



Optimal binary representation via non-convex optimization on tomographic displays

SUYEON CHOI, SEUNGJAE LEE, YOUNGJIN JO, DONGHEON YOO,
DONGYEON KIM, AND BYOUNGHO LEE*

School of Electrical and Computer Engineering, Seoul National University, Gwanak-Gu Gwanakro 1, Seoul, 08826, South Korea

*byoungcho@snu.ac.kr

Abstract: There have been many recent developments in 3D display technology to provide correct accommodation cues over an extended focus range. To this end, those displays rely on scene decomposition algorithms to reproduce accurate occlusion boundaries as well as retinal defocus blur. Recently, tomographic displays have been proposed with improved trade-offs of focus range, spatial resolution, and exit-pupil. The advantage of the system partly stems from a high-speed backlight modulation system based on a digital micromirror device, which only supports 1-bit images. However, its inherent binary constraint hinders achieving the optimal scene decomposition, thus leaving boundary artifacts. In this work, we present a technique for synthesizing optimal imagery of general 3D scenes with occlusion on tomographic displays. Requiring no prior knowledge of the scene geometry, our technique addresses the blending issue via non-convex optimization, inspired by recent studies in discrete tomography. Also, we present a general framework for this rendering algorithm and demonstrate the utility of the technique for volumetric display systems with binary representation.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Since augmented reality and virtual reality emerged as one of the most promising next-generation displays, we have been observing unprecedented rapid growth of 3D display industries. There have been several efforts for realizing an ultimate 3D display system to deliver immersive and realistic experience to users. Among the candidates, head-mounted displays (HMDs) have been settled down as a competitive platform that allows users to experience augmented reality and virtual reality. However, commercially available HMDs have a limitation in providing accurate focus cues. Because the physical focal plane of ordinary HMDs is fixed on the specific depth, users may experience vergence-accommodation conflict (VAC) which may involve visual fatigue and prevent immersive experience [1].

To provide accurate focus cues for mitigating VAC, several display systems for HMDs have been explored and studied. Recently, volumetric displays with dense focal stacks were introduced, which employ a digital micromirror device (DMD). Combining DMD projection system with a liquid crystal display (LCD) panel [2] or a high dynamic range light emitting diode [3, 4], the prototypes generate a large number of focal planes without the significant degradation in frame rate or bit-depth. These prototypes may provide users with accurate and continuous focus cues along the wide depth range. Especially, tomographic displays have shown superior performance in resolution and exit-pupil, compared to other candidates for HMDs with focus cues. Nevertheless, the representation of occlusion boundaries in volumetric scenes has been pointed out as a distinct drawback [2].

Since focal plane images of volumetric displays are synthesized via addition, users may notice artifacts at occlusion boundaries of 3D scenes. For multi-plane displays that also suffer from similar limitation, various computational approaches have been introduced to alleviate the occlusion boundary artifacts [5–16]. For instance, we can find optimal focal plane images by

solving a least squares problem that minimizes errors between ground truth and reconstructed retinal or pupil view images. This method, however, is not applicable for volumetric displays because of a distinct optical principle involved by the DMD system. When the DMD projection system is applied, focal plane images are determined by binary states. Therefore, determination of focal plane images has a binary constraint so that we need to solve a binary least squares problem.

The binary least squares problem is categorized as non-deterministic polynomial-time hardness (NP-hard) problem, which requires more complicated algorithm to solve [17]. Lee et al. presented an algorithm for their system to mitigate occlusion boundary artifacts, which solves relaxation of the binary least squares problem [2]. However, the convergence and the performance of the algorithm are sensitive to an initial condition which is determined by 3D information of reconstructed scenes: It uses in-focus color images and depth maps from tens of viewpoints to set the initial condition. Additionally, the algorithm is not applicable for other volumetric displays that have similar binary constraints. In summary, we note that the binary least squares problem for volumetric displays has not been solved using a stable and adequate algorithm.

