

3D Display Calibration by Visual Pattern Analysis

Hyoseok Hwang, Hyun Sung Chang, Dongkyung Nam, In So Kweon

Abstract—Nearly all 3D displays need calibration for correct rendering. More often than not, the optical elements in a 3D display are misaligned from the designed parameter setting. As a result, 3D magic does not perform well as intended. The observed images tend to get distorted.

In this paper, we propose a novel display calibration method to fix the situation. In our method, a pattern image is displayed on the panel and a camera takes its pictures twice at different positions. Then, based on a quantitative model, we extract all display parameters (i.e., pitch, slanted angle, gap or thickness, offset) from the observed patterns in the captured images. For high accuracy and robustness, our method analyzes the patterns mostly in frequency domain. We conduct two types of experiments for validation; one with optical simulation for quantitative results and the other with real-life displays for qualitative assessment. Experimental results demonstrate that our method is quite accurate, about a half order of magnitude higher than prior work; is efficient, spending less than 2 s for computation; and is robust to noise, working well in the SNR regime as low as 6 dB.

Index Terms—Display calibration, 3D observation model, autostereoscopic display, parameter estimation, rendering correction.

I. INTRODUCTION

DURING the past decade, we have witnessed rapid advancement in 3D display technology. Diverse approaches have attempted to realize 3D scenes on a flat panel display. Among them, a multiview autostereoscopic display provides, at a low cost, immersive 3D environments to multiple users without requirement of wearing special glasses. An autostereoscopic display uses optical elements such as lenticular lenses or parallax barriers to direct each ray from the pixels toward the intended view point [2], [3]. Although the lenses and barriers work differently, i.e., lenses by refracting the rays and barriers by blocking irrelevant rays, they commonly fulfill the same principle that different pixels must be visible in different eye positions. If the left eye and the right eye are located in different viewing zones, they will observe different images and consequently recognize the binocular disparity. Moreover, as the eyes move horizontally, the observed images change slightly, realizing so-called motion parallax. The binocular disparity and the motion parallax are two most important visual cues for human 3D perception [4].

Manuscript submitted January 1, 2016. This work extends the previous work presented at a conference [1].

Hyoseok Hwang and In So Kweon are with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, South Korea (e-mail: {hyoseok.hwang, iskweon}@kaist.ac.kr). H. Hwang is also affiliated with Samsung Advanced Institute of Technology (SAIT), 130 Samsung-ro, Yeongtong-gu, Suwon 16678, South Korea.

Hyun Sung Chang and Dongkyung Nam are with SAIT (e-mail: {hyun.s.chang, max.nam}@samsung.com).

In terms of hardware, autostereoscopic displays are easy to implement. We only need to attach optical elements in front or at the back of a usual 2D display panel. Due to the simplicity as well as the cost efficiency, the autostereoscopic displays have been adopted in a variety of commercial products from 3D display walls to mobile devices [5].

Despite all the advantages, several issues still remain [6]. Above all, visual quality degradation may happen by optical misalignment and/or by crosstalk. If the optical elements are misaligned from the desired parameter values (e.g., slanted angle, pitch, etc.), the observed images get distorted. For the autostereoscopic display, multiple viewpoint images must be prepared, by various means [7] (e.g. multiple camera array, depth image based rendering, image based rendering), and be multiplexed together in a panel image, as shown in Fig. 1. Here, the multiplexing must follow correct view assignment which depends on the display parameters [8]. In practice, however, the parameters are likely to differ from the designed values for many reasons: fabrication/assembly inaccuracy, thermal effect, chronological change, to name a few. Whatever the reason is, even a small error may lead to a significant amount of image distortion.

On the other hand, crosstalk generally refers to a phenomenon whereby light from a pixel, designated for a specific view point, is smeared into adjacent ones. It occurs mainly due to the shape mismatch between the pixels and the optical elements. An autostereoscopic display typically slants the lens or barrier to mitigate moiré artifacts and to divide the resolution reduction by both dimensions. As a result, rectangular pixels can hardly pass through the slanted optical element as a whole; only a fraction can make it, forming a parallelogram jointly with multiple pieces of other pixels. This makes the viewer see not only the intended pixels but also unintended ones. Consequently, images may look blurry or look as if they were multiply-exposed [9]. Strictly speaking, given the slanted structure, the crosstalk is inevitable. But recent research results [10]–[13] show that one may reduce the crosstalk “effects” (i.e., blur, multiple-exposure, etc.) by appropriately adjusting pixel values so that they compensate for the crosstalk. To do this, one needs a precise model of the crosstalk which again requires the knowledge of the actual display parameters.

To summarize, knowing the display parameters to a sufficient level of accuracy is prerequisite to solving the image distortion and crosstalk problems. Therefore, we need to estimate the display parameters at least once, after installing the optical elements to the display, or more frequently (e.g., on-the-fly) whenever it is necessary.

In this paper, we propose a novel method for the display parameter estimation. The proposed method is based on the analysis of pattern images observed from a couple of view points. We investigate what the observed patterns tell us

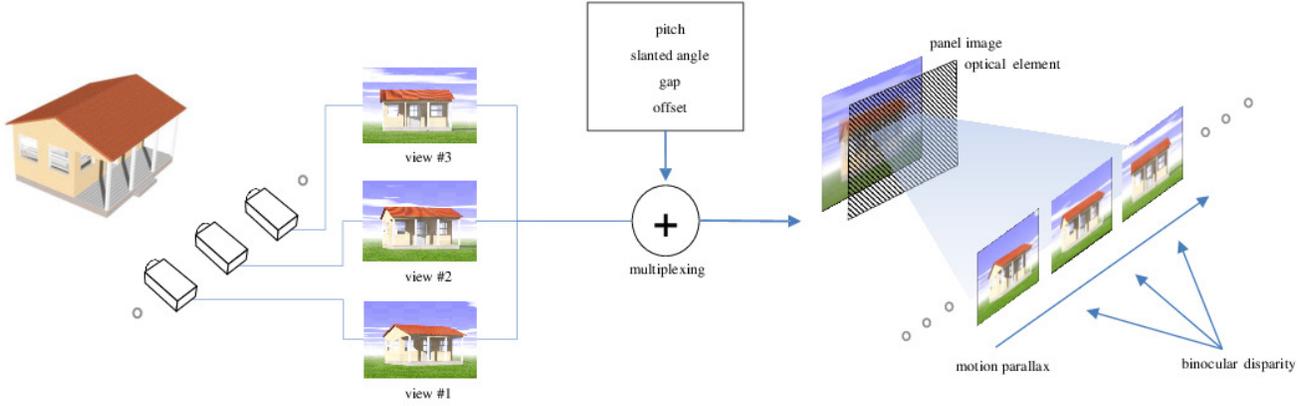


Fig. 1. Overall procedure of realizing 3D scenes in a multiview autostereoscopic display. In this example, we assume that the input images are from a multiple camera array.

about the display parameters and develop a frequency-domain algorithm which robustly works in everyday environment. Unlike prior work, which usually requires many images to be photographed from all viewing zones at the optimal viewing distance, our method only needs two observations and is largely insensitive to the observation positions.

The remainder of this paper is organized as follows. We review the related work in the next section. In Section III, we establish an observation model for images which are seen through the optical elements of the autostereoscopic display. Section IV presents a novel display calibration method, based on the observation model, which robustly estimates the display parameters in frequency domain. Then, we experimentally validate the proposed method in Section V. Finally, Section VI concludes the paper.

II. RELATED WORK

In recent years, plenty of 3D research has focused on the visual quality improvement. We first review several work seeking to minimize the 3D image distortion that occurs due to optical misalignment.

Wang *et al.* [14] propose a method to correct the subpixel positions of the panel image when the amount of optical misalignment is known. The correction is basically equivalent to newly assigning views to all subpixels according to the actual parameters. The results demonstrate the quality improvement of the 3D images; however, the misalignment needs to be somehow measured in advance. Lee and Ra [15] measure the misalignment. They display a periodic color pattern such that the subpixels of the same view have the same color, initially using the designed parameters for the view assignment. Then, they observe the pattern at the optimal viewing distance. If the actual parameters differ from the designed ones, color variations happen. The authors argue that the misalignment errors are specifically related to the number of color variations. They estimate the errors based on the counted number of color variations and re-synthesize the panel image following the estimated parameters. In another work, Lee and Ra [16] attempt to find where, among all viewing

zones, each subpixel looks brightest and then to conduct view assignment accordingly. This approach is supplemented by Zhou *et al.* [11], who concretely show how to find the pixel correspondences between the panel and captured images using structured lights. The scheme can deal with complicated types of misalignments, including inhomogeneous ones, but is not convenient to apply in practice because it requires too many images to be photographed from all viewing directions. Ge *et al.* [17] propose a computer-vision based method to estimate the parameters of VarrierTM, a huge autostereoscopic display tiled with thirty-five panels and as many parallax barriers. Stereo cameras examine left and right images, mimicking the human eyes. Then, the parameters are sequentially adjusted until the cameras find little artifact in both images. Hirsch *et al.* [18] show that the moiré-magnifier effect may be useful for calibrating the angle between the lens array and the panel. They establish a mathematical model of the observation when a specific pattern is observed through lenses. We pursue a similar course in this paper. But Hirsch *et al.* do not show how to automate the calibration and how to deal with all types of display parameters. Here our approach departs from theirs.

