IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, October 2018.

Keyframe-based Photometric Online Calibration and Color Correction

Jan Quenzel, Jannis Horn, Sebastian Houben and Sven Behnke

***Abstract*— Finding the parameters of a vignetting function for a camera currently involves the acquisition of several images in a given scene under very controlled lighting conditions, a cumbersome and error-prone task where the end result can only be confirmed visually. Many computer vision algorithms assume photoconsistency, constant intensity between scene points in different images, and tend to perform poorly if this assumption is violated. We present a real-time online vignetting and response calibration with additional exposure estimation for global-shutter color cameras. Our method does not require uniformly illuminated surfaces, known texture or specific geometry. The only assumptions are that the camera is moving, the illumination is static and reflections are Lambertian. Our method estimates the camera view poses by sparse visual SLAM and models the vignetting function by a small number of thin plate splines (TPS) together with a sixth-order polynomial to provide a dense estimation of attenuation from sparsely sampled scene points. The camera response function (CRF) is jointly modeled by a TPS and a Gamma curve. We evaluate our approach on synthetic datasets and in real-world scenarios with reference data from a Structure-from-Motion (SfM) system. We show clear visual improvement on textured meshes without the need for extensive meshing algorithms. A useful calibration is obtained from a few keyframes which makes an on-the-fly deployment conceivable.**

I. Introduction

Vignetting, i.e., the difference in intensity for equally bright scene points in different parts of the image, is an undesirable property of most dioptric camera systems. It is caused by a non-uniform exposure of different points on the photoelectric chip as a fraction of the light that passed through the lens is blocked by the aperture, or in some cases by another set of lenses. The effect can differ substantially between different lens systems or cameras and may—even if clearly present—be neglected for many applications.

Auto exposure enables cameras to provide useful images under changing lighting conditions. The sensor is exposed over a longer or shorter period of time to prevent images from becoming too dark or bright.

However, when it comes to quantitative image processing, scene reconstruction, optical flow, or SLAM algorithms, one important assumption of many methods is constant intensity for the same scene points. In fact, most applications with moving cameras—including most from the field of robot vision—should be designed with this phenomenon in mind, both to increase the reliability and reducing the algorithm complexity due to normalized input data.

All authors are with the Autonomous Intelligent Systems Group, University of Bonn, Germany {quenzel, houben, behnke}@ais.uni-bonn.de This work was supported by grants BE 2556/7 of the German Research Foundation (DFG) and 608849 of the European Union's 7th FP.

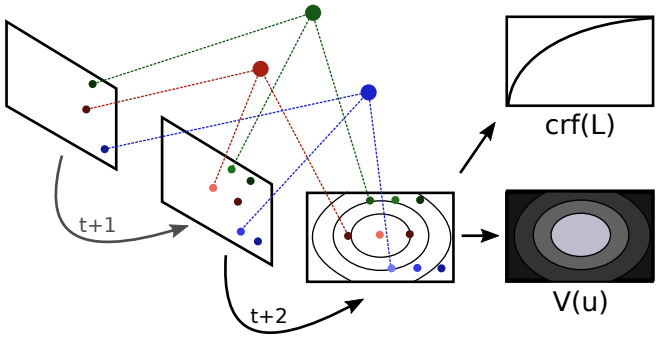


Fig. 1: Calibration principle: Points from the scene (shown in red, blue and green) are triangulated by minimizing the backprojection error over several frames. Due to vignetting $V(u)$, the measured intensity of the corresponding image points will vary depending on the position $\mathbf{u}$ within the frame while the brightness of the entire image is affected by different exposure times k and the camera response function $crf(L)$. The observed attenuations give rise to a camera response and vignetting function.

To correct vignetting, in most cases, a static calibration is computed beforehand, which requires recording an image of a known texture, usually plain white, and measuring the difference in intensity for different image points. The lighting conditions have to be controlled very thoroughly: the texture must be uniformly illuminated and no shadows, e.g., created by the camera itself, must be present. One should also point out that changing the aperture or focus of the camera lens will also affect the vignetting, thus, after calibration these characteristic must not be changed anymore.

The mapping between the radiance $L_{\mathbf{p}}$ of a scene point $\mathbf{p} \in \mathbb{R}^3$ and the measured intensity $I_{\mathbf{u}}$ of the corresponding image point $\mathbf{u} \in \mathbb{R}^2$ is often modeled as a two-part function

$$I_{\mathbf{u}} = f\left(k \cdot V\left(\mathbf{u}\right) \cdot L_{\mathbf{p}}\right). \tag{1}$$

Here, $V : \mathbb{R}^2 \mapsto [0;1]$ is the position-dependent vignetting function, k is the exposure time, and f is the camera response function (CRF) that covers the mapping between the amount of light reaching the chip and its corresponding measurement. For simplicity, V is often considered radially symmetric around the optical center of the image. We do not restrict ourselves to this assumption. In this work, our objective is to estimate all involved photometric parameters.

To this end, we make use of a robust visual SLAM procedure and examine the recorded intensity of well-established map points. Stable triangulation requires the points to be recorded from several camera positions with sufficient parallax, thus, the corresponding image points are spread over

some portion of the camera frames. The intensity of a map point in different regions of the image yields samples for the vignetting function V that is extrapolated with the help of a thin plate spline and a sixth-order polynomial. This approach, illustrated in Fig. 1, allows us to quickly obtain a reliable estimate of V based on only a few map points. It does not rely on any known illumination pattern or scene appearance and can be performed by recording natural, albeit textured, environments with static illumination. We demonstrate our method on an image sequence taken with a micro aerial vehicle (MAV) to reconstruct a chimney wall structure as well as synthetic sequences with artificial photometric disturbances and show that it significantly improves the results.

