Calibration and Correction of Vignetting Effects with an Application to 3D Mapping

Sergey V. Alexandrov, Johann Prankl, Michael Zillich and Markus Vincze

***Abstract*— Cheap RGB-D sensors are ubiquitous in robotics. They typically contain a consumer-grade color camera that suffers from significant optical nonlinearities, often referred to as vignetting effects. For example, in Asus Xtion Live Pro cameras the pixels in the corners are two times darker than those in the center of the image. This deteriorates the visual appearance of 3D maps built with such cameras. We propose a simple calibration method that only requires a sheet of white paper as a calibration object and allows to reliably recover the vignetting response of a camera. We demonstrate calibration results for multiple popular RGB-D sensors and show that removal of vignetting effects using a nonparametric response model results in improved color coherence of the reconstructed maps. Furthermore, we show how to effectively compensate color variations caused by automatic white balance and exposure time control of the camera.**

I. Introduction

Since the release of the Microsoft Kinect, RGB-D cameras came to play a major role in robotic perception. They combine a conventional color camera with a depth sensor and capture both appearance and geometry of observed scenes. The availability of this cheap and rich sensory input pushed the limits of what is possible in many areas, including 3D mapping.

State-of-the-art RGB-D SLAM systems demonstrate ever improving trajectory estimation accuracy [1], surface reconstruction quality [2], and ability to handle large-scale environments [3]. However, little progress has been reported towards improving visual appearance of produced maps.

Most online reconstruction systems employ some form of color averaging. Each surfel or voxel (depending on the map representation) has an associated color value. These values are averaged with new observations as they become available over time. Averaging is usually weighted and additionally a number of ad-hoc rules are introduced to discard unreliable observations. Nevertheless, this approach often leads to visually poor results, as demonstrated in Figure 1 (top).

An implicit assumption behind averaging is that image pixel intensities directly reflect apparent color of object points. In reality, this does not hold due to the nonlinearities involved in the image formation process. Most notably, vignetting effects are responsible for intensity fall-off towards the edges of the image. For example, the corner pixels in the Asus Xtion Live Pro camera are two times darker than those in the center (see Figure 2). Therefore, the result of averaging strongly depends on the image locations at which observations were made.

Vision4Robotics Group, ACIN, TU Wien, Austria {alexandrov, prankl, zillich, vincze}@acin.tuwien.ac.at

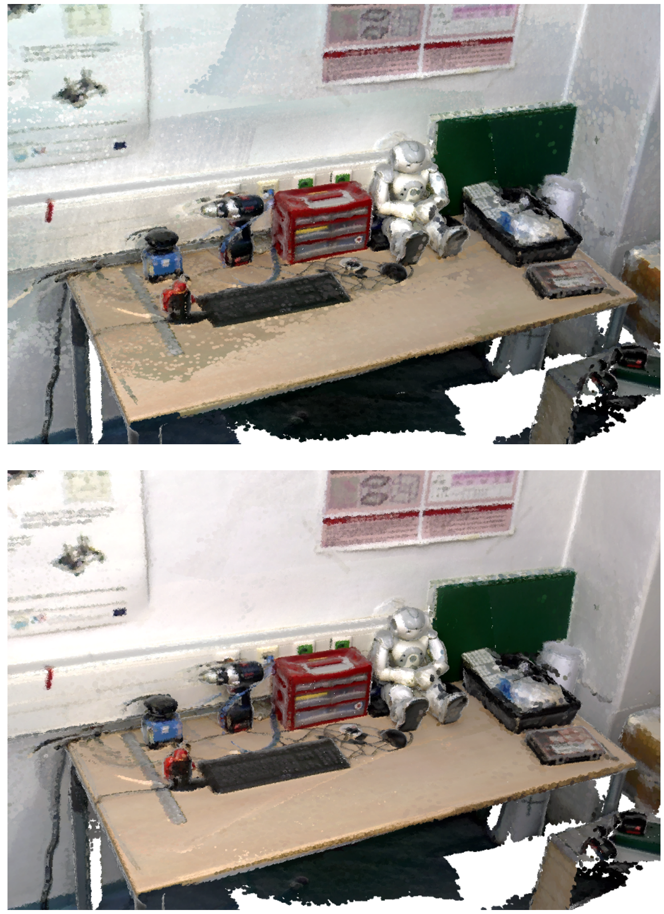


Fig. 1. Top: surfel-based 3D reconstruction obtained using ElasticFusion [2]. Automatic white balance and exposure control were disabled; color artifacts are purely due to vignetting effect. Bottom: reconstruction from the same image sequence, but applying the proposed color correction method.

