A Turnkey Solution to Automatic Calibration and Crosstalk Reduction for Mobile Three-dimensional Display

Kui Ma, Renjing Pei*, Zheng Geng*

Institute of Automation, Chinese Academy of Sciences, China

***These two authors contributed equally to this work, and listed in alphabetical order by their last names**

Abstract

Existing mapping methods for microlens arrays 3D display, especially for mobile devices, require tedious and device-dependent measurement. No turnkey solution exists. We propose a turnkey solution to automatic calibration and crosstalk reduction for mobile 3D display, based on camera capture and optimization that requires no prior knowledge of microlens parameters.

Author Keywords

mobile 3D display, microlens arrays, rendering, crosstalk reduction.

1. Introduction

Three-dimensional (3D) display technologies for producing high fidelity 3D scenes have attracted considerable researches in recent years. The microlens arrays naked eyes 3D display technology, based on microlens arrays overlaid onto 2D screen, is relatively simple to implement and its quality of 3D image reconstruction is often adequate for many consummer electronics devices [1]. Existing rendering methods for the display technology, such as Philips' modular arithmetic [2] or rays back-projection method [3], need tedious and device-dependent measurement for accurate calibrations to produce high quality 3D display.

Moreever, applying the microlens array techniques to mobile phone 3D display presents a set of new challenges. Sizes of both pixels and microlens reach to micron level (e.g. the sizes of a pixel for iPhone 4s is 0.077mm×0.077mm), making it extremely difficult and labor intensive to perform accurate 3D calibration. Furthermore, variations on subpixels' arrangements in different screen types exacerbates the difficulty in streamlining accurate calibration procedures. As a result, no turnkey solution exists for mass production of high quality mobile 3D displays. To solve this problem, we propose a turnkey solution to automatic calibration and crosstalk reduction for mobile 3D display, based on camera capture and optimization. The ultimate goal is to have a turnkey system that automatically produces the optimized 3D rendering for any microlens-based mobile 3D display without knowing parameters of microlens arrays and LCD screen. Our experimental results show that this automatic calibration method is device-independent, can eliminate tedious manual measurement tasks, and lead to effective crosstalk reduction and high quality 3D display.

2. Technical description

Overview: We have conducted the experiment to calibrate a 3D mobile display and produce left (view-0) and right (view-1) views by simulating a human vision system for iPhone 4s, and considering the neighboring subpixels' intensity influence for optimization. Algorithms were developed and the experimental results demonstrated that our method is capable of rendering for arbitrary parameter-unknown mobile display hardware system. A schematic of the proposed technique is shown in Fig 1.

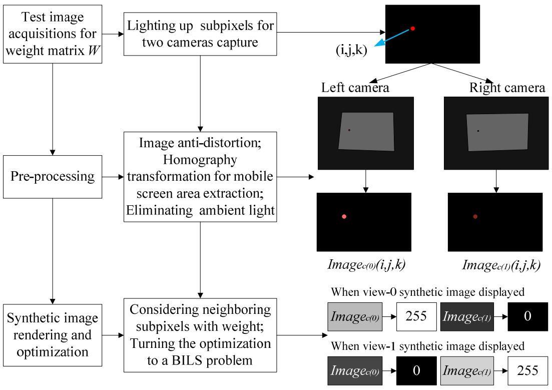


Figure 1. Flow chart of automatic turnkey solution for rendering mobile 3D display.

Step 1 - Test image acquisitions: test image $Image_m$, with resolution of $H \times W$, are shown on mobile display in turns for stereo cameras to capture (Details in section **C**2). The weight matrix W can be indexed from acquired images after pre-processing.

Step 2 - Pre-processing: (1) the original acquisition images are anti-distorted; (2) a homograph transformation is made on the interest region (mobile's screen area) in the anti-distorted image to extract a $H \times W$ image $Image_{c(t)}(t = 0 \sim 1)$; (3) ambient light is eliminated for rudimental light, though a dark enclosure experimental apparatus was built.

