Correspondences. In each image, the detected dot centres are assigned to the 17×17 grid nodes once, before any model is evaluated: a RANSAC pose of the target with a one-to-one (Hungarian) assignment, in which a detection farther than half the grid spacing ($2.5\,\mathrm{mm}$ on the target) from every node stays unassigned. These correspondences are the same for every model, and no model rejects dots afterwards.

lens	target	role	distance [mm]	$f/$	exposure	captured	used
Geometric accuracy (Sec. 5.2)							
Sigma 14 mm	Siemens stars	train	295	22	1.6 s	168	132
Sigma 14 mm	AprilGrid	train	–	22	2 s	408	399
Sigma 14 mm	Thorlabs	test	300	22	1/10 s	142	141
Sigma 8 mm fisheye	AprilGrid	train	186	22	2 s	407	249
Sigma 8 mm fisheye	Thorlabs	test	121	22	1/13 s	154	154
Sigma 8 mm fisheye	Siemens stars	train	–	32	3.2 s	301	162
Sigma 8 mm fisheye	Thorlabs	test	–	32	1/10 s	254	82
Tamron 20 mm	Siemens stars	train	189	22	1.3 s	264	209
Tamron 20 mm	Thorlabs	test	197	22	1/60 s	112	112
Tamron 20 mm	AprilGrid	train	–	22	2.5 s	411	404
Tamron 20 mm	Thorlabs	test	448	22	1/60 s	142	142
Vignetting (Sec. 5.3)							
Sigma 14 mm	Siemens stars	train	762	4	1/8 s	174	166
Sigma 14 mm	flat field	reference	–	4	1/640 s	1	1
Sigma 8 mm fisheye	Siemens stars	train	693	4	1/10 s	160	63
Sigma 8 mm fisheye	flat field	reference	–	4	1/800 s	1	1
Tamron 20 mm	Siemens stars	train	–	2.8	1/50 s	163	138
Tamron 20 mm	flat field	reference	–	2.8	1/2000 s	1	1
Defocus (Sec. 5.4)							
Tamron 20 mm	Siemens stars	train	440	8	1/4 s	340	332
Tamron 20 mm	Thorlabs	test	–	8	1/800 s	190	190
zoom at 48 mm	Siemens stars	train	–	10	0.8 s	171	159
zoom at 48 mm	Thorlabs	test	–	10	1/500 s	54	54
Sigma 14 mm	Siemens stars	train	931	2.8	1/15 s	479	475
Sigma 14 mm	text board	test	–	2.8	1/15 s	106	106

Table A2. The captures of the data set, grouped by experiment, each training set followed by the test set captured with it. *Role* is the training/test split: training images calibrate the camera, test images (Thorlabs) only score it. *Captured*: the images of the set; *used*: those that enter the experiment, *i.e.* for a training set the images in which the geometric pipeline that starts our refinement localizes the board, and for a test set those whose dots are registered. The *used* counts are those of the starts before the geometric image changed to the full-resolution green image (Sec. A.13.2) and are to be regenerated. Distance: the median camera-to-target-plane distance inferred from the estimated poses (–: not yet computed). Of the 106 images of the text board, 63 fit the texture, 22 choose its regularization and 21 test it. *Flat field*: one image per lens through a diffuser (a white acrylic plate over the lens, hood on, lit by diffuse daylight), captured in the same session as the lens's vignetting set (2026-10-28); it is the reference the learned vignetting is compared with, not used in any calibration. The fisheye Siemens-star set of 2026-10-27 (150 images, 118 Thorlabs) is excluded: the start pipeline localized too few of its images.

grid	17×17 dots, 80×80 mm
dot spacing	5.000 mm
dot diameter	0.500 mm
tolerance*	$\pm 5\ \mu$m
substrate	soda-lime glass, 90×90 mm
coating	OD 3 chrome, $0.12\ \mu$m
polarity	clear dots, opaque field

Table A3. Specifications of the Thorlabs R3L3S7N1 target (manufacturer's drawing, 15 Nov 2024). *Of the dot diameter, the dot spacing and the corner-to-corner distance.

Motion. The camera model under test is held fixed. Since the camera rotates about one unknown pivot, a target point $\boldsymbol{X}$ is in view k at $\boldsymbol{R}_k(\boldsymbol{X} - \boldsymbol{c}) + \boldsymbol{d}$, with a rotation $\boldsymbol{R}_k$ per view and the centre $\boldsymbol{c}$ and the offset $\boldsymbol{d}$ shared by all views ($3N + 6$ parameters for N views, instead of $6N$ for free poses). They are fitted by minimizing the robust reprojection error. Constraining the motion in this way keeps the unknown poses from absorbing reprojection error at test time.