Vignetting effects are a combination of natural, pixel, optical, and mechanical vignetting [4]. They are caused by the design of the lens system and imaging sensor. Therefore, vignetting effects are intrinsic for the camera and can be calibrated and compensated. Importantly, such a correction is completely orthogonal to other techniques that might be employed to improve the appearance of 3D reconstructions.

We contribute a calibration method that allows to recover the vignetting response of a color camera. It employs a simple data collection procedure that only requires a sheet of white paper as a calibration object and half a minute of operator's time. Collected data can be used to fit parametric models of vignetting response commonly adopted in color correction literature. However, we advocate usage

Fig. 2. Vignetting effects in an image from the Asus Xtion Live Pro camera. The scene consists of a wash basin and a flat wall. The upper and lower parts of the wall are uniformly painted with two slightly different shades of beige. The left side of the figure presents the original image. Intensity at the corners of the image drops below 50% of the intensity in the center. On the right side of the figure correction has been applied to remove vignetting effects. The upper and lower parts of the wall appear to have uniform colors, as in reality.

of a nonparametric model (a look-up-table with per-pixel correction factors) derived directly from calibration data. We demonstrate that it better captures optical nonuniformity, without overfitting to the calibration data. After correcting the vignetting effects, we show how the color variations caused by automatic white balance and exposure time control of the camera can be effectively compensated.

The software toolbox implementing proposed methods is hosted at `https://github.com/taketwo/radical`.

The paper is structured as follows: Section II gives a brief account of the radiometric image formation process, followed by an overview of related work on color correction and high-quality texturing in 3D reconstruction. Section III describes our proposed data collection and model fitting approach. Section IV explains how to apply the calibrated model to correct vignetting effects and to compensate for variation in exposure time. Experimental evaluation is presented in Section V and the paper is concluded in Section VI.

II. Preliminaries and Related Work

A. Radiometric Image Formation

Image formation in a camera is a complex process. Conceptually, it can be decomposed into two phases. Firstly, the energy emitted by the scene points in the direction of the camera (radiance) passes through the lens system and falls on the image plane. Secondly, the power received on the image plane (irradiance) is converted into an electrical signal and then to discrete pixel values (intensity).

The first phase can be mathematically stated as

$$E_{\mathbf{x}} = V(\mathbf{x}) L_{\mathbf{X}}, \tag{1}$$

where $L_{\mathbf{X}}$ is the radiance of a scene point $\mathbf{X}$, $E_{\mathbf{x}}$ is the irradiance at image location $\mathbf{x}$ (to which the scene point

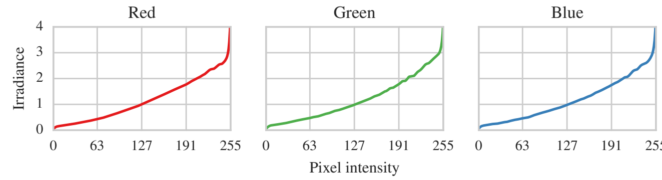


Fig. 3. Inverse radiometric response function of an Asus Xtion Live Pro camera, recovered using the method of Debevec and Malik [5]. The irradiance is recovered up to an unknown scale factor.

projects), and $V(\cdot)$ is a spatially varying optical response of the lens system. Typically, this response has a form of radial fall-off from the center of the image towards the edges. It is a combination of natural, pixel, optical, and mechanical vignetting, which are often collectively referred to as "vignetting" effect [4].

The second phase is given by

$$I_{\mathbf{x}} = f(kE_{\mathbf{x}}), \tag{2}$$

where k is the exposure time, $I_{\mathbf{x}}$ is the pixel intensity, and $f(\cdot)$ is the radiometric response function of the camera.[1] It describes how the amount of light that fell on a unit area of the image sensor maps to a pixel intensity of the output image. It is an invertible nonlinear mapping. A number of approaches to recover this mapping were proposed [5]–[7]. In this work we adopt the method of Debevec and Malik [5], where the inverse response is estimated from a set of images of a static scene taken with different exposure times. Figure 3 demonstrates the recovered response of an Asus Xtion Live Pro camera.

Combining (1) and (2) we get

$$I_{\mathbf{x}} = f(kV(\mathbf{x})L_{\mathbf{X}}) \tag{3}$$

This equation suggests that to obtain a faithful appearance models in 3D maps we need to know $f(\cdot)$ and $V(\cdot)$. This will allow to inverse the image formation process and accumulate scene point radiances instead of pixel intensities.