II. Related Work

There are two main approaches for vignetting correction: The correction can be estimated from a single image or from a sequence of images at different view poses with overlapping fields of view. The first approach has been recently explored [1], [2]. We prefer the latter approach since it reduces the number of necessary assumptions to obtain a viable solution, is more robust, and enables us to obtain all photometric parameters. Historically, this approach was first used for image mosaicking and panoramic photography, in which images are stitched together using visual features such that no seam between images is visible. Simple vignetting functions, like 6th degree polynomials, are often employed [3], [4] or images are blended in overlapping regions [5].

More recent work from Waechter et al. [6] aims at adding textures to a reconstructed mesh. The view that best observed a mesh face with minimal seam towards neighboring faces is selected using graph cuts. Afterwards, the color is adjusted to reduce color differences between patches, first globally and later locally, via Poisson Editing to remove remaining visible seams close to the seam itself. A different approach was taken by Zhou and Koltun [7]. Instead of selecting best-view patches, they optimize the camera poses such that mesh vertices observed in multiple cameras have the same intensity. To deal with more complex distortions, this optimization was augmented by a per-camera grid on the image plane that non-rigidly deforms the image. The final per-vertex color is then obtained using a weighted mean of the observed colors. Yet, mesh faces are not colored, but an improved resolution is obtained via subsampling the mesh to generate more vertices.

The photometric or brightness constancy assumption [8] is used by most direct Visual Odometry systems, e.g., Semi-direct Visual Odometry (SVO) by Forster et al. [9] or Direct Sparse Odometry (DSO) by Engel et al. [10]. Since vignetting and auto-exposure violate the constancy assumption, incorporating photometric calibration improves the accuracy of direct methods as reported by Zheng et al. [11].

Complementary to DSO, Engel et al. [12] created a monocular camera benchmark, including photometric calibration of an industry-grade camera. The CRF was calibrated from a set of 1000 images taken by a statically placed camera while manually changing the exposure time in-between. Vignetting was disregarded at first, and calculated later on from images of a bright colored planar wall with an attached AR marker and approximately Lambertian reflectance. The marker allows them to estimate the camera pose w.r.t. the planar wall. The wall is divided into 1000×1000 points. Given the camera pose, the points can be reprojected into the images to obtain the corresponding intensities. The vignetting is then calibrated together with the unknown irradiance of the wall in an alternating fashion using a Maximum-Likelihood Estimator.

A single white sheet of uniformly illuminated paper is used by Alexandrov et al. [13] as a calibration target for consumer RGB-D sensors. First, the CRF is estimated using the OpenCV implementation of [14]. Then the paper is fixed and the exposure is set to a constant value in order to obtain a bright, yet not overexposed image. Automatic white balance should be disabled before moving the camera around the object to capture enough observations. Illumination is assumed constant for the whole piece of paper, thus, intensity differences result purely from vignetting. The paper is detected via floodfill segmentation and no projection is necessary. After application of the inverse CRF, normalization and inversion of each entry, dense correction factors are obtained. For comparison, Alexandrov et al. [13] evaluated the sixth-order polynomial used by Goldman and Chen [3] and showed the superiority of dense factors.

For field applications, in which a tagged wall or a sheet of paper on a planar table are not easily accessible, the conditions often differ from laboratory settings and focus or aperture have to be adjusted. Hence, tedious recalibration using one of the previous methods is required.

In our previous work on vignetting calibration [15] we used the projection of map points to estimate a sensor-dependent deformation using a thin plate spline. The method worked online and is complementary to further steps needed during reconstruction or dense tracking. There we replaced the known calibration target with an arbitrary, textured environment without any geometric constraints.

Recently, Bergmann et al. [16] optimized for the photometric model as well as the radiance and exposure times in an online calibration method. A KLT tracker was employed to find corresponding patches between consecutive images. The CRF is approximated with the EMoR-model and the polynomial by Goldman and Chen is used for vignetting. The authors separate the parameter estimation into fast exposure estimation, photometric model estimation and radiance estimation. The exposure estimation uses a window of ten images compensated for CRF and vignetting. The radiance within this window is approximated as the mean of the corrected intensities. The model and the radiance estimates are updated in parallel using a window of the last 200 images. The method showed a significant accuracy gain for DSO. However, the authors rely on a completely independent keypoint and motion tracking approach that does not directly benefit the SLAM results but that instead corrects images as input into a fully separate SLAM framework. Our approach, on the other hand, optimizes all parameters jointly and

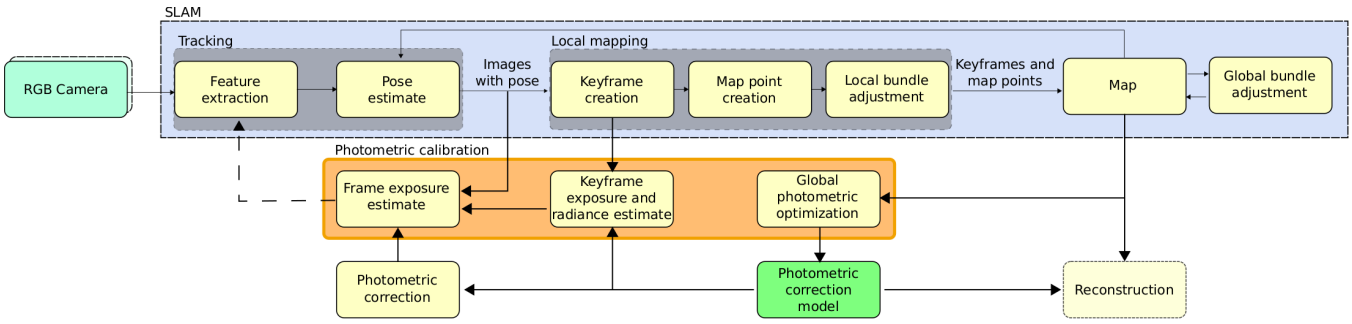


Fig. 2: Method overview: Visual features are tracked from (Stereo-) RGB images in order to estimate the camera pose w.r.t. the map points. Keyframes and triangulated map points are computed by the local mapping and stored in the map. The photometric calibration (orange dashed box) estimates the exposure time for each frame given the current photometric model (green) and the tracked map. The radiance estimate of the map is updated on keyframe creation. The photometric model (CRF and vignetting), keyframe exposure times and mappoint radiance are refined given all matched keyframe observations during global photometric optimization. Sparsely, distributed features are used to estimate a pixel-wise vignetting correction based on a thin plate spline that can be used to correct future images and reconstruction results.

