Optical calibration and distortion correction for a volumetric augmented reality display

Kishore Rathinavel, Hanpeng Wang, and Henry Fuchs

University of North Carolina, Chapel Hill

ABSTRACT

We develop optical calibration and distortion correction for a recently developed DMD-based volumetric augmented reality display. The display is capable of displaying imagery over a large volume — composed of 280 depth planes over a large depth-range (15 cm to 400 cm) and 40 degrees field-of-view. An unintended property of this display is that the field-of-view of the depth planes changes slightly over depth. This can cause distortions, perceptual errors for the perspective depth cue, and reduce the image quality slightly. To address these issues, we develop an optical calibration method and a distortion correction as a post-processing step to our rendering pipeline.

Keywords: Augmented Reality, Calibration, Volumetric

1. INTRODUCTION

Augmented Reality (AR) systems are future digital platforms that have tremendous potential for several applications, e.g., computing, visualization, communication, etc. Developing new AR displays is an active area of research, particularly in the direction of developing AR displays that can present all depth cues, e.g., binocular (trivial to solve), monocular,[1] and occlusion.[2] This is an important research direction because AR displays that present all depth cues will enable an advanced level of integration between our real and digital worlds and will enable several important applications.

Current generation of commercially available AR displays (except for *Magic Leap 1*[*]) typically present binocular depth cues and present the image plane at a fixed depth for each eye. An image at a fixed depth means that the eyes need to be focused at that depth always even though the eyes may converge to different depths to observe digital objects at different depths. In other words, while such displays support binocular depth cues, they do not support monocular depth cues such as accommodation and retinal blur. This can cause double vision and blurry objects, which can lead to reduced task performance.[3] Additionally, this also results in a conflict between the binocular and monocular perception systems (called the vergence-accommodation conflict), which leads to headaches and discomfort.[3] *Magic Leap 1* provides some support for monocular depth cues by presenting imagery at two depth planes, which are too few. Previous studies have shown that our eye's depth-of-field is approximately 0.3 diopters, which means that to present imagery between 0 diopters (infinity) and 3.3 diopters (30 cm), we need at least 11 depth planes.[4]

Recently, several dense binary multifocal displays have been proposed[5–8] with slight differences in operation between each other. The common working principle of these displays is that they synchronize a focus-tunable lens to a high-speed projector and set the focus-tunable lens to sweep the virtual image plane in depth continuously. An interesting, but unintended, feature of such dense multifocal displays is that the field-of-view of the depth-planes changes with depth.[5] This change in field-of-view can cause the following problems:

1. Image distortions: if two digital objects at different depths are expected to line up with each other, they will not. Alternatively, a long object somewhat parallel to the display axis will appear to curve rather than look straight.

kishore@cs.unc.edu

*https://www.magicleap.com/magic-leap-1

2. Reduced spatial resolution: since some of these displays distribute the decomposition of a color voxel over multiple binary voxels at different depths, the slight curve can introduce a blurring effect and lead to a loss in spatial resolution.[5]

3. Incorrect depth cues: Some depth cues, e.g., motion parallax, perspective, relative density, and relative size, will be slightly incorrect due to this distortions.[9]

It may be possible to build a volumetric display where the field-of-view doesn't change, but such a prototype will likely use more optical components. As an alternative, in this paper, we explore a software-based calibration and distortion correction approach.

2. BACKGROUND

The current paper mainly deals with an optical distortion present in the display presented in Rathinavel et al.[5] The same distortion might be present in other similar volumetric displays.[6–8] This section summarizes some details from Rathinavel et al.[5] necessary for the work presented in this paper.

2.1 Display overview

The display is composed of a focus-tunable lens, a high-speed DMD projector that uses an HDR LED as its illumination source. The DMD-projector and the HDR illuminator work together to produce a single-color binary image. The HDR illuminator's color can be changed on a per-binary frame basis in synchronization with the DMD. The focus-tunable lens changes optical power sinusoidally and sweeps the DMD projector's virtual image plane back and forth continuously at 60 Hz. Synchronization between all the components enables a volume composed of single-color binary virtual image planes to be built incrementally over the lens cycle slice-by-slice. The resulting volume is composed of 280 slices of single-color binary images distributed between 15 cm and 400 cm over a 40 degrees field-of-view.