Step 3 - Synthetic image rendering and optimization: we formulated the entire problem into a box-constrained integer least squares (BILS [4]) optimization problem. A weighted value considering neighbor subpixel with window radius r is assigned for the subpixel (See details in sections **C***2* and **C***3*). Initial synthetic image view-0 $(\hat{I}_0)$ and view-1 $(\hat{I}_1)$ for the BILS optimization problem as:

$$\hat{I}_{0(i,j,k)} = \frac{Image_{c(0)}(i,j,k)}{Image_{c(0)}(i,j,k) + Image_{c(1)}(i,j,k)} \tag{1}$$

$$\hat{I}_{1(i,j,k)} = \frac{Image_{c(1)}(i,j,k)}{Image_{c(0)}(i,j,k) + Image_{c(1)}(i,j,k)} \tag{2}$$

where the (i, j, k) denotes the subpixel coordinate on image. The goal of our optimization is to generate a rendering such that when view-0 synthetic image I_0 displayed, the capture image $Image_{c(0)}$ closer to white and $Image_{c(1)}$ to black:

$$\min(|| Image_{c(0)} - 255_{H\times W\times 3}||_1 + || Image_{c(1)} - 0_{H\times W\times 3}||_1) \tag{3}$$

and for view-1 synthetic image I_1 to have a similar display effect:

$$\min(|| Image_{c(1)} - 255_{H\times W\times 3}||_1 + || Image_{c(0)} - 0_{H\times W\times 3}||_1) \tag{4}$$

Once the I_0 and I_1 are obtained, the calibration procedure is completed. For any scene display, it is simple to produce the synthetic image as follow:

$$\text{Syn} = I_0.\times view_0 + I_1.\times view_1 \tag{5}$$

, where $view_t$ (t = 0~1) represents a pair 3D image of the scene and ".×" denotes element-by-element multiplication of two matrices.

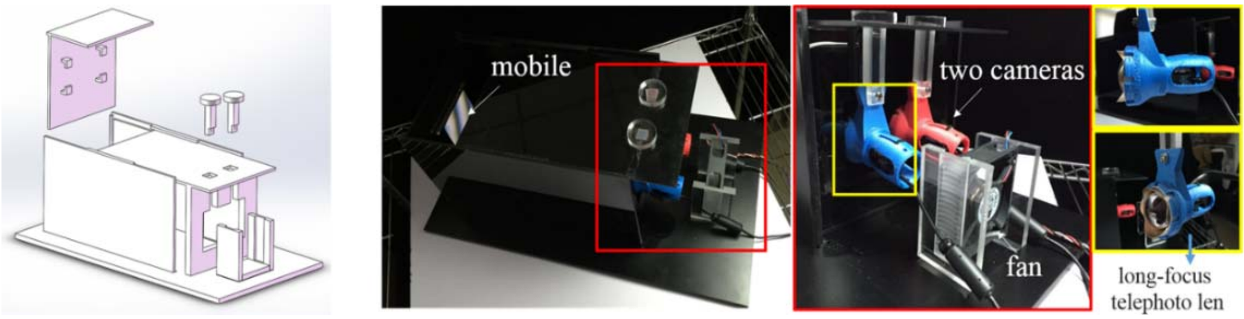


Figure 2. Hermetic experiment box: (a) design drawings; (b) the experimental setups.

(a) (b) (c) (d)

Figure 3. (a) Capture for weight: when the subpixel $(\boldsymbol{i+m, j+n, q})$ is lighted on $\boldsymbol{Image_m}$, it has an intensity influence value $\boldsymbol{w^{(t)}_{i+m,j+n,q}(-m,-n,k)}$ (after normalized) on the neighboring subpixel $(\boldsymbol{i,j,k})$ in $\boldsymbol{Image_{c(t)}}$. As a result, the weight matrix $\boldsymbol{W}$ is acquired; (b) Test images generation; (c) The pattern of picture captured by view-0 camera; (d) view-1 camera

Build experiment platform: We used a pair of stereo cameras with 60 mm baseline separation to simulate human eyes. The standoff distance between cameras and mobile display is approximately 300 mm. We designed a hermetic experiment box (Fig 2) as a dark enclosure for image acquisition. A cooling fan is fixed behind the cameras to solve the overheating problem of the cameras that may cause defocus. To maximize the pixel utilization in the acquired images, we installed a telephoto len in front of each camera.

Different test images are generated on PC and transmitted to mobile for display. The connection between the mobile display and PC is established by custom-designed socket programming in a Local Area Network (LAN).