B. Related Work on Color Correction

The task of building 3D maps with color bears similarities with that of constructing panoramic images and mosaicking. In both cases multiple observations of the same scene point are available and have to be fused. Color correction for image stitching is a well-researched topic; the approaches can be roughly split into two groups.

The first group works directly in the image intensity domain. One of the most sophisticated algorithms is that of Oliveira et al. [8]. They segment images into regions, compute local joint color histograms of these regions, and model them using truncated Gaussians. By comparing the fitted parameters between matching regions they establish color mapping functions, which are then applied locally. Such algorithms are expensive, but can handle images made

[1] Also known as the camera response function (CRF).

by different cameras, with different settings, and under different lighting conditions.

The second group considers the radiometric image formation process. All the functions and quantities involved in (3) are explicitly recovered. The vignetting effect is removed and differences in exposure time are compensated. Goldman and Chen [4] assume that vignetting response is the same for all color channels and is symmetrical around the image center. They parameterize it as a 6^{th} order even polynomial

$$V(\mathbf{x}) = 1 + \sum_{n=1}^{3} \beta_n \left(\mathbf{x} - \mathbf{c}\right)^{2n}, \tag{4}$$

where $\mathbf{c}$ is the center of the image. They use an alternating optimization scheme to fit model coefficients and estimate exposure times and scene point radiances. The same model is used in [9] and [10]. In the latter an elegant way to decouple vignetting response estimation from the rest of the problem is presented, which leads to a more robust and efficient estimation algorithm. Conversely, in the work of Yu [11] no functional approximation of the vignetting response is used. Instead, a look-up-table (LUT) with per-pixel correction factors is obtained by observing a reference object in controlled conditions. He takes a single picture of this object, applies a wavelet denoising scheme, and stores the result in a LUT.

C. Related Work on High-Quality Texturing

Recently, several contributions were made that aim at improving coherency and fidelity of reconstructed color maps. Meilland et al. [12] explored the possibilities for high dynamic range colors in dense SLAM systems. Kerl et al. [13] made an attempt to estimate the pure albedo of the textures observed in RGB-D frames. This allowed them to remove illumination effects from color images.

Maier et al. [14] presented a keyframe-based approach, where in parallel to the volumetric reconstruction a set of RGB-D keyframes is collected as the mapping progresses. These keyframes are upscaled to super-resolution. Once the mapping is completed, a mesh representation is extracted from the TSDF volume and keyframes are mapped onto it.

Zhou and Koltun [15] considered the problem of mapping color images onto geometric reconstructions. Typically, both the images and geometry are noisy, therefore perfect mapping is impossible. They formulate a problem where the camera poses and non-rigid correction functions for images are jointly optimized to maximize photometric consistency. Their method performs global optimization and is inherently offline, but outperforms volumetric blending approaches.

In the related field of object modeling, Narayan and Abbeel [16] have formulated a similar optimization problem, although without frame correction functions. They demonstrate that their method produces coherent color models.

None of the mentioned work explicitly account for the vignetting effects, and only Meilland et al. estimate exposure time differences. Therefore, the color correction technique presented in this paper may be viewed as complementary and can be employed to further improve the appearance of 3D reconstructions.

III. Vignetting Calibration Method

Given the fact that the geometry of the lens is fixed in commodity RGB-D sensors, we make a simplifying assumption that the optical response of the camera does not change. Furthermore, we assume that the individual color channels are uncorrelated and can be treated separately.

First, we estimate the radiometric response function of the camera using the method of Debevec and Malik [5].[2] It returns the inverse of the response function in a form of a look-up table with 2^8 values per channel. This can be used to map pixel intensity to image irradiance and vice versa.

A. Data Collection

Calibration setup requires a flat white object with Lambertian reflectance and a static environment with even lighting. In practice, a sheet of printer paper attached to a desk in a typical office is suitable. Automatic exposure and white balance control in the camera have to be disabled. The exposure time is set to the maximum possible value such that none of the pixels are saturated (or close to saturation where the radiometric response is highly nonlinear). This ensures that the amount of noise in collected images is minimized.

The operator moves the camera around to make sure that the calibration object has been projected to every image pixel location multiple times, thus acquiring as much redundancy in calibration data as possible. Care should be taken to avoid specular reflections and casting shadows on the calibration object. In each incoming frame the object is detected. We use floodfill segmentation with adaptive range and seed it with the center of the object from the previous frame. This simple approach is robust enough for our calibration setup. The intensities of pixels that belong to the object are recorded into buffers according to their spatial location in the image.

By following the prescribed procedure, in under a minute hundreds of observations of the object through each image pixel are accumulated. We compute a per-pixel mean, therefore obtaining a “fused” image $\bar{I}$ where noise is averaged out (see Figure 4a).