Another line of research improves the 3D visual quality by reducing the crosstalk effect. Li *et al.* [10] set up a set of linear equations, each of which represents the intensity mixture of light for each site of subpixels. They assume that a subpixel, at its intended view point, is only affected by two adjacent subpixels of the same color. To obtain the equation coefficients, they measure the crosstalk from the neighbors using an imaging photometer. By solving the linear equations, they find a new intensity value for every subpixel that effectively cancels out the crosstalk effect. Similar approaches follow. Zhou *et al.* [11] consider the same crosstalk model, but they impose the range constraint (i.e., between 0 and 255) on the new intensity values. There might be no solution that exactly satisfies all the linear equations as well as the range constraint. In this reason, Zhou *et al.* formulate the crosstalk cancellation as a constrained least-square problem. To find the solution, they use an iterative algorithm. Wang and Hou [12] do pretty much the same things. But notably they compute,

rather than measure, the equation coefficients by the visible proportion of each subpixel. In more recent work [13], Li *et al.* still assume the same crosstalk model. Instead of imposing the range constraint on the output (i.e., new intensity value), they propose a way to estimate the maximum input range that would keep the output values to be within the valid range. They pre-compress the input dynamic range and simply solve the linear equations. Here, for the computation efficiency, they use inverse filtering, in frequency domain, which exploits the shift-invariance property of the crosstalk model.

Note that, in the crosstalk model, the linear equations actually depend on the display parameters. To compensate for the crosstalk based on the linear equations, we require them to be sufficiently accurate. If not, any attempt for crosstalk removal would not only fail but would also add new artifacts, which strongly necessitates reliable display parameter estimation.

III. OBSERVATION MODEL

In this section, we present a mathematical model for the observation when an image on the panel goes through the optical elements of the autostereoscopic display. For clarity, we particularly consider an autostereoscopic display in which parallax barriers exist in front of the panel.¹ The display has the following four parameters (see Fig. 2):

- (1) Pitch p ;
- (2) Slanted angle α ;
- (3) Gap or Thickness t ;
- (4) Offset σ .

We denote the set of the parameters by Θ , i.e., $\Theta = (p, \alpha, t, \sigma)$. For later use, we also define h and ρ as

$$h = \frac{d}{d-t} \frac{p}{\cos \alpha} \quad (1)$$

$$\rho = \frac{d}{d-t} \sigma, \quad (2)$$

where d is the distance from the panel to the observer's eye. We will use Θ to denote the derivative parameters (h, α, ρ) , not only to denote the primary parameters (p, α, t, σ) , if it makes no confusion.

A. Visible Pixels

The parallax barriers enable every pixel on the panel to have directionality. Based on a projective model, we trace back all the rays that reach the eye position (u, v) at the distance d . Here, for simplicity, we assume a conceptually ideal environment. The panel has infinite size and infinitely high resolution; the barrier slits are infinitesimally narrow; and light does not disperse.

In Fig. 2, we can express the barrier slits as a set of lines, i.e.,

$$S(\Theta) = \left\{ (x', y') : x' = y' \tan \alpha + \sigma + n \frac{p}{\cos \alpha}, \forall n \in \mathbb{Z} \right\}. \quad (3)$$

¹Our subsequent results are not restricted to such type of displays. They are generally applicable to the displays with rear barriers or with lenticular lenses.

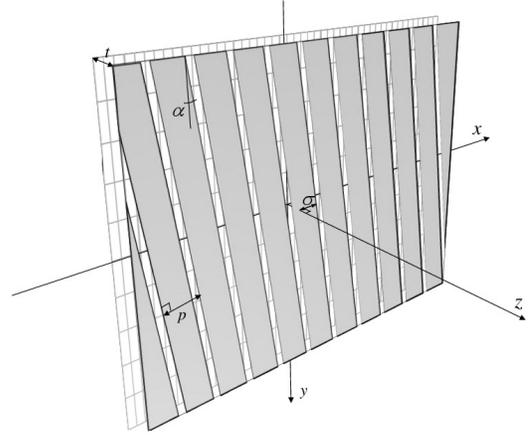


Fig. 2. 3D display with a parallax barrier. The architecture is parameterized by the barrier slit pitch p ; the slanted angle α of the slits; the gap t between the panel and the barrier; and the horizontal offset σ by which the slits are located from the center of the panel.

The ray that passes both the slit (x', y') and the eye position (u, v) must come from a pixel (x, y) such that

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{d}{d-t} \begin{pmatrix} x' \\ y' \end{pmatrix} - \frac{t}{d-t} \begin{pmatrix} u \\ v \end{pmatrix}. \quad (4)$$

By arranging Eq. (4) with respect to (x', y') and subsequently by plugging it into Eq. (3), we obtain the set of pixels that are visible from the eye at (u, v) :

$$P_\gamma(\Theta) = \left\{ (x, y) : x = y \tan \alpha + \rho + \left(n + \gamma(u, v; \Theta) \right) h, \forall n \in \mathbb{Z} \right\}, \quad (5)$$

where

$$\gamma(u, v; \Theta) = \frac{t}{pd} (v \sin \alpha - u \cos \alpha). \quad (6)$$

The eye position is only one-dimensionally parameterized, i.e., via γ . This is because we assumed a line-type barrier that only allows one-dimensional parallax.² Multiple eye positions have a common set of visible pixels if they share the same value of γ . In this reason, we can use γ , what we call *view*, in place of the eye position (u, v) , whenever the exact 2D coordinate is not necessary. Note also that P_γ is periodic, in terms of γ , with the period equal to 1. A space of the 2D eye positions, which forms a single period of γ , i.e., $\gamma \in [0, 1)$, is often called the *primary field of view*. Beyond it, the set of visible pixels simply repeats, extending the effective field of view [19].

We have figured out which pixels are visible at the observer's side. For correct rendering, at the panel's side, the same set of pixels must be controlled to display the correspondent image. This naturally generates the following pixel-view assignment rule:

$$R : P_\gamma(\Theta) \mapsto \gamma. \quad (7)$$

²The same is also true for the cylinder-type lenslet array.

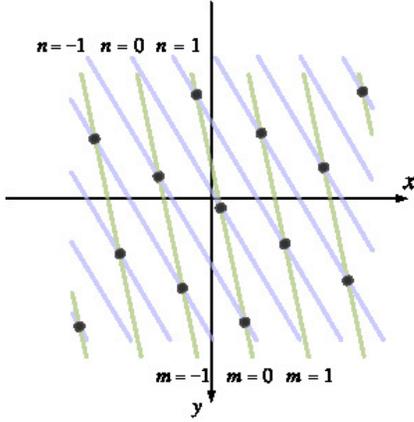


Fig. 3. 2D lattice structure generated by the intersection of $P_{\gamma'}(\Theta_r)$ (greenish lines), and $P_\gamma(\Theta)$ (blueish lines). We assume that $\alpha_r \neq \alpha$. The coordinates of each point are given by Eqs. (10), (11).

We can rewrite Eq. (5) as

$$P_\gamma(\Theta) = \left\{ (x, y) : \frac{x - \rho - y \tan \alpha \pmod{h}}{h} = \gamma \right\}, \quad (8)$$

by eliminating n . The resulting pixel-view assignment function has exactly the same form as van Berkel's [8]. Van Berkel actually derived the function with lenticular lenses, while we did with parallax barriers. In this sense, our derivation complements his work.