Here, we introduce a non-convex optimization algorithm that has advantages in stability and versatility. Our algorithm was inspired by previous researches related to discrete tomography that solves the binary least squares problem for reconstruction of volumetric samples [18–26]. The algorithm can be applied for most volumetric displays using binary modulation to present realistic volumetric scenes [2–4]. Furthermore, the convergence of the proposed algorithm has tolerance for the initial condition. In this study, we first explore the principle of tomographic displays and theoretical background of discrete tomography related to the proposed algorithm. Second, our algorithm is thoroughly analyzed and compared with the related approaches in terms of the convergence, accuracy, and optimization time. In conclusion, we demonstrate the versatility by applying our algorithm to various applications. We believe our work can extend the potential of tomographic display systems.

2. Background

Prior to proposing a way to explore an optimal set of display and backlight patterns for tomographic displays, we need to elaborate on the system and related algorithms. First, we introduce the concept of the tomographic displays. We elucidate the current limitations of tomographic displays, which comes from its unique structure. We would like to overcome the limitations by combining techniques proposed in 3D displays and tomography. Thus, we examine algorithms used for blending on other displays and for reconstruction in discrete tomography (DT) subsequently.

2.1. Tomographic display

Tomographic displays [2] consist of focus-tunable optics (FTO, e.g., focus-tunable lens), a display panel (e.g., LCD or spatial light modulator (SLM)), and fast spatially adjustable backlight (FSAB, e.g., DMD). It uses the temporal multiplexing method with synchronization of FSAB and FTO to construct 3D scenes. Specifically, the focus-tunable lens modulates its focal length in the single cycle (e.g. 60Hz) so that the presentation plane sweeps a certain range of depth. Simultaneously, the DMD illuminates a designated pixel at the appropriate moment, and thus allows the corresponding pixel of the display panel to represent the depth information. Since the RGB image is fixed during the single cycle, focal planes of a tomographic display are correlated with each other. In other words, its binary backlight sequences share the identical RGB image of the display panel.

Figure 1 describes how it depicts a volumetric scene. ‘Binary blending’ technique, which is introduced in Lee et al.’s work, turns on each pixel of FSAB at an appropriate moment with given depth-map information. However, since tomographic displays operate via additive manner, this rendering technique produces artifacts at occlusion boundaries as shown in Fig. 1(a). These

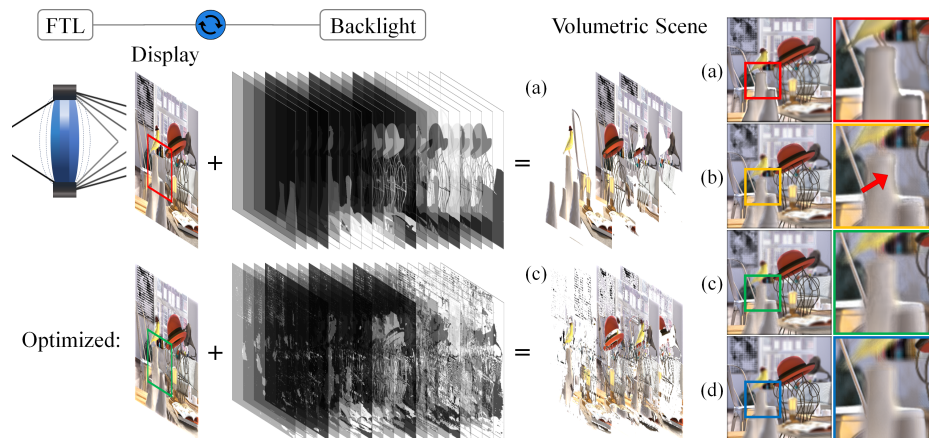


Fig. 1. Principle of tomographic displays. A tomographic display depicts a 3D scene by synchronizing fast spatially adjustable backlight and a focus-tunable lens. (a) When it renders the scene based on its depth information (binary blending), users may notice boundary artifacts since it merges the planes in an additive manner. It can be moderated with a scene decomposition algorithm (b) using SART; but there still remain some defects (*red arrow*), (c) using non-convex optimization algorithm. (d) The ground truth.

artifacts occur because users may notice the misalignment of voxels with depth discontinuity when they shift the viewpoint, and time-multiplexing allows light from a rear plane to pass through front planes as illustrated in Fig. 2. One can conceive of applying optimal blending techniques of multi-plane or light-field displays to our system [6–15], but they are not appropriate since the variables for our problem are restricted by binary constraints of DMD, which makes it a general NP-hard problem.