B. Model Fitting

Equation (3) can be rewritten as

$$f^{-1}(\bar{I}_{\mathbf{x}}) = kV(\mathbf{x})L_{\mathbf{X}} \tag{5}$$

$$V(\mathbf{x}) = \frac{f^{-1}(\bar{I}_{\mathbf{x}})}{kL_{\mathbf{X}}} \tag{6}$$

The exposure time k was fixed and the radiance $L_{\mathbf{X}}$ is the same for each point of the calibration object, therefore the denominator in (6) is a constant. Taking into account that the radiometric response itself is recovered only up to a scale factor [5], we may write that $V(\mathbf{x}) = f^{-1}(\bar{I}_{\mathbf{x}})$. Therefore, after applying the inverse of previously recovered radiometric response function, we obtain a “distilled” vignetting

[2]An open-source implementation is available in the OpenCV library.

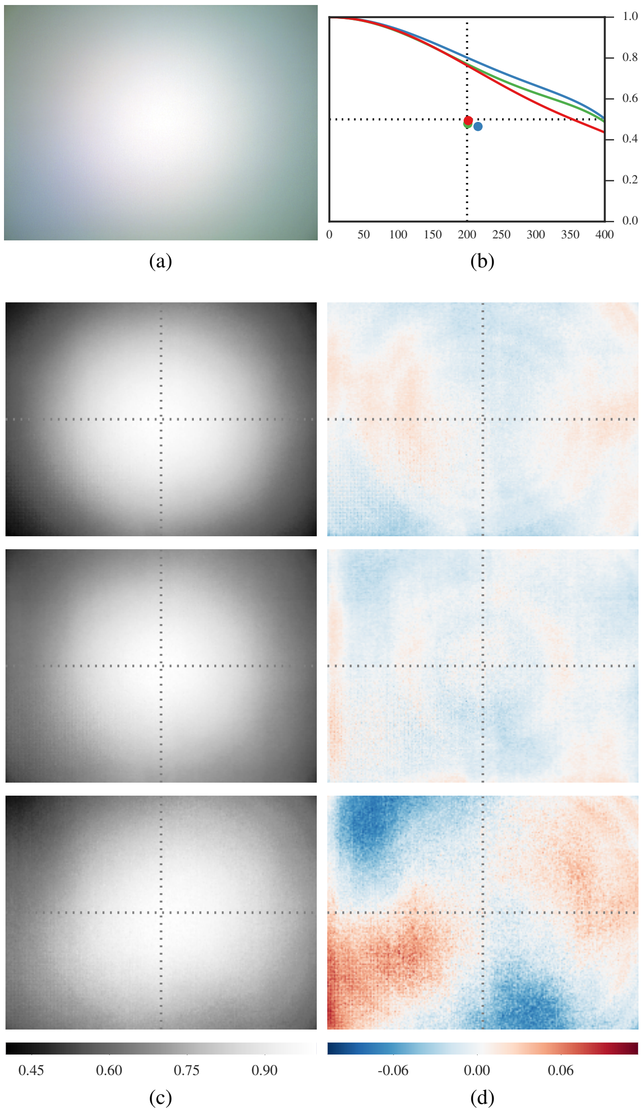


Fig. 4. Calibration example for Asus Xtion Live Pro camera. (a) A "fused" image of a flat uniformly lit white paper. (b) Schematic representation of the polynomial vignetting model fitted to the data. The dots show locations of the centers of symmetry in the image space. Curves represent the amount of radial fall-off starting from 1.0 in the center to below 0.5 in the corners of the image. Different channels are shown with their corresponding colors. (c) Individual channels (Red, Green, Blue) in the "fused" image (normalized). (d) Residuals of the fitted polynomial model.

response. If we normalize it and inverse each element, we get a matrix of correction factors, that can be multiplied with irradiance images to remove the vignetting effect.

Alternatively, a parametric model (4) can be fitted to the data. We would like the error residual $\|I_{\mathbf{x}} - V(\mathbf{x})\|$ to be small for each location $\mathbf{x}$ in the image. Therefore, we minimize the following objective:

$$(\mathbf{c}*, \beta*) = \arg\min_{\mathbf{c},\beta} \sum_i \left\| I_{\mathbf{x}_i} - 1 - \sum_{n=1}^{3} \beta_n \left(\mathbf{x}_i - \mathbf{c}\right)^{2n} \right\|^2 \tag{7}$$

Note that $\mathbf{c}$ is usually fixed to be the center of the image. In this case the model is linear in it's parameters and thus the optimization problem can be solved using ordinary least squares. In our experiments we discovered that for some cameras the center of vignetting is shifted with respect to the image center. In such cases, adding an extra degree of freedom to the model by turning $\mathbf{c}$ into a model parameter allows us to fit the empirical data better. The objective becomes nonlinear, but typically can be solved using the Levenberg-Marquardt algorithm in several iterations.