B. Misalignment Effects

We emphasize that the display parameters must be known for the pixel-view assignment. We denote, by $\Theta_r = (p_r, \alpha_r, t_r, \sigma_r)$ or $\Theta_r = (h_r, \alpha_r, \rho_r)$, the input parameters with which the 3D rendering module performs the view assignment. Let us partition the pixels on the panel by the assigned view. Then, the set of pixels assigned for view γ' becomes equal to $P_{\gamma'}(\Theta_r)$ (see Eq. 7). The 3D rendering works correctly if, for every view $\gamma \in [0, 1)$, all the visible pixels $P_\gamma(\Theta)$ are assigned with the correct view number. We mathematically express this correct 3D rendering condition as

$$P_{\gamma'}(\Theta_r) \cap P_\gamma(\Theta) = \begin{cases} P_\gamma(\Theta), & \gamma' = \gamma \\ \emptyset, & \gamma' \neq \gamma \end{cases} \quad (9)$$

for every $\gamma, \gamma' \in [0, 1)$. One can easily verify that this condition is simply equivalent to that $\Theta_r = \Theta$.

What would result if $\Theta_r \neq \Theta$? We will evaluate $P_{\gamma'}(\Theta_r) \cap P_\gamma(\Theta)$ to see which pixels in the observation are actually correct and which are wrong. Recall that both $P_{\gamma'}(\Theta_r)$ and $P_\gamma(\Theta)$ represent a bunch of lines parallel and equispaced within each set. We still assume the same idealized environment that we used in identifying visible pixels: panel with infinite resolution, nondispersive light, etc. We will relax the conditions in Section IV. If the lines in both sets have "exactly" the same slope (i.e., $\alpha_r = \alpha$), the intersection can form either another set of lines or the empty set. In this paper, we skip considering this case because it is unlikely to happen in practice.

If the lines in $P_{\gamma'}(\Theta_r)$ have a different slope from those in $P_\gamma(\Theta)$ (i.e., $\alpha_r \neq \alpha$), the intersection produces a set of points:

$$x = \frac{mh_r \tan \alpha - nh \tan \alpha_r}{\tan \alpha - \tan \alpha_r} + \frac{(\rho_r + \gamma' h_r) \tan \alpha - (\rho + \gamma h) \tan \alpha_r}{\tan \alpha - \tan \alpha_r}, \quad (10)$$

$$y = \frac{mh_r - nh + \rho_r + \gamma' h_r - \rho - \gamma h}{\tan \alpha - \tan \alpha_r}, \quad (11)$$

where $(m, n) \in \mathbb{Z} \times \mathbb{Z}$. We make the following observations:

Lattice encoding. The points form a lattice pattern in 2D plane (see Fig. 3). The lattice pattern encodes the actual parameters $\Theta = (h, \alpha, \rho)$ in a certain way.

View-invariant structure. The lattice structure is invariant (only up to a shift) to both view parameters γ and γ' .

Mixed views. With γ fixed and γ' varying (from 0 to 1), the lattice pattern makes a continuous shift along the lines of $P_\gamma(\Theta)$. This implies that rainbow-like view blending will appear in the observation. This phenomenon quite much resembles crosstalk. But remember that it happens due to the wrong assignment of views by the rendering module rather than due to light leakage.³

IV. CALIBRATION

Supposing that we do not know the actual display parameters, how can we estimate them with a high level of accuracy? In this section, we provide a method, based on the observation model that we considered in the previous section.

We seek to obtain all physical parameters, i.e., p, α, t, σ , in our calibration. However, a single observation is not sufficient for unambiguous identification. We may easily show this by example. Consider the following two cases: (1) $p = 0.9975 \times \sqrt{10}$, $\alpha = \arctan(1/3)$, $t = 25$, $\sigma = 0$, (2) $p = 0.9950 \times \sqrt{10}$, $\alpha = \arctan(1/3)$, $t = 50$, $\sigma = 0$, while the observer is commonly at $(u, v) = (0, 0)$, distant from the panel by $d = 10,000$. If we evaluate $P_\gamma(\Theta)$ in Eq. (5), it is exactly the same for both cases. Given the same image, the observer sees exactly the same thing for both cases, not being able to distinguish one case from the other. In this study, we make a couple of observations to get rid of the ambiguity.

For the automated calibration, we use a camera as the measurement device. In this context, hereafter, the position (u, v) , view γ , and the distance d denote those of the camera, not of a human eye. We assume that the camera coordinate (u, v, d) can be known from state-of-the-art camera pose estimation techniques (e.g., [20]). Section IV-C explains the procedure.

Although many recent studies have been proposed for the display calibration, they generally require that a number of observations should be made (see Section II). A few exceptions exist. The work by Hirsch *et al.* [18] and by Lee and Ra [15] needs rather a small number of observations. However,

³Certain studies (e.g., [11]) define "crosstalk" in wide terms, while classifying it into *intrinsic* crosstalk and *extrinsic* crosstalk. What we consider here corresponds to the extrinsic one.

the aim of [18] is basically to help a “human” observer to calibrate the angle (α in our notation) in micro levels, so it does not fit our purpose (i.e., automated calibration dealing with all types of display parameters). In contrast, [15] can be made automated and be made applicable, without difficulty, for finding all display parameters. But its drawback lies in the accuracy. Their model, based on counting color change, suffers from a sort of quantization effect. The counting-based measurements are not easy to obtain to a small fractional precision. Besides, their modeling holds only if the actual display parameters are sufficiently close to the designed ones (or the initial ones). Otherwise, aliasing may happen and every color change may not be observed, violating their assumption.

A. Pattern Images

In our study, we seek to directly observe the lattice encoding pattern (as in Fig. 3) without complication. For this purpose, we prepare the all-white image for a single view $\gamma' = 0$ and all-black images for the other views, and interweave them according to our rendering parameters. The design parameters may be available, and we may want to use them for rendering. However, in our perspective, they are merely one of many values that mismatch the actual parameters. Instead of using those values, we choose rendering parameters such that it can make our analysis as simple as possible. In this study, we use

$$h_r = \beta, \quad \alpha_r = 0, \quad \rho_r = \epsilon \quad (12)$$

with β being a multiple of 3. With the choice, the panel image is filled with vertical stripes of the same color, spaced by β and offset by ϵ . Then, the lattice pattern is reduced to (see Eqs. 10 and 11)

$$L_\gamma(x, y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \left\{ \delta(x - m\beta - \epsilon) \cdot \delta\left(y - m\frac{\beta}{\tan\alpha} + n\frac{h}{\tan\alpha} - \tau\right) \right\}, \quad (13)$$

where

$$\tau = \frac{\epsilon - \rho - \gamma h}{\tan\alpha}. \quad (14)$$

In (13), δ denotes Dirac’s delta function.

B. Display Parameter Estimation

We have assumed several ideal conditions in deriving the lattice pattern $L_\gamma(x, y)$. Now, let us relax them one by one.

(i) *The panel has a finite size.* We denote the finite window by $B(x, y)$, which would typically be a rectangular function without special treatment.

(ii) *Barrier slits have a certain width and light may disperse.* The points on the panel spread (see Fig. 4). We denote the point spread function by $H(x, y)$. The shape actually depends on the slit width, slanted angle α , distance d , etc., so $H(x, y)$ is difficult to specify accurately. However, it commonly delivers some low-pass filter (LPF) characteristics.

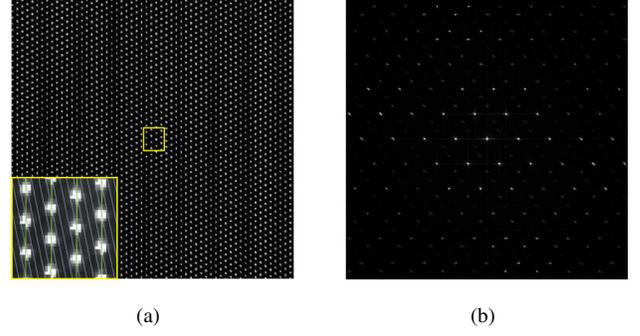


Fig. 4. 2D lattice pattern found in an actual observation. (a) Spatial domain, (b) Frequency domain. The boxed region is shown in magnification, with $P_{\gamma'}(\Theta_r)$ and $P_{\gamma}(\Theta)$ overlaid together (in greenish and blueish color, respectively), for better viewing. Note that a 2D lattice in spatial domain is mapped to another 2D lattice, called *reciprocal lattice* [21], in frequency domain.

(iii) *The observation tends to be contaminated by noise.* We assume, for simplicity, that the noise $N(x, y)$ is independent and identically distributed (i.i.d.) over the pixels.

(iv) *The panel has a finite resolution.* This makes the actual observation be a sampled version of the ideal observation. We postpone considering this nonideality to a little later part.