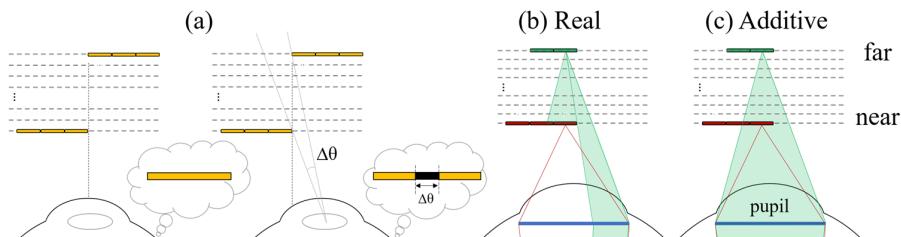


Fig. 2. Illustration of artifacts which occur in tomographic displays. (a) As the viewpoint shifts, the user may notice black artifacts through the gap between voxels. (b) In real environment, the light from a rear plane is blocked by front objects. (c) In additive 3D displays, each plane cannot conceal the light from real planes.

2.2. Related algorithms

Previous techniques on optimization for 3D displays implicitly employ a least squares formulation to recover its volumetric signal from its known linear projections. Since this problem is very similar to that of computed tomography (CT), tomographic perspectives on 3D displays have been introduced with the simultaneous algebraic reconstruction techniques (SART) [27] for the light-field reconstruction [11, 12, 28].

2.2.1. SART-based algorithms

The algebraic reconstruction techniques (ART) are a large family of reconstruction algorithms for solving $\mathbf{y} = \mathbf{Ax}$ in an iterative manner [29]. The SART, which adds longitudinal weighting of the correction terms to the standard ART algorithm, has shown to rapidly converge to a solution among other implementations and has had a major impact in CT applications. It has also been widely used in recovering light-field on 3D displays with multi-layer architectures [6, 9, 11–13].

However, SART shows limited performance when the system is poorly conditioned [6]. In tomography, it corresponds to reconstruction from a small number of projections. To solve this problem, many reconstruction algorithms have been proposed based on the assumption that unknown values are sparse or are made up of only a few gray levels [18–26]. Among the various models, the discrete algebraic reconstruction technique (DART) has been proposed as an effective reconstruction algorithm by alternating iteratively between continuous steps using SART and discretization steps exploiting prior knowledge [20, 21].

2.2.2. The first-order primal-dual algorithm

Since it is not only poorly conditioned but also binary constrained, applying a gradient based algorithm such as SART is not suitable to solve the integer programming problem for DMD. In binary tomography, which is a special case of DT, this problem has been addressed with graphical models, a level-set method, and difference-of-convex (DC) programming [22–26]. The most straightforward idea is to supplement the objective function with a term enforcing binary solutions. To deal with the concave penalty term, a primal-dual sub-gradient algorithm is proposed, decomposing the objective functional [22]. Also, we note that a frequency-domain approach via primal-dual hybrid gradient algorithm (PDHG) is presented for optimizing multi-plane displays [6]. PDHG is a preconditioned version of the first-order primal-dual algorithm which applies to saddle-point problems of the form:

$$\min_x \max_y g(\mathbf{x}) + \langle \mathbf{Kx}, \mathbf{y} \rangle - h^*(\mathbf{x}), \quad (1)$$

where g and h^* are convex and lower-semicontinuous (LSC) functions and the map $\mathbf{K}: \mathcal{X} \rightarrow \mathcal{Y}$ is continuous linear operator. h^* denotes the convex conjugate of a function h . It is shown that Eq. (1) is equivalent to minimizing $g(\mathbf{x}) + h(\mathbf{Kx})$ [30]. This algorithm solves efficiently a large family of convex problems arising in image processing and computer vision. We refer to [30] for this vast literature of saddle-point methods. Furthermore, this scheme can be extended to more general form of minimizing the sum of three convex functions

$$\min_x f(\mathbf{x}) + g(\mathbf{x}) + h(\mathbf{Kx}), \quad (2)$$

where f is convex with β -Lipschitz gradient term. A wide range of problems in image and signal processing, statistics and machine learning can be formulated into this form as many studies have been carried out to solve these problems [30–35].