IV. Color Correction

Having obtained $V(\cdot)$ either in the form of a look-up-table of correction factors, or as a parametric model, we can compensate the vignetting effect in images delivered by the camera. From (3), the corrected intensities are computed as

$$I'_{\mathbf{x}} = f\left(\frac{f^{-1}(I_{\mathbf{x}})}{V(\mathbf{x})}\right) \tag{8}$$

For a given image resolution this is an $O(1)$ operation. If automatic white balance and exposure control in the camera are disabled, this correction is sufficient. Otherwise the differences in exposure time should be estimated.[3]

A. Compensation of Exposure Time Variation

Assume that two images of the same scene are captured with different exposure times from slightly different view-points. We can perform dense matching and find pairs of pixels that correspond to the same world point. Assuming that its radiance is constant, the ratio of exposure times is

$$\frac{k_1}{k_2} = \frac{f^{-1}(I_{\mathbf{x}_1})V(\mathbf{x}_2)}{f^{-1}(I_{\mathbf{x}_2})V(\mathbf{x}_1)}, \tag{9}$$

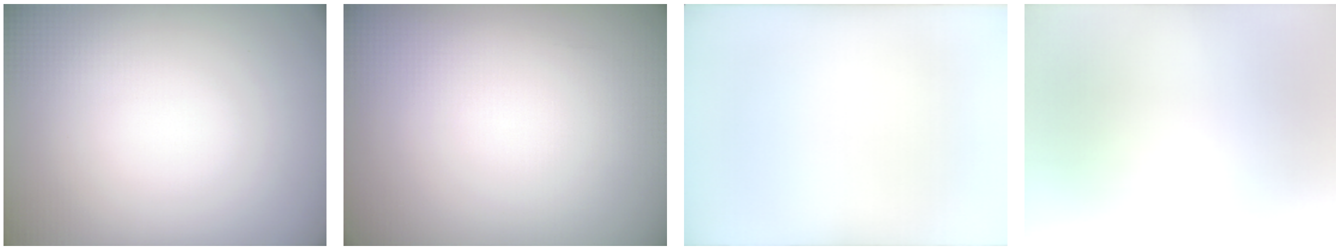
where $\mathbf{x}_1$ and $\mathbf{x}_2$ are coordinates of corresponding pix-els in first and second image respectively. Potentially, we have thousands of pairs of matched points, and in practice they will yield different ratios because of image noise and mismatches. Furthermore, the assumption that the radiance of a scene point is constant does not always hold, e.g. for non-Lambertian surfaces. Therefore, in order to robustify exposure ratio estimation, we propose to use the median of the ratios computed for all pixel correspondences.

V. Experimental Evaluation

We start by presenting an example calibration of an Asus Xtion Live Pro camera. Figures 4a and 4c show the collected dataset. Red and green channel responses are symmetrical around the image center and the polynomial model fits well. However, the response of the blue channel seems to be shifted to the right and demonstrates inhomogeneity that can not be adequately modeled with a polynomial.

Figures 5a and 5b present calibrated vignetting responses for two more Asus cameras. They demonstrate a similar amount of radial fall-off, however the difference between inhomogeneities in the blue channel is clearly visible. Figures 5c and 5d show calibration results for Intel RealSense F200 and R200 cameras. The effects are less severe than in the Asus Xtion Live Pro, however are still noticeable.

[3] We make a simplifying assumption that auto white balance can be viewed as independent per-channel adjustment of exposure time.



(a) Asus Xtion Live Pro 1 (b) Asus Xtion Live Pro 2 (c) Intel RealSense F200 (d) Intel RealSense R200

Fig. 5. Calibration examples for multiple RGB-D sensors. Each image is a "fused" view of a flat uniformly lit white paper.