integrates the photometric calibration more straightforwardly and, in particular, avoids two redundant pose estimation and point tracking procedures.

In this work, we model vignetting by a sixth-order polynomial in combination with a smooth thin plate spline to acquire pixel-wise correction factors. The camera response function is modeled as another thin plate spline with border conditions. We reduce the necessary amount of data by estimating the photometric model on a low number of keyframes while estimating the exposure time of the current frame w.r.t. the tracked map. In summary, the key features of our method are:

- oriented patches around ORB-features from visual SLAM are used,
- thin plate splines combined with a Gamma curve models the camera response function,
- a sixth-order polynomial captures the basic shape of the vignetting whilst local deformation is captured by a thin plate spline,
- the thin plate spline allows for dense approximation of the sparsely distributed correction factors,
- joint optimization of radiance, vignetting, and camera response is performed on keyframes only,
- our algorithm can be used online in real-time on a modern laptop CPU, and
- handling of natural and partially dynamic scenes without uniform illumination.

III. Method

We use an ORB-feature-based Graph-SLAM system [17] with keyframes to obtain the color camera trajectory and triangulate a sparse feature map from visual features. We prefer to use a synchronized stereo camera rig to perform SLAM instead of a monocular camera since the absolute scale is fixed and we avoid monocular scale drift.

We use established feature-to-map-point correspondences for our photometric calibration and generate a larger set of samples by extracting patches around the individual features on their respective scale. Purely black and white pixels are considered invalid and are discarded. Still, the obtained correspondences are sparsely distributed on the sensor. Hence, we model the vignetting function as a smooth thin plate spline (TPS), which allows us to estimate the dense attenuation factors for each pixel and color channel. The vignetting TPS is enhanced with an even sixth-order polynomial for higher accuracy. Another one-dimensional TPS is used for the camera response function with border conditions.

Subsequently, the exposure time of a frame is estimated given the current map. This requires fast radiance updates. Hence, we directly refine the radiance of updated map points after keyframe creation while optimizing the photometric model asynchronously on keyframes.

After introducing our notation, we motivate made assumptions and describe in detail the estimation of the attenuation factors and the correction models using TPS.

A. Notation

We denote sets and matrices with capital letters and vectors with bold lower case letters. Each map point $\mathbf{p}_w = (x, y, z)^\intercal \in \mathbb{R}^3$ is defined in the world frame w, determined by the initial color camera frame. All poses are represented by a transform $T_{F_2F_1} \in \mathbf{SE}(3)$, which maps a point $\mathbf{p}_{F_1}$ from the frame F_1 into the frame F_2. For convenience of notation, we identify $T_{F_2F_1}$ with its 4×4 matrix operating on homogeneous coordinates. The projection of a point $\mathbf{p}_w$ with pose T_F and camera matrix K_F into frame F yields the image coordinates $\mathbf{u} = (u_x, u_y)_F^\intercal$ in the image domain $\Omega \subset \mathbb{R}^2$ via the mapping:

$$g_F(\mathbf{p}_w) : \mathbf{p}_w \to \mathbf{p}_F, \tag{2}$$

$$(\mathbf{p}_F, 1)^\intercal = T_{Fw} \cdot (\mathbf{p}_w, 1)^\intercal, \tag{3}$$

$$\pi_F(\mathbf{p}_F) : \mathbf{p}_F \to \mathbf{u}_F, \tag{4}$$

$$(x, y, z)_F^\intercal = K_F \cdot \mathbf{p}_F, \tag{5}$$

$$\mathbf{u}_F = (x/z, y/z)^\intercal. \tag{6}$$

B. Assumptions

We assume that we observe a static scene with Lambertian reflectance such that the amount of reflected light is independent of the viewing angle. The illumination within the scene should not change over time but may differ locally in the observed scene, i.e., we do not assume a uniformly lit scene. Since we can only obtain the attenuation correction factor for each pixel up to scale, we assume the values to be within $[0, 1]$. We further assume similar attenuation between neighboring pixels.

Obviously, we will need some texture in the images to extract visual features. Furthermore, we assume a given rough factory calibration for the intrinsic camera matrices K_c, lens distortion, and the extrinsic transform between the stereo cameras.

C. Attenuation Factor Estimation

After a part of the scene has been explored and the camera trajectory has been successfully tracked, a global bundle adjustment refines the camera poses T_{cw} and triangulated map point positions $\mathbf{p}_w$ from undistorted feature observations. In a second optimization step, we include refinement of lens distortion and intrinsic parameters given the original feature observations while fixing the extrinsic transformation between a stereo camera pair or keeping the first two poses fixed. Thereby, we obtain a more accurate estimate from a factory calibration, which allows us to establish further correspondences between keyframes which have been previously discarded due to high reprojection errors.