Taking the first three nonidealities into accounts, we may express the actually observed image I_γ as

$$I_\gamma(x, y) = B(x, y) \left(H(x, y) \star L_\gamma(x, y) \right) + N(x, y), \quad (15)$$

where \star denotes the convolution operator. If we define $L_o(x, y)$ as

$$L_o(x, y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \left\{ \delta(x - m\beta) \cdot \delta\left(y - m\frac{\beta}{\tan\alpha} + n\frac{h}{\tan\alpha}\right) \right\}, \quad (16)$$

$L_\gamma(x, y)$ is merely a shift of $L_o(x, y)$, horizontally by ϵ and vertically by τ (This is due to the view invariance property of the lattice structure; see Section III-B). Therefore, we can rewrite Eq. (15) as

$$I_\gamma(x, y) = B(x, y) \left(H(x, y) \star L_o(x - \epsilon, y - \tau) \right) + N(x, y), \quad (17)$$

where τ is a function of γ as given by Eq. (14).

We further eliminate the shift variation by moving our analysis to the frequency domain. As is well known, a shift leads to the phase modulation in frequency domain and does not affect the signal magnitude. In the frequency domain, Eq. (17) is transformed into

$$\hat{I}_\gamma(f_x, f_y) = e^{-2\pi i(f_x \epsilon + f_y \tau)} \hat{B}(f_x, f_y) \star \left(\hat{H}(f_x, f_y) \hat{L}_o(f_x, f_y) \right) + \hat{N}(f_x, f_y), \quad (18)$$

where we used the multiplication-convolution duality.

Generally, a 2D lattice in spatial domain is mapped to another 2D lattice in frequency domain (see [21], [22] for reference). In our case, $\hat{L}_o(f_x, f_y)$ is given by

$$\hat{L}_o(f_x, f_y) = C \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \left\{ \delta \left(f_x - m \frac{1}{\beta} + n \frac{1}{h} \right) \cdot \delta \left(f_y - n \frac{\tan \alpha}{h} \right) \right\}, \quad (19)$$

where C denotes a constant scale factor. If we insert Eq. (19) into Eq. (18) and make a bit of arrangement, we obtain

$$\hat{I}_\gamma(f_x, f_y) = \sum_{m,n} w_{mn} \hat{B} \left(f_x - \frac{m}{\beta} + \frac{n}{h}, f_y - \frac{n \tan \alpha}{h} \right) + \hat{N}(f_x, f_y), \quad (20)$$

where

$$w_{mn} = C e^{-2\pi i m \frac{1}{\beta} \epsilon} e^{-2\pi i n (\frac{\tan \alpha}{h} \tau - \frac{1}{h} \epsilon)} \hat{H} \left(\frac{m}{\beta} - \frac{n}{h}, \frac{n \tan \alpha}{h} \right). \quad (21)$$

We interpret Eq. (20) as follows: The signal part forms a mixture of functions \hat{B} . The functions would be sincs if $B(x, y)$ were a rectangular window. We use a simple trick to make $B(x, y)$ be Gaussian. We multiply Gaussian weights to the pixel values. Then, in our case, the signal part of \hat{I}_γ becomes a mixture of Gaussians, rather than a mixture of sincs, because a Gaussian remains as a Gaussian (with a different bandwidth) in frequency domain. The mixture weights w_{mn} are proportional to \hat{H} , which has the LPF characteristics as aforementioned. This means that the amplitude of a Gaussian is high around the center of the frequency axes and diminishes away from the center.

Finally, we attempt to consider the remaining nonideality that the panel has a finite resolution. In the frequency domain, this poses two implications. First, the signal replicates due to the sampling effect. The replica might be able to corrupt the original signal, behaving like noise. But recall that the Gaussians far from the origin have attenuated weights. When they are replicated at the multiples of the sampling frequency and then stretched to around the origin, the magnitudes become negligible. Therefore, even with the sampling effects, Eq. (18) holds around the origin quite well. Second, the resolution of the observation may be so low that we cannot directly extract the parameters from the observation, to a sufficient level of accuracy. We will soon show how to deal with this issue in our calibration.

Horizontal pitch h , Slanted angle α : Given $\hat{I}_\gamma(f_x, f_y)$, we detect a Gaussian peak around the origin. To accurately localize the peak, we employ paraboloid fitting around the initial ballpark. The paraboloid fitting can be performed quite efficiently because the closed-form solution is available. A similar approach was proposed by Cho *et al.* [23] in calibrating light-field cameras. But our scheme includes a notably distinct feature. Recall that our signal around its peak is almost a Gaussian because we cooked it by means of the window function $B(x, y)$. In log domain, it “indeed” becomes a

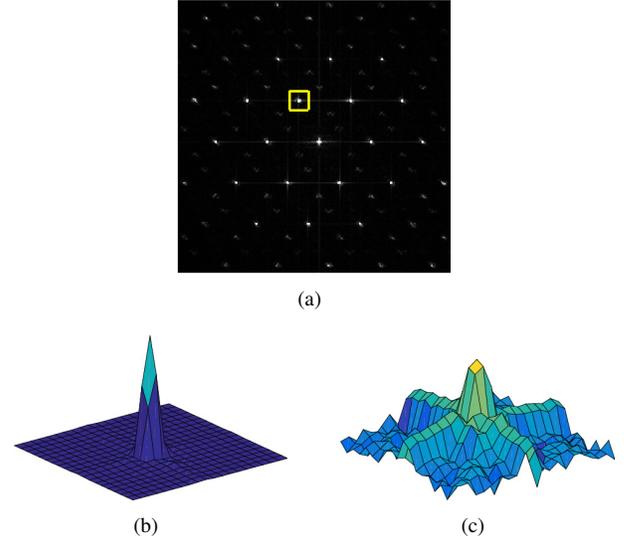


Fig. 5. Accurate peak localization. (a) Initial ballpark peak point and the neighborhood. (b) Magnified plot. The shape around the peak is almost a (coarsely sampled) Gaussian. (c) Magnified plot in log domain. The shape around the peak is almost a (coarsely sampled) paraboloid. The peak point is estimated by the paraboloid fitting in the log domain.

paraboloid (see Fig. 5), so the paraboloid fitting in the log domain should work very well.

With no noise involved, this provides us

$$f_x^* = \frac{m}{\beta} - \frac{n}{h}, \quad f_y^* = \frac{n \tan \alpha}{h} \quad (22)$$

for some $m, n \in \mathbb{Z}$. In fact, noise is well suppressed in frequency domain. We will argue that the signal-to-noise ratio (SNR) around a Gaussian peak is quite high in frequency domain. To see this, let us consider the signal power and the noise power around peaks. First, due to the unitary property of the Fourier transform, the *total* signal power is preserved (this is known as Parseval’s theorem [24]). But the number of lattice points is usually far fewer in frequency domain than in spatial domain (e.g., see Fig. 4), and it implies that the signal power is concentrated on the points (or peaks). On the other hand, noise power is evenly spread over all the frequencies because the noise is assumed to be i.i.d. over the pixels and the Fourier transform is unitary. Therefore, the SNR becomes quite high around a peak in frequency domain (while quite low around a valley).

From Eq. (22), we obtain

$$h = \frac{n}{m/\beta - f_x^*}, \quad \alpha = \arctan \frac{f_y^*}{m/\beta - f_x^*}. \quad (23)$$

If m and n are somehow known, Eq. (23) directly gives us the two parameters h and α . Otherwise, h and α remain in uncertainty, but they are still restricted to be in a discrete set of candidates. To break the ambiguity, we actually multiplex two patterns in a panel image (with different values of β and ϵ). To be specific, we use $\beta = 15$, $\epsilon = 0$ for one pattern and $\beta = 24$, $\epsilon = 1$ for the other. Multiplexed together, these patterns generate the panel image that consists of two sets of stripes, each differently colored and differently spaced. In the observation, the two stripe patterns can easily be separated.

Then, for each pattern $j = 1, 2$, we construct the candidate set $\mathcal{C}_j(h, \alpha)$ using Eq. (23) for $m, n \in \mathbb{Z}$. Then, we intersect the candidate sets $\mathcal{C}_1, \mathcal{C}_2$ to find h and α that can explain both patterns at the same time. In practice, however, noisy factors or numerical imprecisions are likely to make the “exact” intersection be empty. So, we actually compute h and α such that they are close to \mathcal{C}_1 and \mathcal{C}_2 at the same time:

$$(h, \alpha) = \arg \min_{(h', \alpha')} \min_{\substack{(h, \alpha)_j \in \mathcal{C}_j \\ j=1,2}} \sum_{j=1}^2 \left\| (h', \alpha') - (h, \alpha)_j \right\|^2. \quad (24)$$

Pitch p , Gap or Thickness t : To further identify the physical parameters p and t , we make a secondary observation at a different distance from the first one. Let d_k denote the distance of the camera from the panel for the k th observation ($k = 1, 2$). We meant that $d_1 \neq d_2$. We assume that the distances are known by a camera estimation technique (see Section IV-C) and that we have already estimated the slanted angle α and the horizontal pitch h_k for each observation k .