Train and test split. The grid nodes (i, j) are split by the parity of $i + j$ into two interleaved halves, like the colours of a checkerboard. The motion is fitted on one half and the reprojection error is measured on the other; the halves are then swapped and the errors of both are pooled. Every reported residual is therefore of a dot that did not enter the fit, and both halves cover the whole image, the rim included.

Statistics. We report the median, mean, 95th percentile and RMS of the reprojection error in pixels. A model

is evaluated on a dot only if it back-projects the dot and projects its fitted target point; the models of a set are compared on the dots that every one of them evaluates, so all rows of a set rest on the same dots. A model that is not defined over the whole image, *e.g.* a calibration restricted to the area its detections covered or a parametric distortion that does not reach the rim, leaves out the dots it cannot evaluate, which lie at the image border and are the hardest ones; the common dots then favour such a model. We therefore also computed each model's coverage, the share of the registered dots it evaluates, and its error on all of them. Every model of Tab. 2 evaluates at least 99.4 % of the registered dots, except the spline start on the fisheye's AprilGrid set (95.0 %), whose refinement, estimated over the whole lens coverage, evaluates 99.4 %. The median error of a model on all of its dots differs from that on the common dots by at most 0.002 px, and by at most 6 % on the fisheye's AprilGrid set, where the common dots leave out the border.

Confidence intervals. We compare two models of a set, *e.g.* a camera after the radiometric refinement and its geometric start, by the difference of a statistic of the reprojection error (median, mean, 95th percentile or RMS), each computed over the dots that every model of the set evaluates, pooled over all images. The interval of this difference is a paired, nonparametric bootstrap over images. The dots of one image are correlated through its pose, so the image is the resampling unit and not the dot. In each of 10,000 draws, as many images as the set contains are drawn with replacement; the statistic of both models is recomputed on the dots of the drawn images, an image drawn several times counting as often, and their difference is taken. Both models are thus always evaluated on the same images and dots, which removes the variation between images from the comparison. From the 10,000 differences we take the 95 % bias-corrected and accelerated (BCa) interval [20], whose bias correction is the share of draws below the observed difference and whose acceleration is estimated by a jackknife that leaves out one image at a time. The four statistics share the same draws, and the random seed is fixed, so the intervals are reproducible. A difference whose interval excludes zero is significant at the 5 % level. The intervals are computed over the whole image and, separately, over its centre and its edge (radii below 0.4 and above 0.6 of the largest radius).

Results. Tables A4 and A5 give the four statistics of Tab. 2 for every camera model, lens and calibration target, with the difference between the radiometric refinement and its geometric start and the 95 % interval of this difference.

A.14. Runtime

The radiometric refinement of the Sigma 14 mm Siemens-star set (132 photographs of 4 256 × 2 848 px, 16 384 samples each, 2 162 688 observations per iteration; 25 LM iterations over 2 273 free parameters: 1 443 of the camera, 792 of the poses, 29 of the target and 9 of the lighting) takes 703.5 ± 3.7 s (mean ± std over 8 runs, 26.1 s per iteration, writing the calibrated camera included) on one NVIDIA B200 with 8 CPU threads, from its loaded start; the decoding and the geometric start are excluded. Table A6 lists the size of the problem.

Sigma 14 mm, Siemens stars, central spline	
Photographs	132 at 4256 × 2848 px
Samples per photograph	16,384
Observations per iteration	2,162,688
Free parameters (camera / poses / target / lighting)	1,443 / 792 / 29 / 9
LM iterations	25
Time per iteration [s]	26.1
Refinement time [s] (mean ± std, min–max)	703.5 ± 3.7 (700.3–710.5)
Runs	8
GPU, CPU threads	NVIDIA B200, 8

Table A6. Runtime of one radiometric calibration: the refinement of the central spline of the Sigma 14 mm lens on the Siemens-star set, from its loaded geometric start to the written calibrated camera, over 8 runs, each on its own GPU after an untimed warm-up run. The decoding of the photographs and the geometric start are not included.