The basic idea to obtain pairs of expected and measured intensity, as visualized in Fig. 1, is to either project map points into the color image and compare the estimated radiance using Eq. 1 against the measured intensity $I_{\mathbf{u}}$ or directly use the matched keypoints to generate corresponding intensities. We cannot rely on gathering a high number of correspondences per pixel if we want to run our method online. Having multiple pairs for one pixel rarely happens for sparsely distributed map points. Hence, the TPS is a convenient method to interpolate vignetting in-between.

D. Attenuation Model

We use TPS to model local attenuation factors w.r.t. the normalized color image coordinates from $[0, 1]^2$. Due to the excellent fill-in property and the minimal bending energy of these splines, this works even with scattered, sparsely distributed data—in our case the correction factors and corresponding image positions—while giving smooth function approximations with a small number of coefficients. We use the following two-dimensional thin plate polyharmonic spline:

$$h(\mathbf{u}) = p\left(\mathbf{u}\right) + \sum_{i=1}^{N} c_i \cdot \phi\left(\left\|\mathbf{u} - \mathbf{d}_i\right\|\right) \tag{7}$$

with the radial basis function (RBF)

$$\phi\left(r\right) = r^2 \cdot \ln\left(r\right), \tag{8}$$

and the polynomial

$$p\left(\mathbf{u}\right) = \mathbf{v}^\intercal \cdot \begin{pmatrix} 1 \\ \mathbf{u} \end{pmatrix}. \tag{9}$$

Here, $\mathbf{u}$ is the data point—in our case a pixel coordinate—and $\mathbf{d}_i \in \Omega$ is a control point within the image. The parameters $\mathbf{c}$ control the influence of the RBF while $\mathbf{v}$ aids the approximation as a polynomial. One advantage of the TPS is the lack of parameters that have to be tuned since $\mathbf{c}, \mathbf{v}$ are calculated from the given image positions $\mathbf{u}$ and the desired function values, the correction factors $s_{\mathbf{u}}$. Furthermore, TPS is far more flexible compared to a polynomial with the same number of coefficients.

In case of interpolation, one seeks to find the coefficients $[\mathbf{c}, \mathbf{v}]^\intercal$ s.t. the following equations are satisfied:

$$s_i = h\left(\mathbf{u}_i\right), 1 \leq i \leq M. \tag{10}$$

Since the interpolation would require as many RBFs (N) as there are data points (M), this cannot be used efficiently online. Instead, we approximate the underlying function using a grid with a small fixed number of $N = P^H \times P^V$ control points:

$$\underset{\mathbf{c},\mathbf{v}}{\operatorname{argmin}} \sum_{i}^{M} \left\|h\left(\mathbf{u}_i\right) - s_i\right\|^2. \tag{11}$$

On each control point $\mathbf{d}_i$, one RBF is placed statically. We typically choose $P^H, P^V \in \{3, 4, 5, 6, 7\}$, but other choices and different grids are possible as well. Often the following conditions are added:

$$\sum\nolimits_i^N c_i = 0, \sum\nolimits_i^N c_i \cdot d_{i,x} = 0, \sum\nolimits_i^N c_i \cdot d_{i,y} = 0, \tag{12}$$

which ensure that the polynomial p can be approximated.

In regard of the vignetting correction, we replace the polynomial in Eq. 7 with a sixth-order even polynomial of Goldman and Chen [3]:

$$p\left(r\right) = 1 + \sum\nolimits_{i=1}^{3} v_i \cdot r^{2i}, \tag{13}$$

$$h\left(\mathbf{u}\right) = p\left(\sqrt{2}\left\|\mathbf{u} - \mathbf{d}_m\right\|\right) + \sum_{i=1}^{N} c_i \cdot \phi\left(\left\|\mathbf{u} - \mathbf{d}_i\right\|\right). \tag{14}$$

The polynomial p matches the general form of the vignetting function whilst higher-order and local deformations are covered by the RBFs. $\mathbf{d}_m$ is the center position within the unit square and the factor $\sqrt{2}$ normalizes the radius to $[0, 1]$.

E. Camera Response Function

We employ a one-dimensional thin plate spline with a linear function (9) for the camera response. The P control points are equidistantly distributed between zero and one. Since the CRF f needs to interpolate from zero to one, we add constraints that enforce $f(0) = 0$ and $f(1) = 1$. This corresponds to a PDE with Laplace Equation and Dirichlet boundary conditions. Solving a PDE using thin plate splines can be performed by solving a linear system of equations.

The fitting accuracy of thin plate splines can be further improved by using a problem specific function for p instead