2.2.3. Proximal alternating linearized minimization (PALM)

While the algorithms introduced above only deal with convex functions, many optimization problems are often non-convex. It has been suggested that primal-dual first-order methods including alternating direction method of multipliers (ADMM) and PDHG can easily be extended to solve non-convex optimization problems [36]. Besides, the proximal alternating linearized minimization (PALM) [37] considers non-convex, non-smooth problems of the form:

$$\min_{\mathbf{x}, \mathbf{y}} \Psi(\mathbf{x}, \mathbf{y}) := F(\mathbf{x}) + G(\mathbf{y}) + H(\mathbf{x}, \mathbf{y}), \quad (3)$$

where H is smooth and the gradient of itself in each block-variable is globally Lipschitz continuous, while F, G are non-smooth, LSC, and possibly non-convex with two-block variables \mathbf{x} and \mathbf{y} . It can handle more than two-block variables with simple functions G_i [38]. This algorithm is based on a proximal regularization of alternating minimizations combined with a forward-backward splitting. For the objective function which has a Kurdyka-Lojasiewicz (KL) property, the convergence to some critical point is shown regardless of the initialization while the critical point depends on initialization [37].

3. Overall optimization strategy

Recall that obtaining optimal representation on tomographic displays is very challenging since it includes the binary optimization problem, which is in general NP-hard. In this section, we thoroughly describe this problem and divide it into sub-problems.

3.1. Notation

In the following sections, we will use the following notational conventions for clarity. We write column vectors in boldface (e.g. \mathbf{x}) and matrices/tensors in upper-case (e.g. \mathbf{K}). Element-wise multiplication (or the Hadamard product) of signals is denoted by $\mathbf{x} \circ \mathbf{y}$. The $|\cdot|$ symbol denotes the ℓ_1 -norm, while $\|\cdot\|$ denotes the ℓ_2 -norm, which for signals are $\|\mathbf{u}\| = (\sum \mathbf{u}_i^2)^{1/2}$ and $|\mathbf{u}| = (\sum |\mathbf{u}_i|)$. $[\cdot]$ denotes the nearest integer function (or rounding) of a real number. $\mathcal{P}_S(\cdot)$ denotes the projection onto a set S . A set of binary vectors of size M is denoted by $\mathbb{B}^M = \{0, 1\}^M$. Also, we define a set $S_{\mathbb{B}}$ and the indicator function $\delta_S(\mathbf{x})$ for a closed set S as follows:

$$S_{\mathbb{B}} := \{\mathbf{Y} \in \mathbb{B}^{MN^2 \times 2} : \mathbf{Y}_{i1} + \mathbf{Y}_{i2} = 1, \text{ for } i = 1, \dots, MN^2\}, \delta_S(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in S \\ \infty & \text{otherwise} \end{cases}. \quad (4)$$

$S_{[0,1]}$ is defined similar to $S_{\mathbb{B}}$. The proximal operator of a function $f : \mathbb{R}^m \rightarrow \mathbb{R}$ is defined as $\text{prox}_{\gamma f}(\mathbf{v}) = \underset{\mathbf{x}}{\text{argmin}} \gamma f(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{v}\|^2$, where $\gamma > 0$ and $\mathbf{v} \in \mathbb{R}^m$.

3.2. Focal plane reconstruction via multi-view based optimization

In this paper, we assume a tomographic display that supports M focal planes with $N \times N$ resolution. Then, we can define following decision variables:

- $\mathbf{u} \in [0, 1]^{3N^2}$ as the RGB image with $N \times N$ resolution, collapsed into a single vector. It can be considered as 3 gray scale images $\mathbf{u} = [\mathbf{u}_r; \mathbf{u}_g; \mathbf{u}_b]$. For simplicity, we suppose a monochromatic (e.g. red) image \mathbf{u}_r instead of \mathbf{u} in the following sections.
- \mathbf{x} as the binary backlight image vector which is made up of M binary images $[\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_M]$, each of which has the resolution of $N \times N$.
- $\mathbf{z} = [\mathbf{z}_1; \mathbf{z}_2; \dots; \mathbf{z}_M]$ as the unknown optimal focal plane images that are given by the element-wise multiplication of \mathbf{x} and \mathbf{u}_M , where \mathbf{u}_M is a column vector containing M copied of \mathbf{u}_r in row dimensions.