	Median [px]			95th percentile [px]		
Model	Geom.	Radiom.	Δ [95 % interval]	Geom.	Radiom.	Δ [95 % interval]
Tamron 20 mm, AprilGrid (142 test images, 40,999 dots)						
Brown–Conrady	0.119	**0.067**	-0.052 [-0.062, -0.043]	0.356	**0.173**	-0.183 [-0.208, -0.161]
OpenCV-14	0.107	**0.060**	-0.048 [-0.059, -0.037]	0.373	**0.155**	-0.217 [-0.270, -0.180]
Kannala–Brandt	0.117	**0.106**	-0.011 [-0.020, -0.001]	0.446	**0.285**	-0.162 [-0.220, -0.102]
Double Sphere	0.261	*0.286*	$+0.025$ [$+0.000$, $+0.046$]	1.70	**1.16**	-0.545 [-0.773, -0.294]
Central spline	0.040	*0.048*	$+0.008$ [$+0.006$, $+0.010$]	0.131	**0.119**	-0.012 [-0.027, -0.002]
Tamron 20 mm, Siemens stars (112 test images, 32,170 dots)						
Brown–Conrady	0.288	**0.135**	-0.154 [-0.179, -0.132]	0.855	**0.321**	-0.534 [-0.589, -0.480]
OpenCV-14	0.308	**0.131**	-0.177 [-0.203, -0.154]	0.969	**0.337**	-0.633 [-0.704, -0.566]
Kannala–Brandt	0.183	0.176	-0.007 [-0.022, $+0.005$]	0.730	**0.515**	-0.215 [-0.264, -0.164]
Double Sphere	0.219	*0.268*	$+0.049$ [$+0.036$, $+0.062$]	0.717	**0.625**	-0.092 [-0.141, -0.040]
Central spline	0.161	**0.136**	-0.026 [-0.034, -0.017]	0.436	**0.331**	-0.105 [-0.128, -0.084]
Sigma 14 mm, AprilGrid (141 test images, 40,744 dots)						
Brown–Conrady	0.499	**0.377**	-0.122 [-0.144, -0.099]	2.07	**1.78**	-0.289 [-0.360, -0.211]
OpenCV-14	0.476	**0.359**	-0.117 [-0.146, -0.092]	2.16	**1.74**	-0.419 [-0.543, -0.290]
Kannala–Brandt	0.456	**0.351**	-0.105 [-0.125, -0.084]	1.99	**1.79**	-0.202 [-0.308, -0.122]
Double Sphere	1.42	1.34	-0.081 [-0.170, $+0.000$]	5.13	**4.42**	-0.707 [-1.171, -0.334]
Central spline	0.479	**0.331**	-0.148 [-0.189, -0.114]	2.45	**1.76**	-0.691 [-0.895, -0.539]
Sigma 14 mm, Siemens stars (141 test images, 40,744 dots)						
Brown–Conrady	0.687	**0.359**	-0.328 [-0.375, -0.284]	2.73	**1.72**	-1.009 [-1.253, -0.790]
OpenCV-14	0.600	**0.384**	-0.217 [-0.270, -0.170]	2.87	**1.93**	-0.942 [-1.200, -0.703]
Kannala–Brandt	0.567	**0.329**	-0.238 [-0.277, -0.199]	2.37	**1.75**	-0.620 [-0.813, -0.392]
Double Sphere	0.734	**0.461**	-0.273 [-0.315, -0.232]	2.42	**1.94**	-0.484 [-0.708, -0.323]
Central spline	0.504	**0.366**	-0.137 [-0.174, -0.104]	2.35	**1.97**	-0.375 [-0.560, -0.229]
Sigma 8 mm fisheye, AprilGrid (153 test images, 40,928 dots)						
OpenCV-14	1.98	**1.70**	-0.285 [-0.406, -0.177]	9.15	**5.68**	-3.472 [-3.825, -3.029]
Kannala–Brandt	2.00	**1.52**	-0.479 [-0.560, -0.395]	6.80	**5.43**	-1.371 [-1.564, -1.207]
Double Sphere	2.09	**1.49**	-0.600 [-0.671, -0.538]	7.41	**5.45**	-1.965 [-2.126, -1.842]
Central spline	1.99	**1.45**	-0.543 [-0.654, -0.433]	7.58	**5.51**	-2.076 [-2.491, -1.702]
Sigma 8 mm fisheye, Siemens stars (82 test images, 22,973 dots)						
OpenCV-14	0.123	**0.079**	-0.044 [-0.065, -0.025]	0.818	**0.440**	-0.378 [-0.584, -0.224]
Kannala–Brandt	0.113	**0.063**	-0.050 [-0.066, -0.037]	0.590	**0.254**	-0.336 [-0.437, -0.237]
Double Sphere	0.109	**0.067**	-0.042 [-0.059, -0.029]	0.591	**0.289**	-0.302 [-0.391, -0.238]
Central spline	0.121	**0.100**	-0.020 [-0.039, -0.002]	0.996	**0.538**	-0.458 [-0.717, -0.309]

Table A4. Median and 95th percentile of the reprojection error on the held-out target at the geometric start and after the radiometric refinement, with their difference Δ and its 95 % bootstrap interval. Bold: significantly lower than the start; italic: significantly higher.