2.2 Previous rendering pipeline

A rendering pipeline was developed to calculate the colors and binary image patterns to be displayed at high-speed in each lens cycle. The rendering pipeline starts by rendering an RGB image and a linearized depth map of the virtual scene. This RGBD data is converted to a volume composed of color voxels, which are decomposed to a displayable volume comprising binary image patterns and illumination values for each binary image. Each color voxel is decomposed into some binary voxels distributed over a few depth-planes such that these binary voxels will lie on a single perspective projection line so that they get integrated onto the same retinal or camera pixel.

2.3 Limitations of previous rendering pipeline

The previous rendering algorithm developed for the display assumes a constant field-of-view for each of the 280 binary slices. However, a mathematical model of the ray propagation through the optics shows that the field-of-view changes gradually over the depth-range. In this paper, we address this limitation by a one-time calibration process and introducing a post-rendering distortion correction step.

3. APPROACH FOR CALIBRATION AND DISTORTION CORRECTION

Our approach for calibration and distortion correction assumes that the depth-planes are centered on and perpendicular to the display's optical axis. The reason we assume this is because it is difficult to measure the depth of one (or a few nearby) display voxels. To measure the depth, we would need a camera with a very narrow depth-of-field such that it can discern the difference in depth between any two adjacent depth-planes — this is not practical. Besides, the human vision's depth of field is 0.3 diopters, and our depth-plane spacing is dense enough to be well within the 0.3 diopters threshold.[5]

Below we discuss each stage of our one-time calibration procedure and our post-rendering distortion correction step.

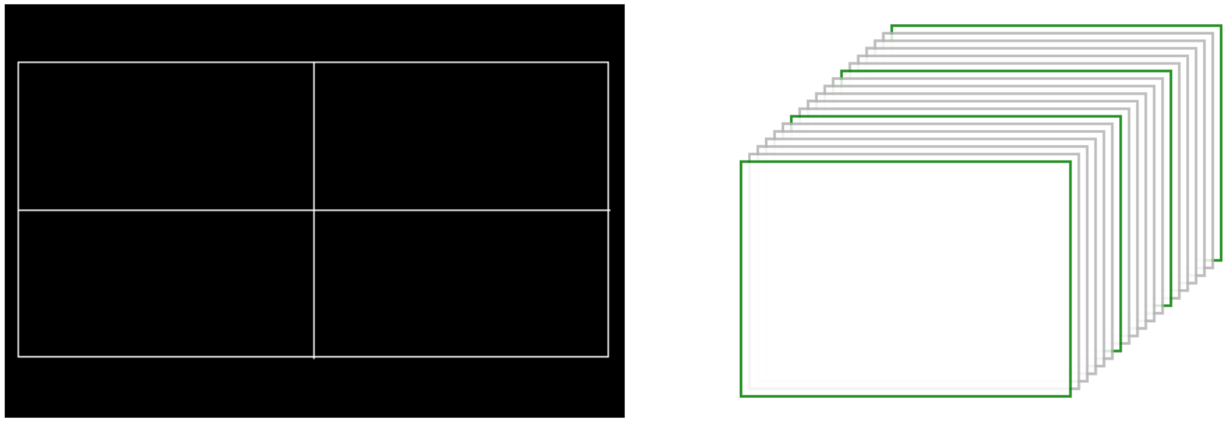

Figure 1. (Left) Calibration image that can be placed at multiple depth-planes. (Right) Calibration volume which is mostly composed of fully-black images (depicted by gray border) with a few calibration images (depicted by green border).

3.1 Synthetic volume for calibration

To use in the calibration steps, we generate a synthetic volume composed of mostly black images interspersed with images composed of a centered rectangle, centered horizontal line, and centered vertical line are bright. An example calibration image is shown in Fig. 3 (left). This image is placed at a sparse set of depth planes. Fig. 3 (right) is a concept diagram explaining the sparse locations of the calibration images — gray bordered depth-planes are fully black images whereas green bordered depth-planes contain the calibration pattern shown in Fig. 3 (left). Note that unlike the depiction in Fig. 3 (right), our display's volume is much denser (composed of 280 depth-planes).