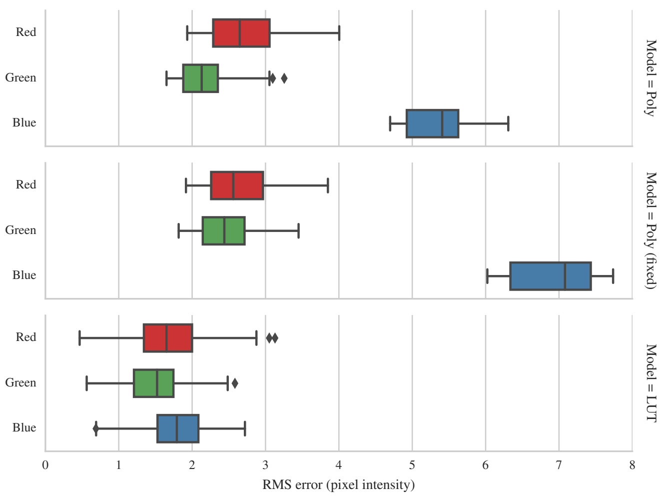


Fig. 6. Distribution of RMS errors between vignetting responses predicted by different calibrated models and real responses collected using the calibration procedure. The error unit is one pixel intensity level.

A. Vignetting Calibration Repeatability

In this subsection we quantitatively assess different parametric and nonparametric vignetting response models. Using the procedure described in Section III we collected 10 fused responses of the same Asus Xtion Live Pro camera. For each response we fitted a polynomial model without and with fixed center, and produced a LUT. Then, for each calibrated model we computed RMS error between it's predicted response and other 9 fused responses (not used for model fitting). Figure 6 presents the distribution of computed RMS errors per model type per channel. We observe that LUT captures the vignetting response well and for every channel consistently yields the smallest error on average. For the red and green channels both variants of polynomial model perform similarly, which indicates that the response is indeed symmetrical around the image center. For the blue channel the model with fixed center is worse. Notable is the large gap between nonparametric and parametric models in the blue channel.

B. Qualitative Assessment

Figures 1 and 7 present surfel-based 3D reconstructions obtained using ElasticFusion [2] without and with vignetting removal (using LUT). In both cases, automatic white balance and exposure control of the camera were disabled, so the color artifacts are purely due to vignetting effects.

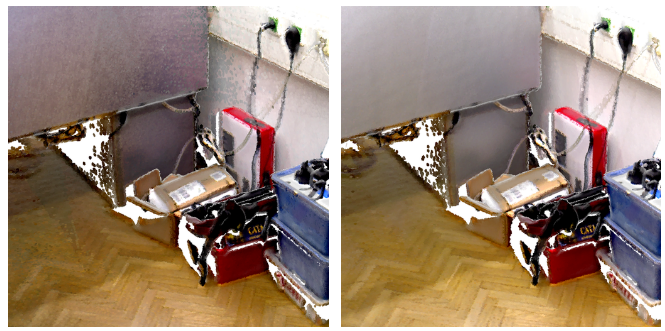


Fig. 7. Left: surfel-based 3D reconstruction obtained using ElasticFusion [2]. Automatic white balance and exposure control were disabled; color artifacts are purely due to vignetting effects. Right: reconstruction from the same image sequence, but with vignetting effect removed using LUT.

Results of applying both vignetting removal and exposure compensation are presented in Figure 8. The input datasets consist of a sequence of RGB-D frames (between 10 to 25), that are taken by a moving camera with automatic white balance and exposure control enabled. The frames are aligned using the ICP algorithm and then merged using volumetric a octree representation.

C. Indirect Quantitative Assessment

We demonstrate indirectly that removal of vignetting effect leads to more consistent and uniform color maps. In ElasticFusion a dense surfel map of the scene is maintained. Each incoming frame is aligned with the current map through minimization of geometric and photometric errors between the frame and the rendered view of the scene. Then every pixel of the incoming frame is associated with a single (or none) existing surfel through projective association. Existing surfels are then fused with the new data, i.e. their position, normal, and color are averaged with the new measurements. Unless removed, vignetting effect gets averaged in, therefore biasing estimated colors and reducing overall texture quality.

We modified the mapping software so that after the correspondences were estimated but before averaging takes place, we compute the L_1 distance between the current color of the surfel and the color of the pixel sample it will be averaged with. If vignetting effects were absent and