Model	Mean [px]			RMS [px]		
	Geom.	Radiom.	Δ [95 % interval]	Geom.	Radiom.	Δ [95 % interval]
Tamron 20 mm, AprilGrid (142 test images, 40,999 dots)						
Brown–Conrady	0.147	**0.077**	−0.069 [−0.080, −0.059]	0.185	**0.092**	−0.093 [−0.110, −0.081]
OpenCV-14	0.144	**0.070**	−0.075 [−0.090, −0.063]	0.201	**0.083**	−0.118 [−0.145, −0.098]
Kannala–Brandt	0.159	**0.124**	−0.034 [−0.052, −0.020]	0.214	**0.149**	−0.065 [−0.092, −0.043]
Double Sphere	0.473	**0.429**	−0.044 [−0.085, −0.007]	0.875	**0.696**	−0.179 [−0.228, −0.128]
Central spline	0.051	*0.055*	+0.003 [+0.001, +0.006]	0.067	0.065	−0.002 [−0.007, +0.001]
Tamron 20 mm, Siemens stars (112 test images, 32,170 dots)						
Brown–Conrady	0.350	**0.151**	−0.199 [−0.225, −0.177]	0.430	**0.176**	−0.253 [−0.282, −0.228]
OpenCV-14	0.386	**0.152**	−0.234 [−0.261, −0.209]	0.483	**0.181**	−0.302 [−0.333, −0.273]
Kannala–Brandt	0.257	**0.213**	−0.044 [−0.061, −0.028]	0.342	**0.261**	−0.081 [−0.103, −0.060]
Double Sphere	0.279	0.296	+0.017 [−0.001, +0.033]	0.358	0.344	−0.013 [−0.036, +0.008]
Central spline	0.190	**0.153**	−0.037 [−0.045, −0.029]	0.230	**0.180**	−0.050 [−0.060, −0.041]
Sigma 14 mm, AprilGrid (141 test images, 40,744 dots)						
Brown–Conrady	0.716	**0.568**	−0.148 [−0.168, −0.128]	0.997	**0.825**	−0.172 [−0.201, −0.146]
OpenCV-14	0.712	**0.544**	−0.168 [−0.198, −0.142]	1.02	**0.791**	−0.233 [−0.284, −0.191]
Kannala–Brandt	0.671	**0.542**	−0.128 [−0.147, −0.110]	0.956	**0.802**	−0.154 [−0.185, −0.126]
Double Sphere	1.95	**1.80**	−0.145 [−0.237, −0.061]	2.88	**2.59**	−0.298 [−0.387, −0.210]
Central spline	0.770	**0.530**	−0.240 [−0.288, −0.199]	1.22	**0.797**	−0.428 [−0.525, −0.343]
Sigma 14 mm, Siemens stars (141 test images, 40,744 dots)						
Brown–Conrady	0.968	**0.546**	−0.422 [−0.485, −0.367]	1.42	**0.796**	−0.625 [−0.761, −0.512]
OpenCV-14	0.942	**0.598**	−0.344 [−0.405, −0.288]	1.50	**0.884**	−0.611 [−0.736, −0.503]
Kannala–Brandt	0.825	**0.515**	−0.310 [−0.360, −0.263]	1.23	**0.767**	−0.459 [−0.570, −0.361]
Double Sphere	0.926	**0.650**	−0.276 [−0.324, −0.235]	1.17	**0.905**	−0.262 [−0.333, −0.206]
Central spline	0.780	**0.594**	−0.186 [−0.235, −0.148]	1.20	**0.901**	−0.301 [−0.403, −0.226]
Sigma 8 mm fisheye, AprilGrid (153 test images, 40,928 dots)						
OpenCV-14	2.98	**2.18**	−0.801 [−0.963, −0.651]	4.14	**2.80**	−1.336 [−1.536, −1.144]
Kannala–Brandt	2.59	**2.01**	−0.578 [−0.649, −0.506]	3.33	**2.65**	−0.682 [−0.750, −0.614]
Double Sphere	2.75	**1.99**	−0.763 [−0.830, −0.698]	3.59	**2.64**	−0.949 [−1.023, −0.878]
Central spline	2.70	**1.96**	−0.746 [−0.868, −0.638]	3.69	**2.63**	−1.059 [−1.306, −0.897]
Sigma 8 mm fisheye, Siemens stars (82 test images, 22,973 dots)						
OpenCV-14	0.280	**0.164**	−0.115 [−0.188, −0.077]	2.62	**1.09**	−1.533 [−2.953, −0.132]
Kannala–Brandt	0.182	**0.089**	−0.093 [−0.122, −0.070]	0.265	**0.123**	−0.141 [−0.181, −0.107]
Double Sphere	0.181	**0.098**	−0.083 [−0.109, −0.064]	0.265	**0.151**	−0.114 [−0.148, −0.086]
Central spline	0.256	**0.182**	−0.074 [−0.113, −0.042]	0.472	0.502	+0.029 [−0.153, +0.342]

Table A5. Mean and RMS of the reprojection error, as Tab. A4.