3.2 Pre-calibration: Aligning camera's and display's optical axis

To demonstrate our calibration and distortion correction, we first need to align the recording camera's axis to coincide with the display's axis. An explanation for this follows:

Recall that in our display, each color voxel is decomposed into some binary voxels such that these binary voxels will lie on a single perspective projection line such that they get integrated onto the same retinal or camera pixel. Since multifocal plane displays are view-dependent displays, it is necessary to track the pupil position and the decomposition also needs to be view-dependent. But, in this paper, we do not use an eye-tracker, and we assume that all the depth planes are centered and perpendicular to the display's optical axis. If we assume the eye's axis to be aligned with the display's axis, the display's axis is the only line that will not need calibration or distortion correction. Hence, we need to align the camera's and display's axis.

To align the camera's axis with the display's axis, the stack of synthetic images is displayed and the camera's position was manually adjusted until the centered-horizontal and centered-vertical lines aligned. When aligned properly, the image seen is shown in Fig. 4 (left).

3.3 Calibration

Our calibration approach is to sample the field-of-view for a sparse set of depth-planes and interpolate the scaling factor that needs to be applied to each depth-plane to ensure a constant field-of-view across the entire volume. This happens in these steps:

1. To sample the field-of-view for each depth-plane, we place the image shown in Fig. 3(left) at that depth plane, capture the image seen by a camera, and measure the number of pixels between the left edge of the outer rectangle to the right edge of the outer rectangle. Say this results in a set $\{\theta_i\}$ where θ_i is the field-of-view of the i^{th} depth-plane and $i \in [1, ..., M]$, where M is the number of depth planes where we sample the field-of-view. In our experiment, $M = 10$.

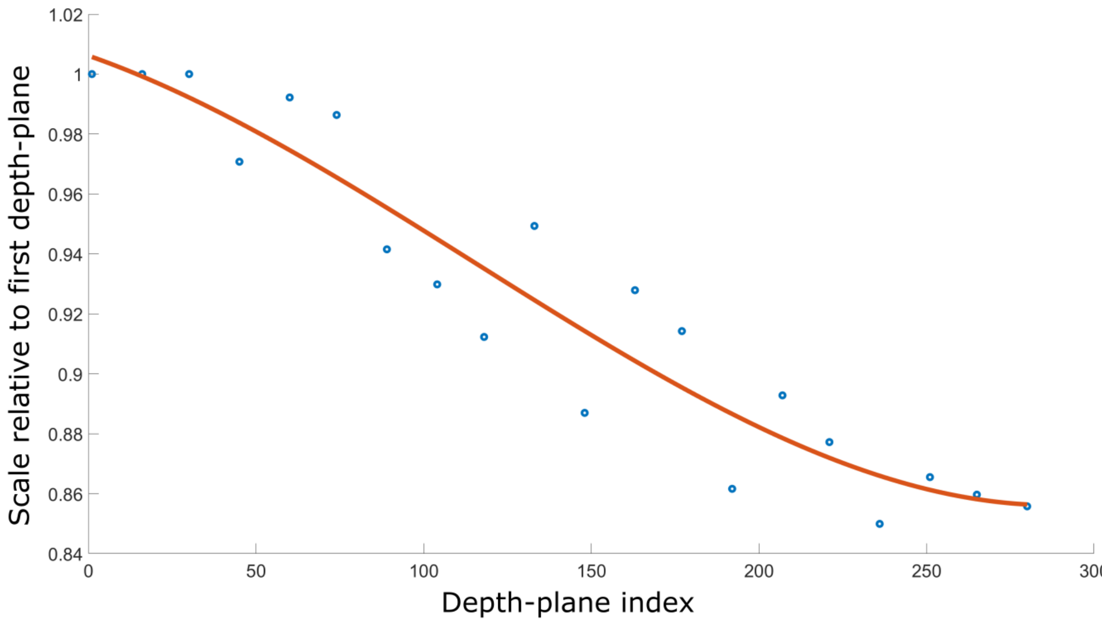

Figure 2. Plot of the relative scale as a function of the depth-plane's index, i.e., $s(j)$ mentioned in Sec. 3.3

2. We calculate the relative scaling factor for these depth planes as $s(i) = \frac{\theta_i}{\theta_1}$, i.e., the first (nearest) depth-plane is assumed to have a scaling factor of 1 and all other depth-planes are assumed to be scaled relative to this.