For each observation, the horizontal pitch is related to the physical pitch p and gap t by (see Eq. 1)

$$h_k = \frac{d_k}{d_k - t} \cdot \frac{p}{\cos \alpha}, \quad k = 1, 2. \quad (25)$$

It is straightforward to solve Eq. (25). The solution is given by

$$p = \frac{h_1 h_2 (d_2 - d_1) \cos \alpha}{d_2 h_1 - d_1 h_2}, \quad (26)$$

$$t = \frac{d_1 d_2 (h_1 - h_2)}{d_2 h_1 - d_1 h_2}. \quad (27)$$

Offset ρ, σ : The offset parameters only affect how much the observed image $I_\gamma(x, y)$ is shifted from the lattice pattern $L_o(x, y)$. Given all other display parameters as well as the camera pose, the amount of the shift can be robustly estimated by finding ζ that best matches $I_\gamma(x + \epsilon, y + \zeta)$ to $L_o(x, y)$. In frequency domain, the operation is implemented by

$$\tau = \arg \max_{\zeta} \left| \sum_{m,n} \hat{I}_\gamma \left(\frac{m}{\beta} - \frac{n}{h}, \frac{n \tan \alpha}{h} \right) \cdot e^{2\pi i (m \frac{1}{\beta} \epsilon + n (\frac{\tan \alpha}{h} \zeta - \frac{1}{h} \epsilon))} \right|. \quad (28)$$

After computing τ , we can estimate the parameter ρ as (see Eq. 14)

$$\rho = \epsilon - \tau \tan \alpha - \gamma h. \quad (29)$$

where γ is evaluated in advance, by Eq. (6), on the basis of the known parameters. Then, we can subsequently find the offset σ , by Eq. (2),

$$\sigma = \frac{d - t}{d} \rho. \quad (30)$$

C. Camera Pose Estimation

We assume that the intrinsic camera parameters (e.g., focal length, principal point, radial distortion coefficients) are known. Otherwise, camera calibration (e.g., see [25]) can be conducted in advance to the 3D display calibration.

To estimate camera poses, we exploit a state-of-the-art method in that field, called *homography decomposition* [20]. We detect four corners of the display in the captured image and find homography between the world coordinates and image coordinates. The homography can be decomposed into the intrinsic parameter matrix and the camera pose matrix. As assumed, we know the intrinsic parameter matrix, so we can compute the camera pose matrix by multiplying its inverse to the homography. The camera pose matrix tells us the camera rotation and translation relative to the origin (i.e., panel center) of the world coordinate. The translation vector exactly corresponds to the camera position (u, v, d) .

If the camera is rotated, the shape of the display may appear geometrically distorted in the captured image. For our display calibration, we need it to be rectified. This could be done with a usual image rectification technique using the computed rotation matrix. In this study, we use an equivalent yet simpler scheme that warps the four corners to the vertices of a rectangle of the panel resolution.

V. EXPERIMENTS

We conduct two types of experiments to evaluate the performance of the proposed calibration scheme. In the first set of experiments, we generate a synthetic dataset from POV-Ray simulation [26], for diverse “virtual” 3D displays. The ground-truth display parameters are available for the synthetic dataset, so we can quantitatively evaluate the estimation error. We also run the experiments in noisy settings and see how robust the proposed scheme is. In the second set of experiments, we apply the proposed scheme to real-life autostereoscopic displays. For those displays, the actual display parameters are generally unknown, which makes it difficult to directly compute the estimation error. In this case, we evaluate the performance in the visual aspect of the observed images when rendering is done using the estimated display parameters. We assess the visual quality in subjective terms as well as two objective measures (i.e., peak-signal-to-noise ratio, structural similarity index).⁴

A. Synthetic Dataset Calibration

For the synthetic dataset, we build a simulation environment based on a ray tracing tool called POV-Ray [26]. We consider three types of 3D displays as listed in Table I where the designed parameters $\Theta_o = (p_o, \alpha_o, t_o, \sigma_o)$ are shown together. For misalignment, we add perturbations to the designed parameters. We assume that the perturbations are uniformly

⁴Some test images are from public domain (Big Buck Bunny ©Blender Foundation, UCSD/MERL Light Field Repository [27], The Stanford Light Field Archive [?], MIT Synthetic Light Field Archive [?]), and others are graphically generated by ourselves.

TABLE I
3D DISPLAYS IMPLEMENTED IN OUR SIMULATION ENVIRONMENT
UNITS IN PARAMETER SPEC.: MILLI-METERS (LENGTH), DEGREES (ANGLE)

Display	Size (Resolution)	Optical elements	Designed parameters			
			p_o	α_o	t_o	σ_o
FHD55B	55 in (1 920×1 080)	Barrier	1	18	4	0.5
UHD32B	32 in (3 840×2 160)	Barrier	0.5	10	2	0.2
WQXGA10L	10 in (2 560×1 600)	Lenticular	0.1	12	1	0.05

distributed over the following range:

$$\frac{|p - p_o|}{p_o} < 0.01, \quad \frac{|\alpha - \alpha_o|}{\alpha_o} < 0.01, \\ \frac{|t - t_o|}{t_o} < 0.01, \quad \frac{|\sigma - \sigma_o|}{p_o} < 0.01. \quad (31)$$

For each type of display, we implement 10 instances, each with independent random perturbations.

We capture two images for each display instance at different distances. In the simulation, we place the camera at some random positions, deviated up to ± 50 mm in all directions, rotated by a small random angle φ ($|\varphi| < 1^\circ$) along a random axis. The captured images have an ultra-high-definition (UHD) resolution, i.e., $3\,840 \times 2\,160$. Given two captured images, we first conduct rectification as well as the camera pose estimation (see Section IV-C). The rectified images are moved to frequency domain, using the discrete Fourier transform, in which the display parameters are estimated sequentially through several steps (see Section IV-B). Table II presents the estimation results and makes comparisons with those obtained with the calibration based on prior work [15]. We have implemented [15] by ourselves. The original scheme, based on counting color changes, can be used to estimate h and α , but it does not actually give a way to estimate p and t . In our implementation, we compute p and t just as in the proposed method (following Eqs. 26, 27), on top of their estimate of h and α . For the offset parameter σ , we follow their strategy of taking additional pictures after correcting p , t , α . Specifically, we only turn on the center view pixels and photograph the observed images from all viewing zones. The offset parameter can be easily computed by spotting which of the pictures is the brightest. In the table, we observe that the proposed scheme works quite well, showing several times higher accuracy than prior work.

We run similar experiments in noisy settings. We add Poisson noise to simulate photon shot noise in usual photographic images [28]. The results are shown in Fig. 6. As expected, the mean absolute errors increase with the level of noise, but the rate is not high. The errors are maintained quite low even with a significant amount of noise. The offset parameter σ turns out to be most vulnerable to the noise. This is probably due to error accumulation; the estimate of the offset parameter depends on those of the other parameters in our calibration. We plan to handle this issue, error accumulation, in our future work.

TABLE II
PARAMETER ESTIMATION ERRORS
THE SHOWN NUMBERS ARE THE MEAN AND STANDARD DEVIATION
(IN PARENTHESIS) OF THE ABSOLUTE ERRORS
OVER 10 INSTANCES PER EACH TYPE OF DISPLAY.

Calibration based on prior work [15]				
Display	Estimation errors (length: mm, angle: $^\circ$)			
	$ \Delta p $	$ \Delta \alpha $	$ \Delta t $	$ \Delta \sigma $
FHD55B	0.0099 (0.0254)	4.18×10^{-4} (5.99×10^{-4})	0.1639 (0.2399)	0.0460 (0.0154)
UHD32B	0.0356 (0.0818)	1.94×10^{-4} (3.58×10^{-4})	0.1455 (0.2680)	0.0173 (0.0043)
WQXGA10L	0.0042 (0.0077)	1.02×10^{-5} (7.76×10^{-6})	0.1980 (0.1340)	0.0048 (0.0063)

Calibration based on the proposed method				
Display	Estimation errors (length: mm, angle: $^\circ$)			
	$ \Delta p $	$ \Delta \alpha $	$ \Delta t $	$ \Delta \sigma $
FHD55B	0.0027 (0.0011)	4.21×10^{-5} (3.96×10^{-5})	0.0296 (0.0250)	0.0281 (0.0175)
UHD32B	0.0017 (0.0005)	2.98×10^{-5} (1.77×10^{-5})	0.0304 (0.0190)	0.0069 (0.0040)
WQXGA10L	0.0061 (0.0011)	4.80×10^{-6} (4.17×10^{-6})	0.1643 (0.0530)	0.0069 (0.0019)

B. Real-Life Display Calibration

We apply the proposed method on real-life autostereoscopic displays as well.