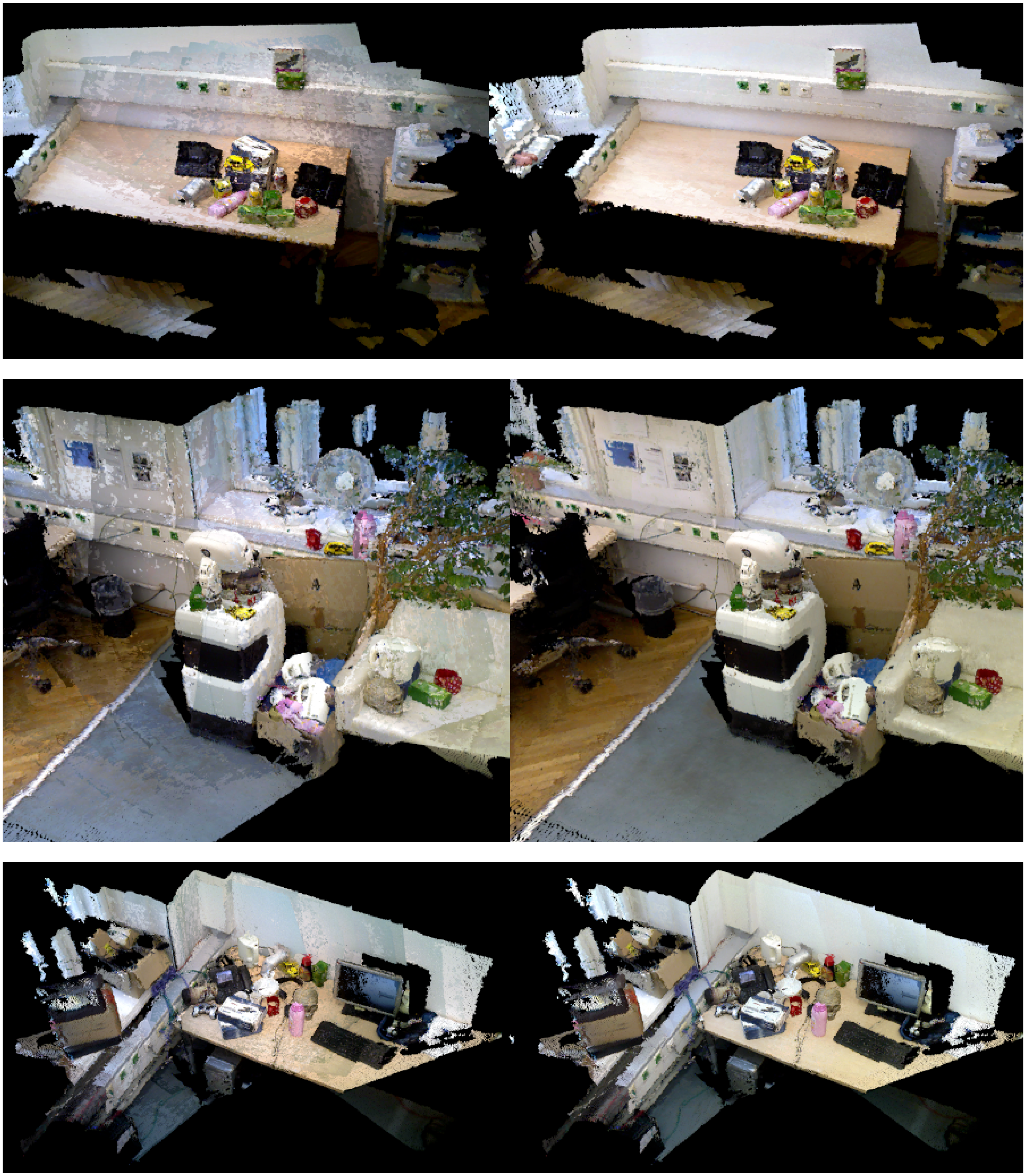


Fig. 8. Examples of 3D reconstructions using octrees from sequences of RGB-D frames corrupted with vignetting effects and with varying exposure time. On the left reconstructions without color correction are shown. On the right are reconstructions with vignetting removal and exposure time compensation.

the incoming frames were always perfectly registered to the existing model, we would expect these distances to be normally distributed around zero. In reality, due to image noise and imperfect registration, this distribution will be shifted to the right and have a tail. We processed an RGB-D sequence (the same as used to produce Figure 1) without and with vignetting removal and recorded the distribution of the L_1 distances between colors. These distributions are

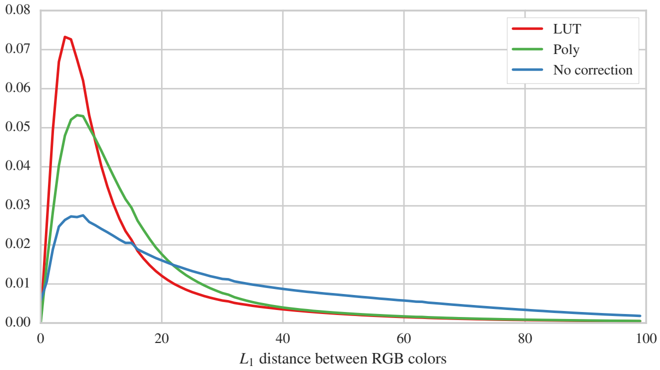


Fig. 9. Histogram of L_1 distances between colors of surfels and image pixels they are being merged with. The number of samples is close to 10^8.

presented in Figure 9. When no correction is applied, the distribution is severely skewed, which indicates that the colors being averaged are often far apart. Vignetting removal with the nonparametric model gives a large improvement.