3. We estimate the scaling factor for all depth-planes as the function $s(j), j \in [1, 280]$ as a cubic interpolation of the data-points $\{(s(i), i)\}, i \in [1, M]$. This interpolation is shown in Fig. 3.3 where the blue circles are the sparse samples and the red curve is the estimated $s(j)$ function.

3.4 Distortion correction

To correct for the distortion introduced by the changing field-of-view, we need to scale each depth-plane by the inverse of its scaling factor, i.e., by $\frac{1}{s(j)}$.

4. RESULTS

Fig. 4 (Top row) shows the before correction and after correction images for the calibration volume. Notice how the rectangles do not line up in the before-calibration image, but they line up correctly in the after-calibration image. Another synthetic volume was generated composed of a point grid and placed at different depths. Fig. 4 (Bottom row) shows the before and after images for the point-grid volume. Notice how the points appear as lines in the before-calibration image but nearly appear as points in the after-calibration image. Both images of Fig. 4 were taken with a very narrow aperture camera so that the images at the different depths will appear clearly. If we were to take these images with a wider aperture setting, only one of the depth-planes would be in-focus, and the others would be out-of-focus and appear very blurry preventing us from verifying whether our method works. Images in Fig. 4 appear to have severe chromatic artifacts, but this happens only when the aperture of the camera is very small; for a wider aperture setting, the chromatic artifacts are not present, and all visible pixels are bright white. We do not fully understand the source of these chromatic artifacts.