Weight matrix acquisition: In the optimization process, weighted value considering neighbor subpixel are assigned for subpixels on I_0 and I_1. We discuss herein the acquisition of weight matrix W.

$Image_{c(t)}(i,j,k)$, $Image_{c(t)}(i,j)$, $Image_m(i,j,q)$, $Image_m(i,j)$ represent the subpixel of $Image_{c(t)}$, pixel of $Image_{c(t)}$, subpixel of $Image_m$, pixel of $Image_m$ respectively, where $i = 1 \sim H$, $j = 1 \sim W$, $k = 1 \sim 3$, $q = 1 \sim 3$. A square neighborhood window $\omega_{Image_m(i,j)}$ centered around every pixel $Image_m(i,j)$ with window radius r is defined. $Image_{c(t)}(i,j)$ will be affected, when any subpixel $Image_m(i,j,q)$ in $\omega_{Image_m(i,j)}$ lighted (Fig 3(a)). At the subpixel level, it can be computed using the relation:

$$Image_{c(t)}(i,j,k) = \sum_{m=-r}^{r} \sum_{n=-r}^{r} \sum_{q=1}^{3} w^{(t)}_{i+m,j+n,q}(-m,-n,k) Image_{m_{(i+m,j+n,q)}} \tag{6}$$

, where $w^{(t)}_{i+m,j+n,q}(-m,-n,k)(t = 0 \sim 1)$ (after normalized) is the intensity influence value at the neighboring subpixel (i,j,k) on $Image_{c(t)}$, when the subpixel $(i+m, j+n, q)$ lighted at maximum value (255) on $Image_m$. (m,n) is the coordinate of the window $\omega_{Image_m(i,j)}$ with the origin at the center of window. The W is indexed from the $w^{(t)}_{i+m,j+n,q}(-m,-n,k)$.

In our test, the mobile's screen is divided into several windows π (safety window) with size of N pixels × N pixels. Two subpixels with pitch of N pixels won't influence each other. Instead of $H \times W \times 3$ test images, only (Fig 3(b)) $N \times N \times 3$ test images are transformed to the mobile. For each $Image_m$, only one subpixel in each window π is lighted in order. For W obtain, $N \times N \times 3$ test images are displayed on the mobile and captured successively. The acquisition images are then processed, and the weight matrix W is obtained from all the normalized weight values for each subpixel.

Synthetic image rendering and optimization: For a target mobile display, when the view-0 synthetic image I_0 is displayed on the screen, view-0 camera will capture a white image and the right catch a black one. The optimization formula can be expressed as:

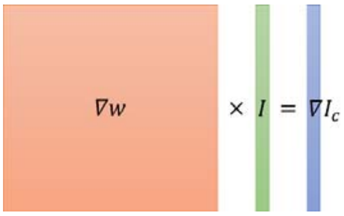
$$min \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{3} \left(255 - Image_{c(0)}(i,j,k) + Image_{c(1)}(i,j,k)\right) \tag{7}$$

We use $\nabla I_c(i,j,k)$ to record the difference of the two views for each subpixel and $\nabla w_{i+m,j+n,q}(-m,-n,k)$ denotes the weight value difference, that is,

$$\nabla I_c(i,j,k) = Image_{c(1)}(i,j,k) - Image_{c(0)}(i,j,k)$$

$$= \sum_{m=-3}^{3} \sum_{n=-3}^{3} \sum_{q=1}^{3} (w^{(1)}_{i+m,j+n,q}(-m,-n,k) - w^{(0)}_{i+m,j+n,q}(-m,-n,k)) I_{0(i+m,j+n,q)} = \sum_{m=-3}^{3} \sum_{n=-3}^{3} \sum_{q=1}^{3} \nabla w_{i+m,j+n,q}(-m,-n,k) I_{0(i+m,j+n,q)} \tag{8}$$

Our goal is to find a $I_{0(i+m,j+n,q)} \in [0,255]$ which makes $\nabla I_c(i,j,k)$ most close to 255. For each (i,j,k), we obtain a constraint as Eq.(7), and the overall problem turns into a BILS problem:



,

, where ∇w is a sparse matrix and has a full rank, with a dimension of $(H \times W \times 3, H \times W \times 3)$. I is constrained to the box $B_I = \{I \in Z^{1\times H} : \mathbf{L} \le I \le \mathbf{U}\}$, where $\mathbf{L} = 0 \times \mathbf{I_{1\times H}}$ and $\mathbf{U} = 255 \times \mathbf{I_{1\times H}}$, and $\nabla I_c = \mathbf{U}$. $I_{1(i+m,j+n,q)}$ is obtained in a similar fashion.