We optimize these parameters with target images from distinct viewpoints, which shows similar performance with that of focal-stack decomposition methods [6–8] when the configured exit-pupil size is equal to the pupil diameter [2, 15]. This approach has advantages in terms of memory and speed [39]. This can be interpreted as a tomographic reconstruction of 3D images from 2D projections [11, 18–21, 29]. Note that our following approach is still valid for retinal optimization using the least squares formulation and can be extended to stereoscopic displays [7–9]. We can derive the inverse problem for multi view-based optimization as in previous studies [2, 9–14] with the cost function:

$$\underset{\mathbf{u}, \mathbf{x}}{\text{minimize}} \|\mathbf{A}\mathbf{z} - \mathbf{t}\mathbf{b}'\|^2, \quad \text{s.t. } \mathbf{z} = \mathbf{u}_M \circ \mathbf{x}, \quad 0 \leq \mathbf{u}_M \leq 1, \quad \mathbf{x} \in \mathbb{B}^{MN^2}. \quad (5)$$

Here, we define:

- $\mathbf{A} \in \mathbb{R}^{k^2 N^2 \times M N^2}$ as the sparse projection matrix from M focal layers to $k \times k$ perspective view images.

- t as a number of times backlight was activated during a single cycle. It implies brightness as well as degrees of freedom of backlight illumination for our optimization scheme.

- $\mathbf{b}' = [\mathbf{b}'_1; \mathbf{b}'_2; \dots; \mathbf{b}'_{k^2}] \in \mathbb{R}^{k^2 N^2}$ as a set of target perspective view images. We assume that each of the view image has the same resolution with a focal plane for simplicity. We will use \mathbf{b} instead of $t\mathbf{b}'$ for simplicity in the following sections.

Since we have to solve Eq. (5) on \mathbf{u}_r and \mathbf{x} , we follow a general alternating least squares (ALS) approach, inspired by previous work [2]. To do so, we use some matrix factorization for the derivation of the least squares problem on each vector \mathbf{u}_r and \mathbf{x} : \mathbf{z} , which is the Hadamard product of \mathbf{u}_M and \mathbf{x} , can be represented by a multiplication of a vector by a matrix as

$$\mathbf{Az} = \mathbf{A}(\mathbf{u}_M \circ \mathbf{x}) = \mathbf{AUx} = \mathbf{K_Ux}, \quad (6)$$

$$= \mathbf{A}(\mathbf{x} \circ \mathbf{u}_M) = \mathbf{AXu}_r = \mathbf{K_Xu}_r. \quad (7)$$

By replacing \mathbf{z} in Eq. (5) with Eqs. (6) and (7), we can bring up sub-problems for \mathbf{u}_r and \mathbf{x} respectively. We update each variable in an alternating manner, by fixing one variable and optimizing for the another. The overall scheme for our system is described in Algorithm 1. See the next section for a more detailed description of optimization methods for the sub-problems.

Algorithm 1: Overall optimization scheme for tomographic displays

```

1 set  $\mathbf{u}$  as the RGB image from the center viewpoint,  $\mathbf{x}$  using binary blending;
2 for  $n = 1, 2, \dots, \dots, \text{maxIter}$  do
3   if  $n \leq \text{maxIter\_backlight}$  then
4     - update backlight with fixed  $\mathbf{u}$ ; using Algorithm 3 (Section 4.2)
5     if  $\text{mod}(n, \text{iter\_round}) == 0$  then
6       - round  $\mathbf{x}$ ; (Section 4.4)
7     end
8   end
9   - update display with fixed  $\mathbf{x}$ ; using Algorithm 2 (Section 4.1)
10 end
```

4. Detailed description of the optimization scheme

In this section, we delve into each step of optimization scheme introduced in Section 3.2. We use the first-order primal-dual algorithm for updating gray-scale display values (Section 4.1) and the PALM for updating binary backlight values (Section 4.2). We also discuss exploiting prior knowledge with PD splitting and a novel rounding scheme.