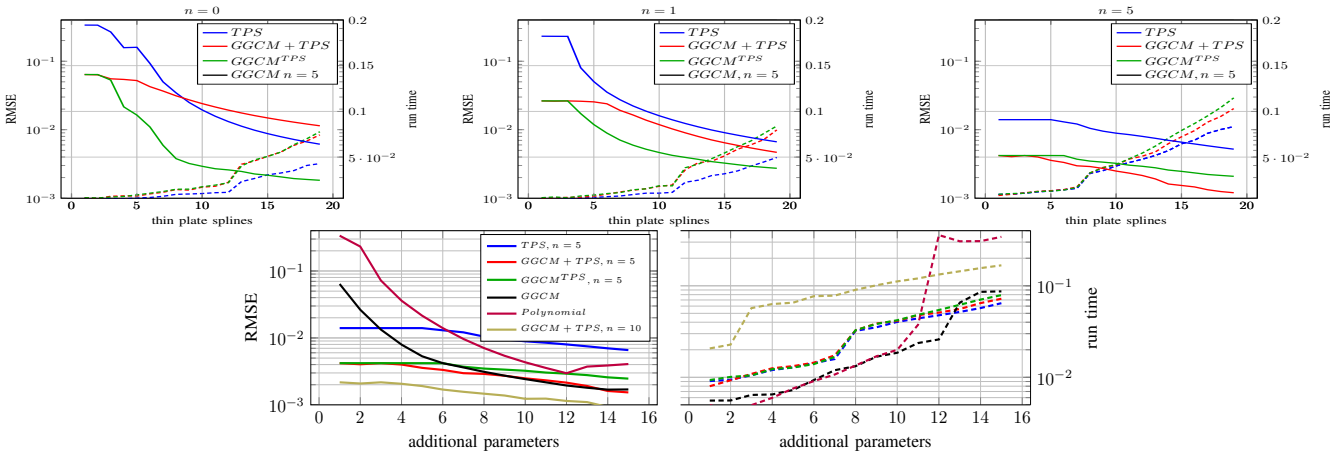


Fig. 3: Results of the proposed CRF-models. Top row from left to right: the polynomial degrees are 0, 1 and 5. Dashed curves showing runtime according to right-hand axis label. Bottom row: the RMSE and runtime for varying number of additional parameters.

of the polynomial (9). Hence, we employ the Generalized Gamma Curve Model (GGCM) [18]:

$$p_f(x) = \sum_{i=0}^{n} v_i \cdot x^i, \tag{15}$$

$$f_{GGCM}(L) = L^{p_f(L)}, \tag{16}$$

$$g_{GGCM}(I) = I^{1/p_f(I)}. \tag{17}$$

We propose two new CRF model functions based on the GGCM. Firstly, we use the model f and add a number of polyharmonic thin plate splines (this corresponds to replacing p in (7) with f_{GGCM}. We call this model $GGCM + TPS$. Secondly, we use the GGCM model with a more variable thin plate spline instead of the polynomial in (15). We name this model $GGCM^{TPS}$. We do not choose the function g_{GGCM} due to the inversion of the polynomial. The reader may note that it is not the correct inverse of f_{GGCM}, except for a pure gamma curve $(n = 0)$. Yet, given an estimate of the polynomial it is easy to calculate a corresponding g_{GGCM}. Once an updated CRF model is available, we sample the CRF equidistantly and optimize for the parameters of the inverse model f^{-1} using g_{GGCM}. So far, we have not seen the necessity to explicitly enforce monotonicity as the results were monotone with $P \in [5, 20]$ control points. An alternative method for monotone TPS is given in [19].

F. Image Correction

Given the solution to Eq. 11, we obtain the fitted TPS by evaluating Eq. 14 for each pixel. In order to remove the vignetting, the inverse camera response function needs to be applied to the pixel intensity, followed by multiplication with the inverse attenuation factor and exposure time:

$$I_{c,\mathbf{u}} = f^{-1}(I_{\mathbf{u}}) / [V(\mathbf{u}) \cdot k]. \tag{18}$$

Here, $I_{c,\mathbf{u}}$ denotes the corrected pixel intensity. A look-up-table (LUT) for $f^{-1} \in [0, 255]$ can be easily obtained due to the strict monotonicity of the CRF whereas the inverse attenuation factors require pixelwise evaluation of the TPS only once.

G. Keyframe-based Photometric Calibration

A photometric calibration is obtained and incrementally refined from a number of keyframes. We optimize the full parameter set similar to [16]. The difference between the measured image intensity $I_{\mathbf{u}}$ and the righthand side of Eq. 1 is jointly minimized for all involved parameters, given an initial guess for the radiance $L_{\mathbf{p}}$, the vignetting $V(\mathbf{u})$, the exposure time k and the camera response function f:

$$\underset{\mathbf{c},\mathbf{v},\mathbf{k},\mathbf{L_p}}{\operatorname{argmin}} \sum_{i,j,o=1}^{N,M,O} \rho_h \left(\left\| w_j \left[I_{\mathbf{u}_j} - f\left(k_i V(\mathbf{u}_j) L_{\mathbf{p}_o}\right) \right] \right\|^2 \right). \tag{19}$$

Here, N denotes the number of keyframes, M the number of observations, and O the number of map points. The Huber robust loss function ρ_h with $\alpha = 0.2/255$ is employed for robustness against outliers. An additional weighting term $w_j = \frac{\eta}{\eta + \left\| \nabla I_{\mathbf{u}_j} \right\|^2}$ with $\eta = 1$ downweighs high gradient pixels. We follow the suggestion to use a patch centered at the feature location. The feature orientation is used to extract an unrotated patch of size 5×5 at the corresponding keypoint scale. Saturated (255) and blank pixels (0) are removed before scaling the intensities to unit range.

The exposure time evaluation of individual frames is done using the current radiance estimation of the tracked map points:

$$\underset{k}{\operatorname{argmin}} \sum_{i=1}^{N} \rho_h \left(\left\| w_i \cdot \left[\frac{f^{-1}(I_{\mathbf{u}_i})}{V(\mathbf{u}_i)} - k \cdot L_{\mathbf{p}_i} \right] \right\|^2 \right). \tag{20}$$

This requires an always up-to-date estimate for the radiance of each map point. As full optimization is infeasible under real-time constraints, we refine the exposure time k of a keyframe and the radiance of all its tracked and newly created map points $L_{\mathbf{p}}$ on its creation. The full optimization (19) runs asynchronously after a constant number of keyframes were created.

TABLE I: Mean RMSE of vignetting, exposure and mean improvement of consistent feature matches on Monte Carlo sampled synthetically deteriorated sequence of ICL-NUIM [20] without loop closure.