First, we calibrate a 32 in ultra-high-definition (UHD) display (with resolution $3\,840 \times 2\,160$) that has a parallax barrier between the panel and the back-light unit. For the camera, we use Point Grey's Flea[®]3 of the resolution $4\,096 \times 2\,160$. We take two photos, each nearly at $d_1 \approx 700$ mm and at $d_2 \approx 1\,000$ mm, respectively. The calibration is conducted on a PC workstation equipped with Intel Xeon E5-1660 processor and 16GB memory. Given the input images, the entire process has only taken as little as 1.99 s. Fig. 7 exhibits the visual quality improvement by calibration for five example images.

The images in Fig. 7 have been photographed nearly at the center position and nearly at the optimal viewing distance, while the 3D rendering is performed according to different sets of display parameters – (b) with the designed parameters; (c) with the parameters calibrated based on prior work [15]; (d) with the parameters calibrated based on the proposed method. In Fig. 7(b), the edges in wall bricks, rabbit's ear, elephant's nose, and door frames appear as if multiply-exposed, and the Chinese character on the background of *Happy Buddha* suffers from severe geometric distortion. In Fig. 7(c), the distortions have been much alleviated but still remain. They almost disappear in Fig. 7(d). Fig. 7 also shows the visual quality metrics in terms of peak-signal-to-noise-ratio (PSNR) as well as structural similarity (SSIM) index [29]. Both measures require a reference image for the quality assessment. For the reference, we must be able to observe the misalignment-artifact-free image through the optical elements. To do this, we carefully identify the view γ corresponding to the camera position and subsequently display the viewpoint image without giving any disparity. Then, what the observer will actually get is a 2D image, not a 3D image, but nevertheless it delivers

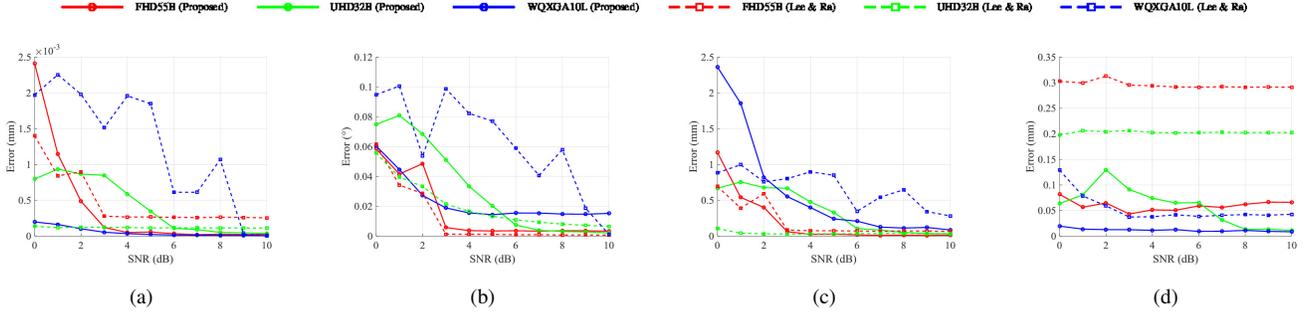


Fig. 6. Mean absolute errors for display parameter estimation in noisy settings. (a) Pitch p , (b) Slanted angle α , (c) Gap or Thickness t , (d) Offset σ . For each sub-figure, we compare the performance between the calibration based on prior work [15] and the proposed calibration method. We find that the errors increase with the noise but not quite rapidly.

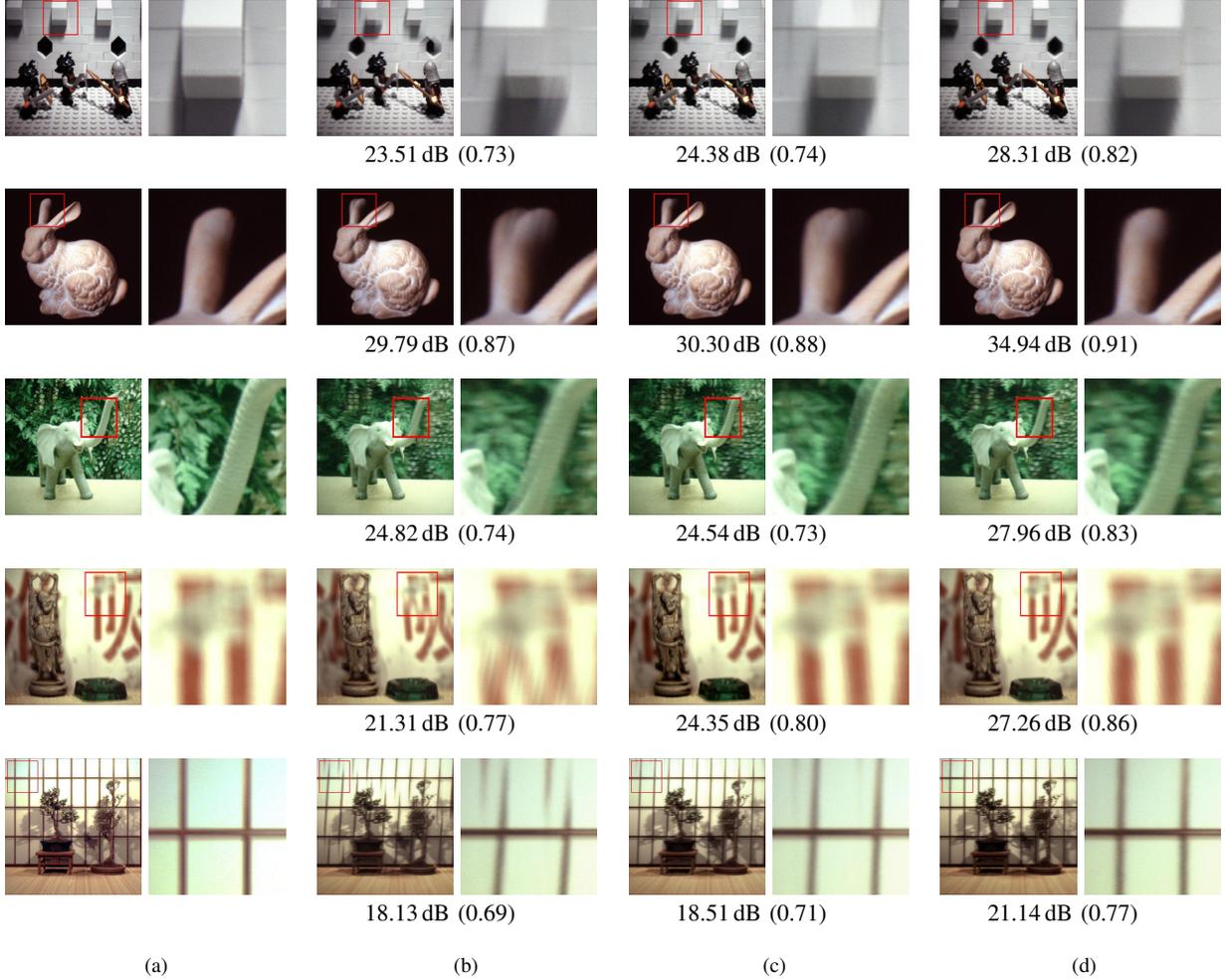


Fig. 7. Example images observed on the 32 in 3D display. The images *Lego Knights*, *The Stanford Bunny*, *Elephant*, *Happy Buddha*, *Bonsai* (from top to bottom) are rendered using different sets of parameters and photographed nearly at the center position. (a) Reference, (b) No-calibration, (c) Calibration based on prior work [15], (d) Calibration based on the proposed method. The quality improvement by the proposed calibration is easily noticeable. Two objective measures (PSNR in dB and SSIM index on a scale of 0 to 1) are given below each image. For better comparison, some regions are shown in magnification.

the correct target contents (in reduced resolution but without misalignment effects) for the specific viewpoint where the camera is located. The proposed scheme consistently shows higher scores, in both objective criteria, than the others (by 4–7 dB in PSNR, 0.06–0.1 in SSIM index on average). An interesting point is that, in this case, the maximum difference among the parameters is quite small – smaller than $1 \mu\text{m}$ for

pitch, smaller than 0.1° for slanted angle. But as manifested in the figure, the visual quality varies a lot, indicating why accurate calibration is vital to 3D displays.