4.1. Display update via primal-dual algorithm

The optimization problem for the SLM image is very similar to that of light-fields [9–14]. With fixed \mathbf{x} and Eqs. (5) and (7), the optimal presentation of display \mathbf{u}_r can be found by solving the following constrained linear least squares problem:

$$\underset{\mathbf{z}}{\text{minimize}} \|\mathbf{Az} - \mathbf{b}\|^2 = \underset{\mathbf{u}_r}{\text{minimize}} \|\mathbf{AXu}_r - \mathbf{b}\|^2, \quad \text{s.t. } 0 \leq \mathbf{u}_r \leq 1. \quad (8)$$

Our constrained convex problem can be put into Eq. (2) by taking g , h and \mathbf{K} as follows:

$$f = 0, \quad g(\mathbf{u}) = \delta_{[0,1]}(\mathbf{u}), \quad h(\mathbf{v}) = \frac{1}{2} \|\mathbf{v} - \mathbf{b}\|^2, \quad \mathbf{K} = \mathbf{AX}. \quad (9)$$

We observe that this strategy is similar to that of Ref. [6] using PDHG but the function choice for the linear composite term is different which enables us to avoid calculating the inverse operator of the form $(I + \alpha \mathbf{A}^T \mathbf{A})^{-1}$ [33]. In this method, the linear operators are applied explicitly without any inversion. Also, we simply derive the proximal operator of g and h as follows:

$$\text{prox}_{\gamma g}(\mathbf{u}) = \mathcal{P}_{[0,1]}(\mathbf{u}) = \min(\max(0, \mathbf{u}), 1), \quad (10)$$

$$\text{prox}_{\delta h}(\mathbf{u}) = (\mathbf{u} + \delta \mathbf{b}) / (1 + \delta). \quad (11)$$

With the Moreau's identity: $\text{prox}_{\delta h^*}(\mathbf{u}) = \mathbf{u} - \delta \text{prox}_{h/\delta}(\mathbf{u}/\delta)$, we update \mathbf{u} by taking the first step of the primal-dual algorithm without its over-relaxation step, following Algorithm 2.

Algorithm 2: Display update using the primal-dual algorithm

```

1 function Update_display(SLM image  $\mathbf{u}_n$ , backlight image  $\mathbf{x}_n$ )
2 - calculate  $\mathbf{K}$  with  $\mathbf{A}$  and  $\mathbf{x}_n$  using Eqs. (7) and (9);
3 -  $\mathbf{v} \leftarrow \text{prox}_{\delta h^*}(\delta \mathbf{K} \mathbf{u}_n)$ ;
4 -  $\mathbf{u}_{n+1} \leftarrow \text{prox}_{\gamma g}(\mathbf{u}_n - \gamma \mathbf{K}^T \mathbf{v})$ ;
5 return  $\mathbf{u}_{n+1}$ ;

```

4.2. Backlight update via non-convex optimization

Here, we propose a novel optimization scheme for finding optimal binary sequences for DMD, inspired by recent works in discrete tomography [20–22]. We formulate the binary optimization problem as a non-convex minimization problem by adding a binary error term, to have the backlight values converged to binary values during the iteration. To this end, we apply the PALM [37] and the three-operator splitting scheme using the primal-dual algorithm [35].

As in Section 4.1, we can derive the inverse problem of optimizing for the binary backlight value \mathbf{x} with fixed \mathbf{u} and a new pseudo-projection matrix $\mathbf{K} = \mathbf{A}\mathbf{U} \in \mathbb{R}^{k^2 N^2 \times MN^2}$:

$$\underset{\mathbf{x}}{\text{minimize}} \frac{1}{2} \|\mathbf{A}\mathbf{z} - \mathbf{b}\|^2 = \underset{\mathbf{x}}{\text{minimize}} \frac{1}{2} \|\mathbf{K}\mathbf{x} - \mathbf{b}\|^2, \quad \text{s.t. } \mathbf{x} \in \mathbb{B}^{MN^2}. \quad (12)$$

Note that the set \mathbb{B}^{MN^2} is not convex. Consequently, many algorithms from convex optimization that can be adopted on optimizing other displays cannot be used directly for this problem [2, 6–15], yet our previous work uses SART and a tailored rounding technique (Section 4.4). In that case, as expected, there is a dominant degradation in performance because it must be rounded at the end (see Fig. 3). Finding a binary solution for the problem of the large size typically requires relaxation such as linear or semidefinite programming relaxation [17, 40].