Model	Radial polynomial	TPS	TPS + radial polynomial
Vignetting RMSE	0.05267	0.04728	**0.04617**
Exposure RMSE	0.04188	0.03473	**0.03345**
Improvement [%]	11.6	**13.25**	11.98

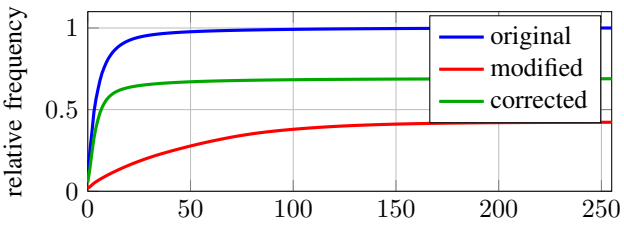


Fig. 4: Cumulative histogram of the pairwise Manhattan distances between fused points after dense reconstruction using COLMAP [21] on the ICL-NUIM dataset. A higher curve is better.

IV. Evaluation

We test our approach in different synthetic and real scenarios. RGB-D sequences from the synthetic dataset of ICL-NIUM [20] are modified to exhibit vignetting and sinusoidally varying exposure times. The real-world sequences are captured with a stereo rig attached to a micro aerial vehicle. The rig consists of two synchronized FLIR Blackfly S BFS-U3-51S5 color cameras with a resolution of 2448×2048 pixel. The cameras were calibrated using Kalibr [22]. We activated auto exposure without gain during our experiments and recorded the images at 22 Hz as well as the exposure time for comparison.

Our estimates will be compared on real-world sequences against the methods by Engel et al. [12], Alexandrov et al. [13], and Bergmann et al. [16]. These sequences were recorded in a hallway with stonework and in a lab. All calculations were performed on an Intel Core i7-6700 HQ with 32 GB RAM running Ubuntu 16.04. We start vignetting and CRF calibration after ten keyframes have been created and placed the control points on a 4×5 regular grid to incorporate the aspect ratio.

Camera Response Function: We evaluate our TPS-CRF on the 201 response curves of real-world cameras within the DoRF-Dataset [23]. We perform a least squares fit for each camera and evaluate the Root Mean Squared Error (RMSE). Minimization is implemented using Ceres-Solver [24].We limit the number of tested TPS parameters to 20 and the polynomial order to 15 and evaluate three different CRFs: TPS with polynomial p as well as with the GGCM (16), denoted as TPS+GGCM, and GGCM in which the polynomial is replaced with a TPS, denoted as $GGCM^{TPS}$. We report the RMSE and the running time in Fig. 3 for some configurations. The functions were always initialized to linear interpolate between zero and one.

We showed that all the presented models can successfully fit real-world camera response functions—even though the total number of parameters increases with additional thin plate splines. We have found that there exists a trade-off between the time it takes to fit a higher-order polynomial model and the number of TPS parameters. Hence, we can lower the polynomial degree by adding a number of TPS and obtain a better RMSE while using less time to fit the model.

We obtained the best results for the combination of TPS+GGCM followed by $GGCM^{TPS}$ while the classical TPS performed worst and took longest to optimize. The fastest results were obtained using $GGCM^{TPS}$. We attribute the accuracy of TPS+GGCM compared to only GGCM to the additional flexibility from the thin plate splines. The deficit of the original TPS stems from the polynomial, which is not an appropriate model for Gamma-like curves. Still, it reaches the same error as GGCM $(n = 15)$ with 60 control points while taking three times as long.

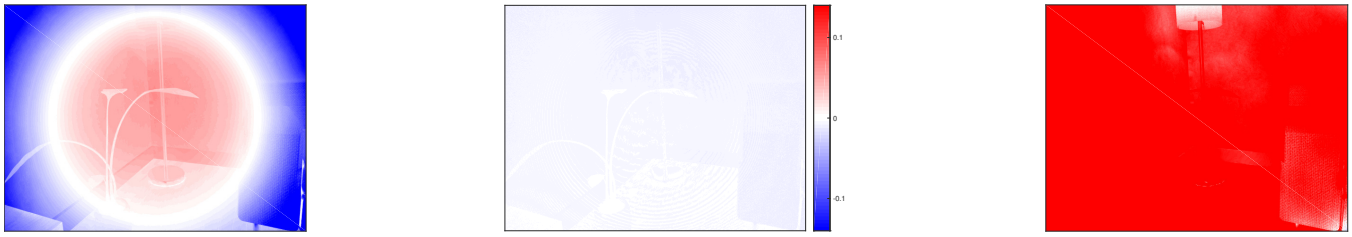
Vignetting Correction: We perform multiple experiments with different vignetting masks. The first vignetting mask is the ideal case with a pure sixth-order polynomial (13) that originates from the center of the image. The second vignetting mask has a randomly shifted origin while the third mask is slightly deformed by locally consistent noise. We compare the obtained results using a pure polynomial against using only TPS [15] and our combination (14).[1] We employ the multi-view stereo pipeline of COLMAP [21] to create a dense reconstruction given only the keyframes selected before. The reconstruction is run with the *original*, the *modified* and the *corrected* images. The modified images are the altered images exhibiting vignetting and synthetic sinusoidal exposure changes as described above. We applied our estimated correction on the modified images to obtain the corrected images. The difference in mean Manhattan distance between fused points is visualized in Fig. 4. A smaller distance is preferable since fused points are more similar. The reconstruction from corrected images follows the original graph closely for small differences (< 10) whereas the modified sequence exhibits larger differences.

We further deteriorated the vignetting by moving the vignetting origin away from the image center and added low spatial-frequency noise. We jointly optimized the vignetting and the exposure time on the same keyframes and computed the RMSE. We repeated this procedure one hundred times for the sixth-order radial polynomial (13), the original TPS (7) and our radial TPS (14). The results are reported in Tab. I. As expected our radial TPS reported the best results, followed by the original TPS and the radial polynomial. An improved vignetting estimate simultaneously reduces the difference between estimated and correct exposure time, but increasing the number of thin plate splines reduced the RMSE at the expense of increased optimization and run time. After optimization, we corrected all keyframes and recomputed matching features. We checked the consistency of all matches

[1] An accompanying video is available at `https://www.ais.uni-bonn.de/videos/IROS_2018_photometric_calibration`.