Next, we calibrate a tablet PC with 2560×1600 pixels on its 10 in panel. A detachable lenticular sheet serves as the frontal cover of the tablet (see Fig. 8(a)). With the cover flipped open, the tablet normally behaves as a 2D display, but it turns into

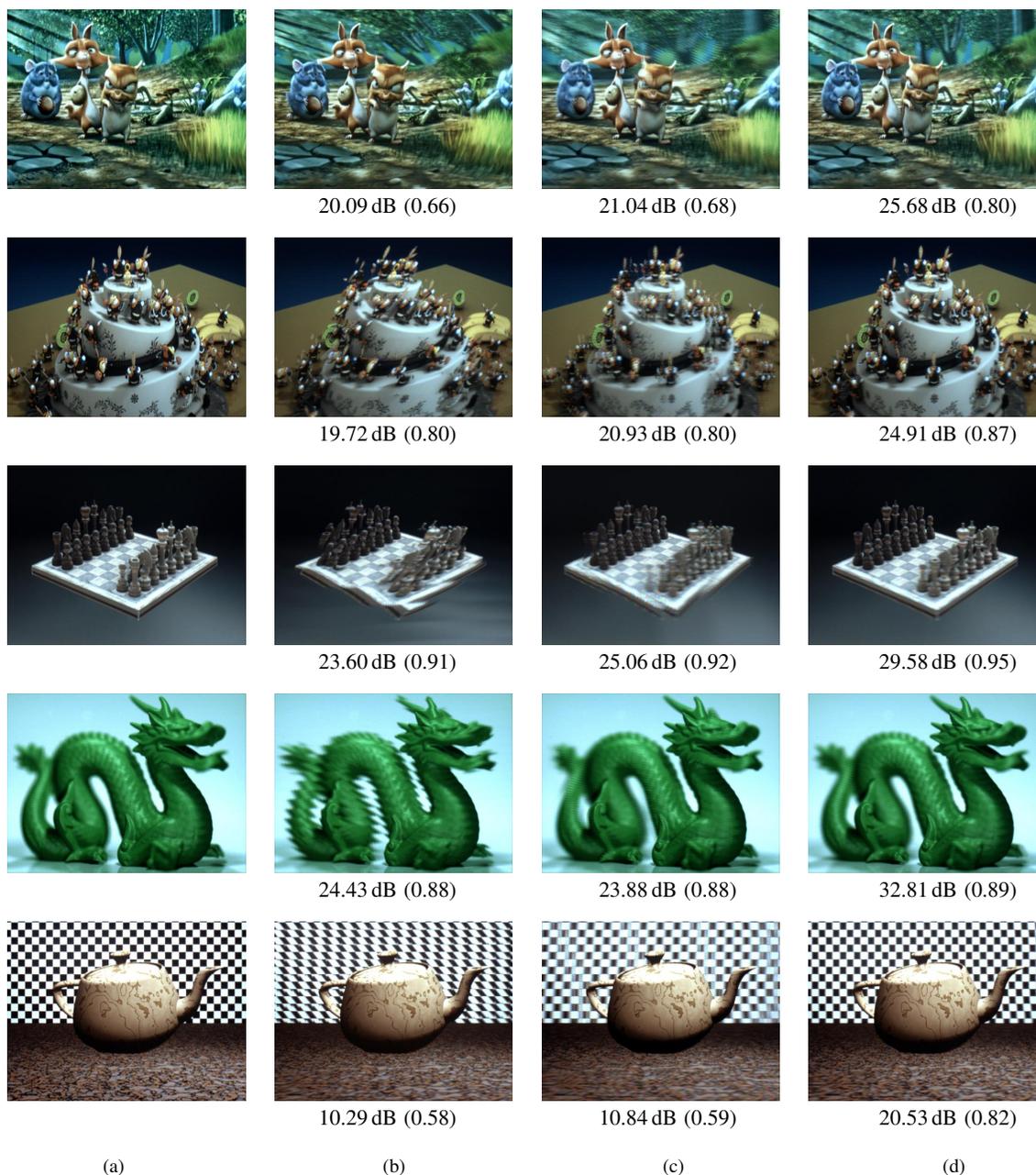


Fig. 9. Example images observed on the 10 in mobile 3D display. The images *Big Buck Bunny*, *Cake*, *Chess*, *Green Dragon*, *Teapot* (from top to bottom) are rendered using different sets of parameters and photographed nearly at the center position. (a) Reference, (b) No-calibration, (c) Calibration based on prior work [15], (d) Calibration based on the proposed method. The quality improvement by the proposed calibration is clearly noticeable, particularly along edges and in the objects with large binocular disparities. Two objective measures (PSNR in dB and SSIM index on a scale of 0 to 1) are given below each image. Notice that more distortions happen without calibration, in comparison with Fig. 7, which is due to a larger amount of optical misalignment. However, when calibrated using the proposed scheme, the resulting images have good visual quality nearly equal to those in Fig. 7.

a 3D display when the cover is closed. This kind of design enables 2D/3D switching at little cost. But, on each flip, the lenticular sheet is likely to move from where it was previously, demanding a new instantaneous calibration. We can use the built-in camera of the tablet PC, together with a mirror, for the calibration, as conceptually depicted in Fig. 8(b). Except that the captured images need to be flipped horizontally, the calibration procedure remains the same. Fig. 9 shows four example images rendered with the calibrated display parameters. The image distortion has remarkably decreased

in comparison with two other schemes: (i) no-calibration, (ii) calibration based on prior work [15]. Fig. 10 shows example images photographed at multiple view points to verify the motion parallax along the horizontal direction. We see that, in the figure, the recognizability of the motion parallax is also enhanced quite much by fixing the rendering parameters.

VI. CONCLUSION

Despite the recent advancement, 3D displays have faced several issues to solve, regarding the visual quality, for active

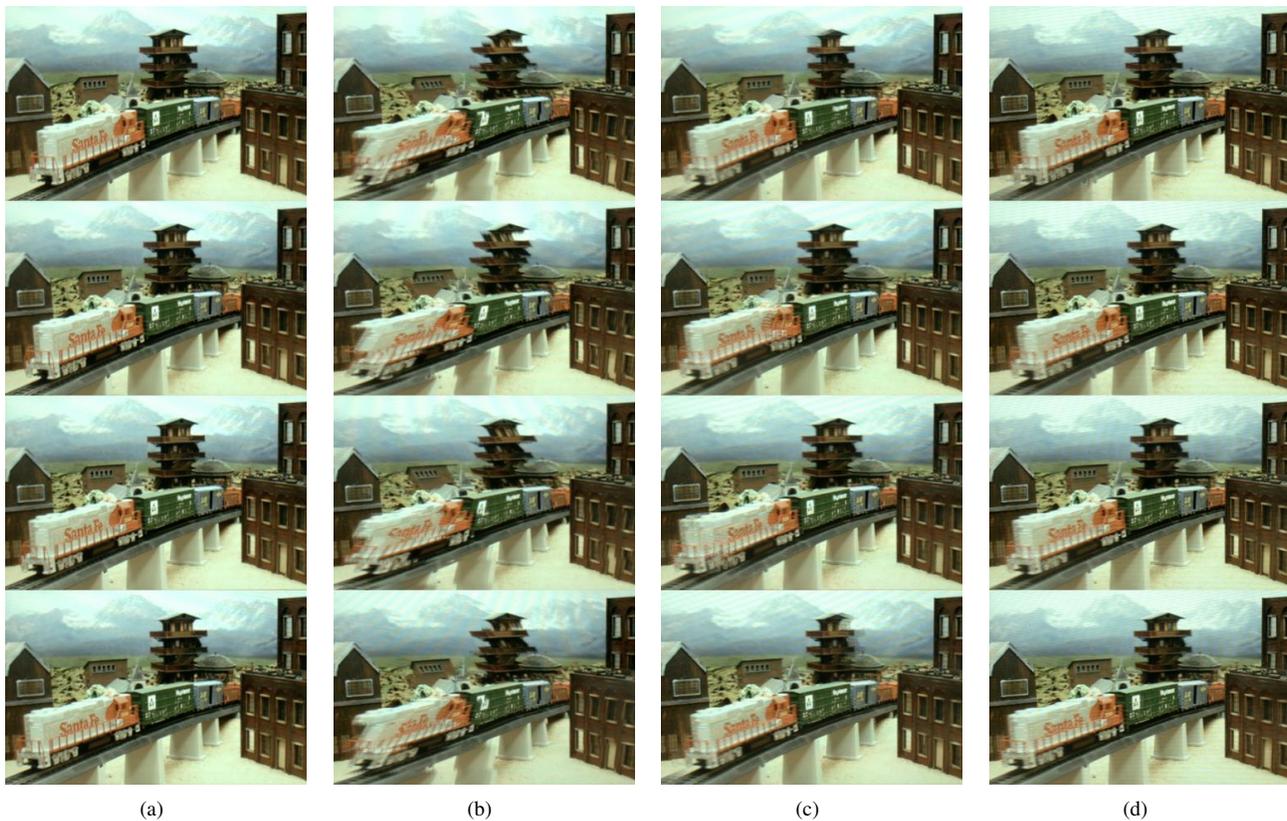


Fig. 10. Example images observed at four different view points. The image *Train* is rendered using different sets of parameters. (a) Reference, (b) No-calibration, (c) Calibration based on prior work [15], (d) Calibration based on the proposed method. The motion parallax is clearly observed in (d). The train, buildings, and mountains move by different amounts and/or in different directions as the view point changes.