To begin with, we conceive a relaxation for Eq. (12) based on the constraints $\mathbf{x} \in [0, 1]^{MN^2}$ with the minimum squared distance between \mathbf{x} and \mathbb{B}^{MN^2}

$$d(\mathbf{x}, \mathbb{B}^{MN^2}) := \min_{\mathbf{w} \in \mathbb{B}^{MN^2}} \|\mathbf{x} - \mathbf{w}\|^2 = \sum_{i=1}^{MN^2} \min_{\mathbf{w}_i \in \mathbb{B}} (\mathbf{x}_i - \mathbf{w}_i)^2. \quad (13)$$

Note that this distance can be regarded as an additional functional for optimization, enforcing binary solution. For the error formulation, instead of finding optimal \mathbf{w} for Eq. (13), we add an auxiliary binary coefficient matrix $\mathbf{Y} \in \mathbb{B}^{MN^2 \times 2}$ considering all the cases of square term $\mathbf{Y}_{ik}^2 (\mathbf{x}_i - \mathbf{w}_i)^2$ for $\mathbf{w}_i = 0, 1$. We only "turn on" \mathbf{Y}_{ik} at the smaller square term in the row so that the sum of each row in \mathbf{Y} is unity. With an optimal $\mathbf{Y} \in \mathcal{S}_{\mathbb{B}}$, we can rewrite Eq. (13) as follows:

$$E_{\text{binary}}(\mathbf{x}, \mathbf{Y}) := d(\mathbf{x}, \mathbb{B}^{MN^2}) = \min_{\mathbf{Y} \in \mathcal{S}_{\mathbb{B}}} \sum_{k=1}^2 \sum_{i=1}^{MN^2} \mathbf{Y}_{ik}^2 (\mathbf{x}_i - \mathbb{B}_k)^2. \quad (14)$$

Here, \mathbb{B}_k , which is another representation of \mathbf{w} , denotes the k th element of \mathbb{B} : $\mathbb{B}_1 = 0, \mathbb{B}_2 = 1$. If we relax \mathbf{Y} from $S_{\mathbb{B}}$ to $S_{[0,1]}$, \mathbf{Y} can be considered as a new optimization variable we can handle in continuous domain, existing on a standard simplex with the implication of the probability that the corresponding \mathbf{x}_i takes the value of unity. It is shown that if and only if \mathbf{x} is an element of \mathbb{B}^{MN^2} , the optimal \mathbf{Y} is an element of $S_{\mathbb{B}}$ [21]. Now we consider $E_{binary}(\mathbf{x}, \mathbf{Y})$ as our binary error formulation. This objective functional ensures smooth coupling of \mathbf{x} and \mathbf{Y} for numerical optimization. We refer to Ref. [21] for a thorough analysis on this error term.

Finally, we combine the coupling term, given by Eq. (14), a constraint on $\mathbf{Y} \in S_{[0,1]}$, and the original data fidelity term, given by Eq. (12). This results in a non-convex optimization problem with respect to (\mathbf{x}, \mathbf{Y}) .

$$E(\mathbf{x}, \mathbf{Y}) := \underbrace{\frac{1}{2} \|\mathbf{K}\mathbf{x} - \mathbf{b}\|^2}_{F(\mathbf{x})} + \underbrace{\delta_{S_{[0,1]}}(\mathbf{Y})}_{G(\mathbf{Y})} + \underbrace{\frac{\alpha}{2} \sum_{k=1}^2 \sum_{i=1}^{MN^2} \mathbf{Y}_{ik}^2 (\mathbf{x}_i - \mathbb{B}_k)^2}_{H(\mathbf{x}, \mathbf{Y})}. \quad (15)$$