3. Results

As an example, we display a number pattern "0" on view-0 and "1" on view-1. The results are shown in Fig 4(a). Fig 4 (b) is the photo of a synthetic image observed through the microlens arrays at the view-0, which is almost "0" and the "1" is hardly seen.

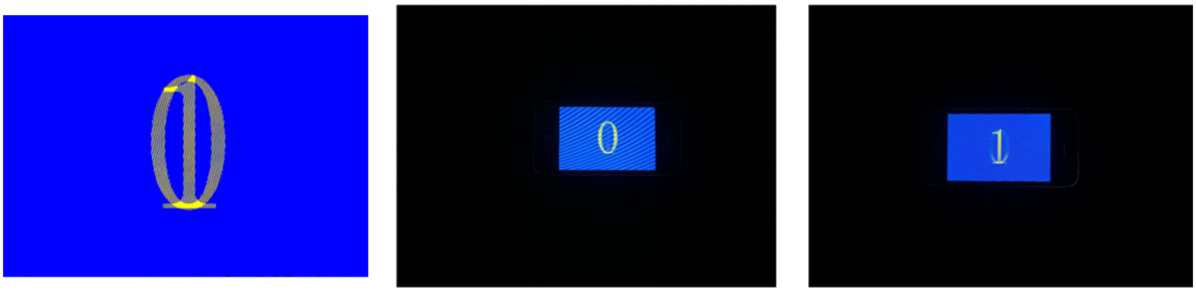
Crosstalk measurement: A camera was moved horizontally to measure the intensity in a range of viewing position (in Fig 5(a)). We lightened each view image to its maximum intensty in turn on the display screen. In the tests, only subpixels corresponding to the certain view image were turned on at maximal intensity and all other subpixels were turned off. Then, we measured each captured image's intensity respectively. The results are shown as in Fig 5(b). We calculate the crosstalk at two viewpoint as [5]:

$$\text{Crosstalk}(\%) = \text{leakage}/\text{signal} \times 100 \qquad (9)$$

In our test, the crosstalk at view-0 is 7.27% and the crosstalk at view-1 is 12.40%.

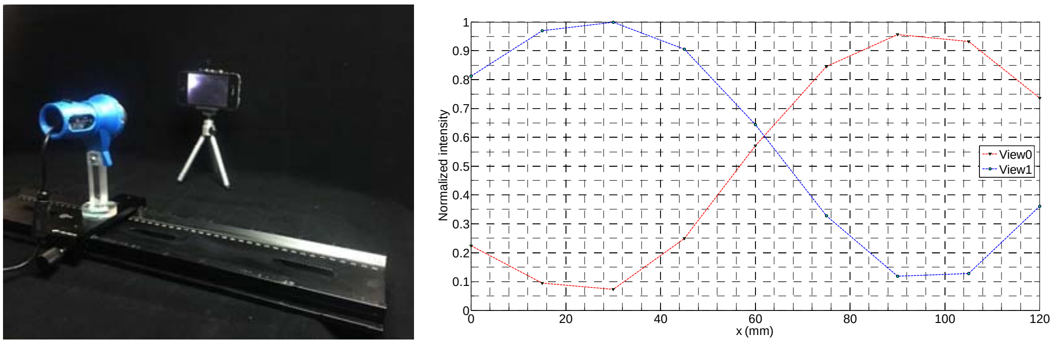
FOV Evaluation: We evaluation the actual Field of View (FOV) α' by a digital SLR camera, a tripod and a tilt board (Fig 6(a)). The target angle α between two views with the center of screen is about $11.42^o\ (= 2 \times (\arctan((60mm/2)/300mm)) \times 180^o/\pi)$. We set a subpixel on the synthetic image in view-0 write zone. When the slope angle of tilt board is adjusted from 0mm to 7.88mm, shown as Fig 6(d) and (e), we see the color of subpixel change from write to black in Fig 6(b) and (c). The distance between the differential head and the center of ball is 80 mm. The α' can be calculated as:

$\alpha' = 2 \times (\arctan(7.88\ mm/80\ mm)) \times 180^o/\text{PI} = 11.25^o$, with only 0.17^o out compare for α.