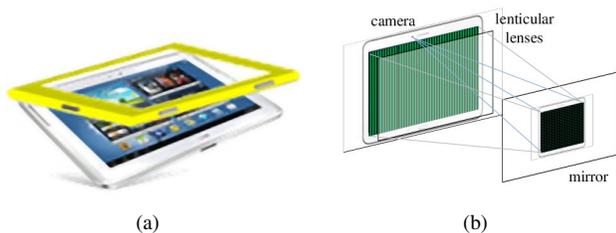


Fig. 8. Mobile 3D display calibration using on-board camera. (a) 3D display implemented on a tablet PC. (b) Calibration using the on-board camera and a mirror. The tablet has a detachable lenticular cover for 2D/3D switching. With the cover closed, the tablet behaves like a 3D display. The calibration can be performed with the on-board camera. A pattern image is displayed on the panel, and then a couple of images reflected by the mirror are captured with the camera. Each time, the mirror must be at different positions. Except that the images need to be flipped horizontally, the calibration procedure remains the same.

proliferation. Some of them are fundamental, unavoidable even if the display behaves without error (e.g., resolution reduction, light leakage). But others are not, while being as important issues. In this paper, we dealt with one such matter – correcting image distortions caused by optical misalignment.

First, we established an observation model when the observer looks at images through the optical elements of the 3D display. We also considered what the observation would be like if misalignment happens. Then, we proposed a 3D display calibration which decodes the correct display parameters from

the observation. Given two photos of a pattern image displayed on the 3D panel, each taken at different positions, the analysis is conducted in frequency domain. All procedures are fully automated and the computation time spent on estimating all display parameters (i.e., pitch, slanted angle, gap, and offset) is less than two seconds. The efficiency of the proposed method makes it applicable on-the-fly whenever it is necessary. In a set of experiments with synthetic dataset, the proposed calibration method showed quite high accuracy, with the estimation error 0.0031° for slanted angle, $0.02\ \mu\text{m}$ for pitch, $74\ \mu\text{m}$ for offset, $13\ \mu\text{m}$ for gap on average, a half order of magnitude higher than prior work. With real-life displays, the proposed method has also demonstrated a significant improvement of visual quality of the observed images.

REFERENCES

- [1] H. Hwang, J. Park, H. S. Chang, Y. J. Jeong, D. Nam, and I. S. Kweon, "Lenticular lens parameter estimation using single image for crosstalk reduction of three-dimensional multi-view display," in *Proc. SID Symp. Dig. Tech. Papers*, 2015, pp. 1417–1420.
- [2] N. A. Dodgson, "Optical devices: 3D without the glasses," *Nature*, vol. 495, no. 7441, pp. 316–317, 2013.
- [3] H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the art in stereoscopic and autostereoscopic displays," *Proc. IEEE*, vol. 99, no. 4, pp. 540–555, 2011.
- [4] N. A. Dodgson, "Analysis of the viewing zone of multiview autostereoscopic displays," in *Proc. of SPIE: Stereoscopic Displays and Virtual Reality Systems IX*, vol. 4660, 2002.

- [5] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. Von Kopylow, "A survey of 3DTV displays: Techniques and technologies," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1647–1658, 2007.
- [6] S. Winkler and D. Min, "Stereo/multiview picture quality: Overview and recent advances," *Signal Process.: Image Commun.*, vol. 28, no. 10, pp. 1358–1373, 2013.
- [7] K. Mueller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View synthesis for advanced 3D video systems," *EURASIP J. Image Video Process.*, vol. 2008, no. 1, pp. 1–11, 2008.
- [8] C. van Berkel, "Image preparation for 3D-LCD," in *Proc. of SPIE: Stereoscopic Displays and Virtual Reality Systems VI*, vol. 3639, 1999.
- [9] T. Järvenpää and M. Salmimaa, "Optical characterization of autostereoscopic 3-D displays," *J. Soc. Inform. Display*, vol. 16, no. 8, pp. 825–833, 2008.
- [10] X.-F. Li, Q.-H. Wang, D.-H. Li, and A.-H. Wang, "Image processing to eliminate crosstalk between neighboring view images in three-dimensional lenticular display," *J. Display Technol.*, vol. 7, no. 8, pp. 443–447, 2011.
- [11] M. Zhou, H. Wang, W. Li, S. Jiao, T. Hong, S. Wang, X. Sun, X. Wang, J.-Y. Kim, and D. Nam, "A unified method for crosstalk reduction in multiview displays," *J. Display Technol.*, vol. 10, no. 6, pp. 500–507, 2014.
- [12] X. Wang and C. Hou, "Improved crosstalk reduction on multiview 3D display by using BILS algorithm," *J. Appl. Math.*, 2014.
- [13] D. Li, D. Zang, X. Qiao, L. Wang, and M. Zhang, "3D synthesis and crosstalk reduction for lenticular autostereoscopic displays," *J. Display Technol.*, vol. 11, pp. 939–946, 2015.
- [14] Q.-H. Wang, X.-F. Li, L. Zhou, A.-H. Wang, and D.-H. Li, "Crosstalk reduction by correcting the subpixel position in a multiview autostereoscopic three-dimensional display based on a lenticular sheet," *Appl. Opt.*, vol. 50, no. 7, pp. B1–B5, 2011.
- [15] Y.-G. Lee and J. B. Ra, "Image distortion correction for lenticula misalignment in three-dimensional lenticular displays," *Opt. Eng.*, vol. 45, no. 1, pp. 017007–017007, 2006.
- [16] —, "New image multiplexing scheme for compensating lens mismatch and viewing zone shifts in three-dimensional lenticular displays," *Opt. Eng.*, vol. 48, no. 4, pp. 044001–044001, 2009.
- [17] J. Ge, D. Sandin, T. Peterka, T. Margolis, and T. DeFanti, "Camera based automatic calibration for the Varrier-system," in *Proc. IEEE Computer Vision and Pattern Recognition Workshops*, 2005, pp. 110–110.
- [18] M. Hirsch, D. Lanman, G. Wetzstein, and R. Raskar, "Construction and calibration of optically efficient LCD-based multi-layer light field displays," in *J. Phys. Conf. Ser.*, vol. 415, no. 1, 2013, p. 012071.
- [19] S.-P. Du, P. Didyk, F. Durand, S.-M. Hu, and W. Matusik, "Improving visual quality of view transitions in automultiscopic displays," in *Proc. ACM Int. Conf. Computer Graphics and Interactive Techniques*, 2014.
- [20] E. Malis and M. Vargas, "Deeper understanding of the homography decomposition for vision-based control," INRIA, Tech. Rep., 2007.
- [21] C. Kittel, *Introduction to Solid State Physics*. John Wiley & Sons, Inc., 2005.
- [22] P. Agniel, "Subsampling models and anti-alias filters for 3-D automultiscopic displays," Master's thesis, Dept. Elect. Comput. Eng., Boston Univ., Boston, MA, 2004.
- [23] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, "Modeling the calibration pipeline of the Lytro camera for high quality light-field image reconstruction," in *Proc. IEEE Int. Conf. Computer Vision*, 2013, pp. 3280–3287.
- [24] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*. Prentice Hall, 1996.
- [25] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 1330–1334, 2000.
- [26] D. K. Buck and A. A. Collins, "POV-Ray – the persistence of vision raytracer," [Online]. Available: <http://www.povray.org/>.
- [27] UCSD/MERL, "Light field repository," [Online]. Available: <http://vision.ucsd.edu/datasets/lfarchive/lfs.shtml>.
- [28] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*. John Wiley & Sons, Inc., 1991.
- [29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, pp. 600–612, Apr. 2004.