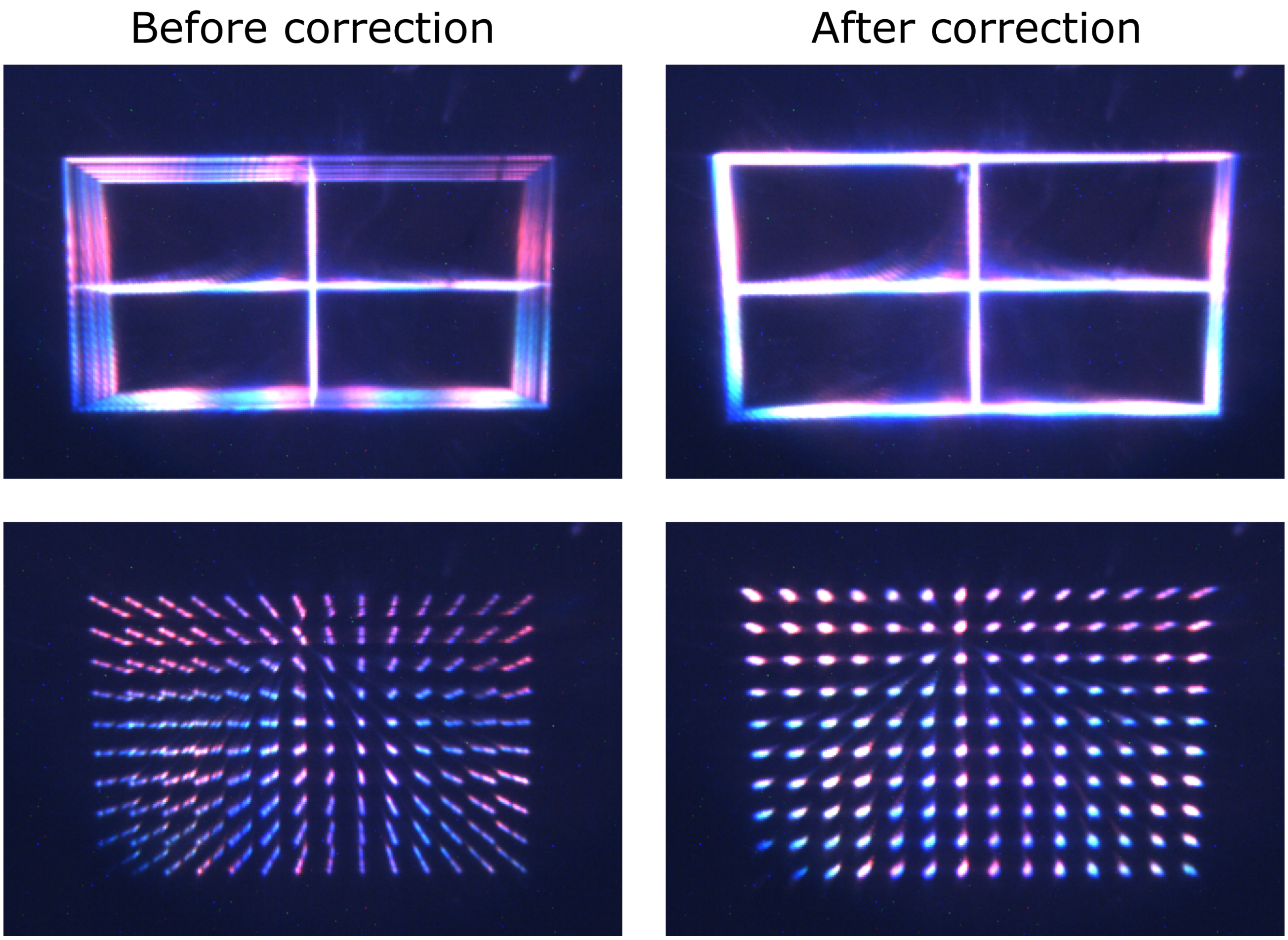

Figure 3. Demonstration of our calibration and distortion-correction approach. For these images, the display is displaying a volume across a large depth-range (15 cm to 400 cm). To ensure that all the images in this large depth-range are clearly visible, the aperture of the camera was set to the smallest setting. The chromatic artifacts seem to occur only for this narrow aperture setting.

5. DISCUSSION

Limitation of our approach Our approach does not address optical distortions that change across different lens cycles. To track optical distortions across lens cycles, we need a sophisticated lens tracking technology. Currently available lens tracking technologies are insufficient because they only track the focal length of the lens.

Future Work An interesting area of future work is to redesign the optics for a volumetric AR display such that it has a fixed field-of-view across all depths.

ACKNOWLEDGMENTS

This research was supported by NSF grant 1405847 ("II-New: Seeing the Future: Ubiquitous Computing in EyeGlasses"), and by a research gift from Intel Corporation, and by the BeingTogether Center, a collaboration between Nanyang Technological University (NTU) Singapore and University of North Carolina (UNC) at Chapel Hill. The BeingTogether Center is supported by the Singapore National Research Foundation, Prime Minister's Office, under its International Research Centers in Singapore Funding Initiative.

REFERENCES

[1] Hua, H., "Enabling Focus Cues in Head-Mounted Displays," *Proceedings of the IEEE* **105**, 805–824 (May 2017).

[2] Rathinavel, K., Wetzstein, G., and Fuchs, H., "Varifocal occlusion-capable optical see-through augmented reality display based on focus-tunable optics," *IEEE transactions on visualization and computer graphics* **25**(11), 3125–3134 (2019).

[3] Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S., "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of vision* **8**(3), 33–33 (2008).

[4] Campbell, F. W., "The depth of field of the human eye," *Optica Acta: International Journal of Optics* **4**(4), 157–164 (1957).

[5] Rathinavel, K., Wang, H., Blate, A., and Fuchs, H., "An extended depth-at-field volumetric near-eye augmented reality display," *IEEE TVCG* **24**(11), 2857–2866 (2018).

[6] Choi, S., Lee, S., Jo, Y., Yoo, D., Kim, D., and Lee, B., "Optimal binary representation via non-convex optimization on tomographic displays," *Optics express* **27**(17), 24362–24381 (2019).

[7] Lee, S., Jo, Y., Yoo, D., Cho, J., Lee, D., and Lee, B., "Tomographic near-eye displays," *Nature communications* **10**(1), 2497 (2019).

[8] Chang, J.-H. R., Kumar, B., and Sankaranarayanan, A. C., "Towards multifocal displays with dense focal stacks," in [*SIGGRAPH Asia 2018 Technical Papers*], 198, ACM (2018).

[9] Cutting, J. and Vishton, P., "Perceiving layout and knowing distances: The interaction, relative potency, and contextual use of different information about depth," in [*Perception of Space and Motion*], Epstein, W. and Rogers, S., eds., ch. 3, 69–117, Academic Press (1995).