(a) (b) (c)

Figure 4. (a) The synthetic image, and the result of a synthetic image observed through the microlens arrays (b) at the view-0; (c) at the view-1.



(a) (b)

Figure 5. (a) Experimental platform for crosstalk measurement; (b) The result of intensity measurement.

Figure 6. (a) The Experimental platform for FOV evaluation: the digital SLR camera is 300mm distant from the mobile. We first set the differential head as 0 mm ((d)), where a "brightest" picture can be seen. Then we adjust the differential head until the "darkest" picture appears ((c)) and the degree scale of the differential head is 7.88mm ((e)).

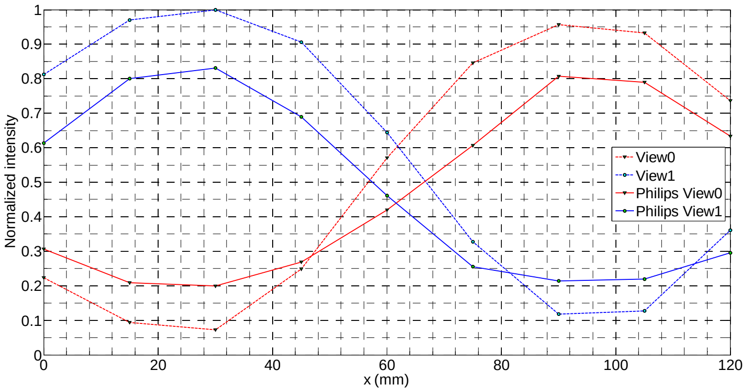


Figure 7. The result of intensity measurement by our method and Philips's method.

4. Impact

Superior performance in cross talk reduction: We compared the crosstalk measurements for our method's intensity and Philips' respectively. Results are show in Fig 7. According to the Eq. (7), the crosstalk of Philips's method is 24.06% at view-0 and 26.53% at view-1, which are larger than our crosstalk- 7.27% at view-0 and 12.40% at view-1.

Easy operation for mass production of mobile 3D displays: The existing rendering methods for the display technology, such as Philips' modular arithmetic or rays back-projection method, need accurate calibrations for setting up each display device. These methods are labor intensive, error prone. A minor error may lead to unsatisfied 3D display results, left view and right view are not completely separated easily as a result. Our proposed method eliminates the manual calibration steps and optical parameters of individual 3D display device are not known a prior. This is desiable for mass production - just put the mobile in the box, start computer program for camera capture, then a mobile screen is set up for displaying high quality 3D scene. This method could have significant impact on the advances in mobile 3D displays.

5. Acknowledgements

This work has been supported by project of National Natural Science Foundation of China (Grant No. 61605240) and another two projects of National High-tech R&D Program (863 Program) of China (Grant No. 2012AA011903 and Grant No. 2015AA015905) of Institute of Automation, Chinese Academy of Sciences (CASIA).

6. References

[1] Geng, Jason. "Three-dimensional display technologies." Advances in optics and photonics 5.4 (2013): 456-535.

[2] Van Berkel, Cees. "Image preparation for 3D LCD." Electronic Imaging'99. International Society for Optics and Photonics, 1999.

[3] R. Pei, Z. Geng and Z. Zhang. "A Novel Optimization Method for Lenticular 3D Display Based on Light Field Decomposion," in Journal of Display Technology, vol. 12, no. 7, pp. 727-735, Jul. 2016. Doi: 10.1109/JDT.2016.2517243.

[4] X.-W. Chang and Q. Han, "Solving box-constrained integer least-squares problems," IEEE Transactions on Wireless Communications, vol. 7,no. 1, pp. 277–287, 2008.

[5] Woods, Andrew J. "Crosstalk in stereoscopic displays." Journal of Electronic Imaging 91 (2012): 3286