TABLE II: RMSE of all optimized parameters on synthetically deteriorated sequence of ICL-NUIM [20] w/o loop closure.

Model	CRF Vignetting	$GGCM^{TPS}$ Radial polynomial	$GGCM^{TPS}$ TPS	$GGCM^{TPS}$ TPS + radial polynomial	[16] Radial polynomial
RMSE	Exposure	0.0748510	0.0331069	**0.0292366**	0.1760950
	Vignetting	0.0558832	0.0380453	**0.0366498**	0.1029400
	CRF	0.0309349	0.0267781	**0.0209797**	0.1468000
$RMSE_{10}$	Exposure	0.0186301	**0.0126985**	0.0140462	0.0567196



(a) Modified (b) Corrected (our) (c) Corrected ([16])

Fig. 5: Difference between original and corrected image for image 858 from ICL-NUIM living room 2 [20]. Vignetting is evident in the modified image (left). Our correction (middle) successfully reduces vignetting, exposure change, and removes the response function. The estimated exposure is too high using the method of [16] (right).

using the ground truth poses. Surprisingly, the number of correct correspondences increased after correction by around $12\,\%$ and thereby improves the overall system accuracy.

Synthetic Datasets: A drawback of the previously mentioned TPS-CRF is its missing closed-form invertibility. Hence, we choose to use the $GGCM^{TPS}$ as our CRF model for the integrated tests on the synthetic and real-world datasets, which is also quite fast to optimize. The TPS uses five control points and a second-degree polynomial. We evaluate all three vignetting models and use a grid size of 4×5 for the TPS. The RMSE for the exposure ratio, CRF and vignetting is reported in Tab. II. Additionally, we evaluated the approach of Bergmann et al. [16] with default parameters and number of active frames set to the sequence length. We observed a strong drift in the exposure estimate. This is evident in Fig. 5 (right), where the difference between the original and the corrected estimates is visualized. Hence, we also report the $RMSE_{10}$ over a smaller window of ten frames. We attribute the improved results of our method to the joint optimization, in contrast to alternating between radiance and photometric parameters, as well as more robust keypoints.

Real-World Datasets: We followed the prescribed calibration procedures for the methods by Engel et al. [12] and Alexandrov et al. [13]. During our tests, we found the white-paper method to be sensible to lighting conditions, e.g., mixtures of artificial and natural light while the method by Engel et al. may produce non-monotonic camera response functions. The corresponding inverse CRF curves are visualized in Fig. 7. Fig. 6 shows the reported exposure ratio on the lab sequence for the first camera in our stereo rig and our estimated exposure times for keyframes (green dots) and for the approach by Bergmann et al. [16]. Similar results are obtained for the second camera. The data is aligned as proposed by Bergmann et al. [16]. The sample texture in Fig. 8 extracted from the wall sequence shows clear visual improvements. The seams disappear and the colors become more uniform.

V. Conclusions

We presented a fast and easy-to-use photometric calibration method that is based on a visual SLAM system running online without the need for white or evenly illuminated surfaces, calibration targets, or known scene geometry. We employ thin plate splines with a sixth-order polynomial for approximating the attenuation factors w.r.t. the image position to deal with sparsely distributed scaling estimates, and to obtain pixel-wise vignetting correction factors. The experimental results substantiate that the calibration converges quickly and effectively corrects vignetting and likewise estimates the camera response function, exposure times, and scene radiance. The fitting approach works well with different models of varying complexity and, thus, allows us to cover non-standard camera configurations as well. Due to the straightforward implementation and fast convergence, our contribution can serve as a general initialization stage for robot vision algorithms on mobile platforms that can then quickly adapt to the current camera setup.

References

[1] Y. Zheng, S. Lin, C. Kambhamettu, J. Yu, and S. Kang, "Single-image vignetting correction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2243–2256, 2009.

[2] L. Lopez-Fuentes, G. Oliver, and S. Massanet, *Revisiting Image Vignetting Correction by Constrained Minimization of Log-Intensity Entropy*. Springer International Publishing, 2015.

[3] D. Goldman and J. Chen, "Vignette and exposure calibration and compensation," in *Proc. of the IEEE Int. Conference on Computer Vision (ICCV)*, 2005.

[4] H. Lauterbach, D. Borrmann, and A. Nüchter, "Towards radiometrical alignment of 3D point clouds," *Proc. of the Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. (ISPRS)*, 2017.

[5] Z. Zhu, J. Lu, M. Wang, S. Zhang, R. Martin, H. Liu, and S. Hu, "A comparative study of algorithms for realtime panoramic video blending," *arXiv:1606.00103*, 2016.