VI. Conclusions and Future Work

In this contribution we presented a study of vignetting effects in multiple consumer RGB-D cameras. A simple and effective calibration method was proposed that yields consistent calibration results. Application of the learned calibration model results in more visually pleasing color models.

We showed quantitatively that removal of vignetting effects gives better matches between surfel colors in the process of mapping. This hints that a possible future work may include studying the impact of vignetting removal on the tracking performance in dense SLAM systems. Indeed, typically the camera pose is estimated through alignment of current frame and rendered model view by optimizing a joint geometric and photometric cost function. Improved color coherence of the 3D model might increase the basin and rate of convergence during optimization of the photometric error.

Experimental evaluation confirmed adequacy of the proposed modeling technique and the assumptions that have been made. Nevertheless, in future work we may explore correlations between color channels due to debayering, as well as influence of thermal and environmental conditions on the vignetting response of the camera.

Acknowledgments

The work presented in this paper has been funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 600623 ("STRANDS") and No. 610532 ("SQUIRREL"). We thank the anonymous reviewers for helpful comments.

References

[1] T. Whelan, M. Kaess, H. Johannsson, M. Fallon, J. J. Leonard, and J. McDonald, "Real-time Large-scale Dense RGB-D SLAM with Volumetric Fusion," *The International Journal of Robotics Research*, 2014.

[2] T. Whelan, S. Leutenegger, R. Salas-Moreno, B. Glocker, and A. Davison, "ElasticFusion: Dense SLAM Without A Pose Graph," in *Proc. of RSS*, 2015.

[3] A. Dai, M. Nießner, M. Zollhöfer, S. Izadi, and C. Theobalt, "BundleFusion: Real-time Globally Consistent 3D Reconstruction using On-the-fly Surface Re-integration," *arXiv preprint*, 2016.

[4] D. B. Goldman and J.-H. Chen, "Vignette and Exposure Calibration and Compensation," in *Proc. of ICCV*, 2005, pp. 899–906.

[5] P. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," in *Proc. of SIGGRAPH*, 1997, pp. 1–10.

[6] T. Mitsunaga and S. K. Nayar, "Radiometric Self Calibration," in *Proc. of CVPR*, vol. 1, 1999, pp. 374–380.

[7] M. D. Grossberg and S. K. Nayar, "Determining the Camera Response from Images: What is Knowable?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1455–1467, 2003.

[8] M. Oliveira, A. D. Sappa, and V. Santos, "A Probabilistic Approach for Color Correction in Image Mosaicking Applications," *IEEE Transactions on Image Processing*, vol. 24, no. 2, pp. 508–523, 2015.

[9] C. Doutre and P. Nasiopoulos, "Fast Vignetting Correction and Color Matching for Panoramic Image Stitching," in *Proc. of ICIP*, 2009.

[10] S. J. Kim and M. Pollefeys, "Robust Radiometric Calibration and Vignetting Correction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 562–576, 2008.

[11] W. Yu, "Practical Anti-vignetting Methods for Digital Cameras," *IEEE Transactions on Consumer Electronics*, vol. 50, pp. 975–983, 2004.

[12] M. Meilland, C. Barat, and A. Comport, "3D High Dynamic Range Dense Visual SLAM and its Application to Real-time Object Relighting," in *Proc. of ISMAR*, 2013, pp. 143–152.

[13] C. Kerl, M. Souiai, J. Sturm, and D. Cremers, "Towards Illumination-invariant 3D Reconstruction using ToF RGB-D Cameras," in *Proc. of 3DV*, 2014, pp. 39–46.

[14] R. Maier, J. Stuckler, and D. Cremers, "Super-resolution Keyframe Fusion for 3D Modeling with High-Quality Textures," in *Proc. of 3DV*, 2015, pp. 536–544.

[15] Q.-Y. Zhou and V. Koltun, "Color Map Optimization for 3D Reconstruction with Consumer Depth Cameras," *ACM Transactions on Graphics*, vol. 33, pp. 1–10, 2014.

[16] K. S. Narayan and P. Abbeel, "Optimized Color Models for High-Quality 3D Scanning," in *Proc. of IROS*, 2015, pp. 2503–2510.