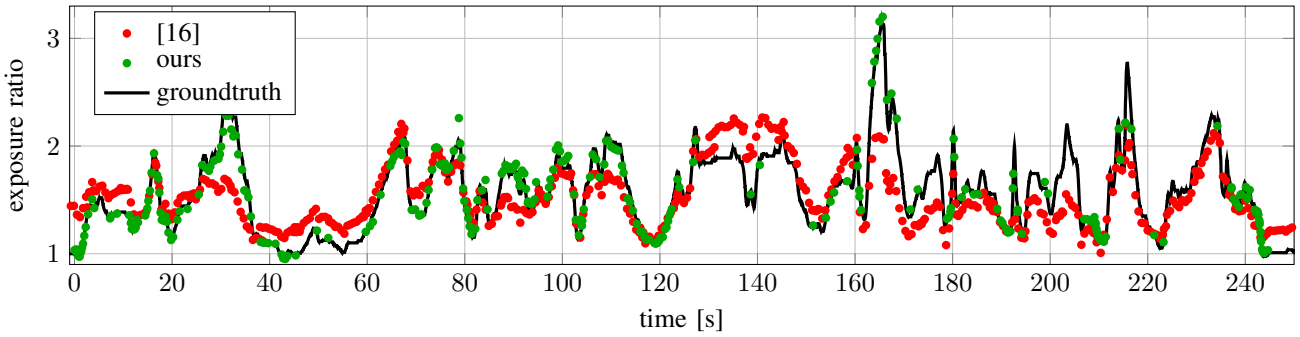


Fig. 6: Exposure ratios on the lab sequence for one camera. The estimated ratio of the keyframes (dots) follows accurately the real exposure ratio of the camera. Shown are the optimized keyframe estimates.

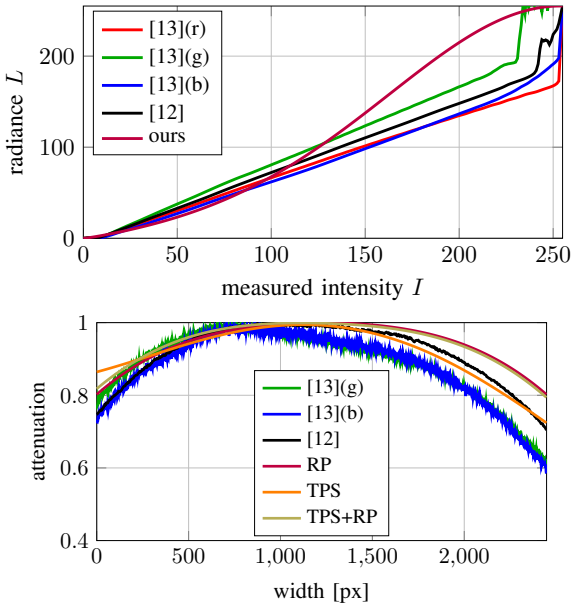
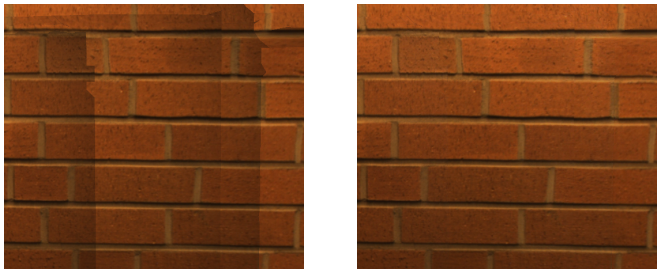


Fig. 7: Estimated inverse CRFs (top) and cross section of vignetting masks (bottom) along the central row. Results using method [13] are reported for each color channel (r,g,b). The vignetting estimates for radial polynomial (RP), TPS and the combination (TPS+RP) and our CRF were calculated on the lab sequence.



(a) Uncorrected (b) TPS + radial polynomial

Fig. 8: Reconstructed texture of a brick wall. The uncorrected approach exhibits brightness differences. Our corrected approach shows substantial improvement.

[6] M. Waechter, N. Moehrle, and M. Goesele, "Let there be color! Large-scale texturing of 3D reconstructions," in *Proc. of the European Conference on Computer Vision (ECCV)*, 2014.

[7] Q. Zhou and V. Koltun, "Color map optimization for 3D reconstruction with consumer depth cameras," *ACM Trans. Graph.*, 2014.

[8] P. Seonwook, T.Schöps, and M. Pollefeys, "Illumination change robustness in direct visual slam," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2017.

[9] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2014.

[10] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611–625, 2018.

[11] X. Zheng, Z. Moratto, M. Li, and A. I. Mourikis, "Photometric patch-based visual-inertial odometry," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2017.

[12] J. Engel, V. Usenko, and D. Cremers, "A photometrically calibrated benchmark for monocular visual odometry," *arXiv:1607.02555*, 2016.

[13] S. Alexandrov, J. Prankl, M. Zillich, and M. Vincze, "Calibration and correction of vignetting effects with an application to 3D mapping," in *Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, 2016.

[14] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. of the Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 1997.

[15] J. Quenzel, M. Nieuwenhuisen, D. Droeschel, M. Beul, S. Houben, and S. Behnke, "Autonomous MAV-based indoor chimney inspection with 3D laser localization and textured surface reconstruction," *Journal of Intelligent & Robotic Systems (JINT)*, 2018.

[16] P. Bergmann, R. Wang, and D. Cremers, "Online photometric calibration of auto exposure video for realtime visual odometry and SLAM," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2018.

[17] S. Houben, J. Quenzel, N. Krombach, and S. Behnke, "Efficient multi-camera visual-inertial SLAM for micro aerial vehicles," in *Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, 2016.

[18] T. Ng, S. Chang, and M. Tsui, "Using geometry invariants for camera response function estimation," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[19] F. Utreras and M. Varas, "Monotone interpolation of scattered data in R^s," *Constructive Approximation*, vol. 7, no. 1, pp. 49–68, 1991.

[20] A. Handa, T. Whelan, J. McDonald, and A. Davison, "A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2014.

[21] J. Schönberger, E. Zheng, M. Pollefeys, and J. Frahm, "Pixelwise view selection for unstructured multi-view stereo," in *Proc. of the European Conference on Computer Vision (ECCV)*, 2016.

[22] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS)*, 2013.

[23] M. Grossberg and S. Nayar, "What is the space of camera response functions?" in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[24] S. Agarwal, K. Mierle, and Others, "Ceres solver," [online] ceres-solver.org, 2016.