With modeling F , G , and H of Eq. (3) as Eq. (15), we minimize this non-convex function using PALM as illustrated in Algorithm 3, and derive the proximal operator of F and G as

$$\text{prox}_{\tau F}(\mathbf{p}) = \underset{\mathbf{x}}{\text{argmin}} \frac{1}{2\tau} \|\mathbf{x} - \mathbf{p}\|^2 + \frac{1}{2} \|\mathbf{A}\mathbf{U}\mathbf{x} - \mathbf{b}\|^2 + \delta_{[0,1]}(\mathbf{x}), \quad (16)$$

$$\text{prox}_{\sigma G}(\mathbf{Q}) = \mathcal{P}_{S_{[0,1]}}(\mathbf{Q}). \quad (17)$$

We refer to Ref. [41] for the efficient evaluation of Eq. (17). However, Eq. (16) cannot be derived as an analytic form without using an inversion. Thus, we numerically compute the proximal operator of F using the three-function splitting scheme [35] (Algorithm 4), with modeling f , g , h of Eq. (2) as

$$f(\mathbf{x}) = \frac{1}{2\tau} \|\mathbf{x} - \mathbf{p}\|^2, \quad g(\mathbf{x}) = \delta_{[0,1]}(\mathbf{x}), \quad h(\mathbf{y}) = \frac{1}{2} \|\mathbf{y} - \mathbf{b}\|^2, \quad \mathbf{K} = \mathbf{A}\mathbf{U}. \quad (18)$$

When $1/\tau = 0$, it reduces to the optimization problem in Section 4.1.

Algorithm 3: Backlight update using PALM

- 1 **function** Update_backlight(backlight image \mathbf{x}_n , dual variable \mathbf{Y}_n , SLM image \mathbf{u}_n)
 - 2 - calculate \mathbf{K} with \mathbf{A} and \mathbf{u}_n using Eqs. (6) and (12);
 - 3 - $\tau_n \leftarrow 1/(\alpha_n \max_i \sum_{j=1}^2 \mathbf{Y}_{ij})$;
 - 4 - $\mathbf{x}_{n+1} \leftarrow \text{prox}_{\tau_n F}(\mathbf{x}_n - \tau_n \nabla_{\mathbf{x}} H(\mathbf{x}_n, \mathbf{Y}_n))$;
 - 5 - $\sigma_n \leftarrow 1/(\alpha_n \max_{i,k} (\mathbf{x}_i - \mathbb{B}_k)^2)$;
 - 6 - $\mathbf{Y}_{n+1} \leftarrow \text{prox}_{\sigma_n G}(\mathbf{Y}_n - \sigma_n \nabla_{\mathbf{Y}} H(\mathbf{x}_{n+1}, \mathbf{Y}_n))$;
 - 7 **return** $\mathbf{x}_{n+1}, \mathbf{Y}_{n+1}$;
-

Algorithm 4: Proximal operator using primal-dual three-operator splitting [35]

- 1 **function** $\text{prox}_{\tau F}(\mathbf{p})$
 - 2 start with any $(\mathbf{x}_0, \mathbf{y}_0)$;
 - 3 **while** stop criterion is not met (Section 5) **do**
 - 4 $\mathbf{x}_{n+\frac{1}{3}} \leftarrow \mathbf{x}_n - \gamma \nabla f(\mathbf{x}_n)$;
 - 5 $\mathbf{y}_{n+1} \leftarrow \text{prox}_{\delta h^*}(\mathbf{y}_n + \delta \mathbf{K}\mathbf{x}_n)$;
 - 6 $\mathbf{x}_{n+\frac{2}{3}} \leftarrow \text{prox}_{\gamma g}(\mathbf{x}_{n+\frac{1}{3}} - \gamma \mathbf{K}^T \mathbf{y}_{n+1})$;
 - 7 $\mathbf{x}_{n+1} \leftarrow 2\mathbf{x}_{n+\frac{2}{3}} + \gamma \nabla f(\mathbf{x}_{n+\frac{2}{3}}) - \gamma \nabla f(\mathbf{x}_{n+\frac{1}{3}})$;
 - 8 **end**
 - 9 **return** \mathbf{x}_{n+1} ;
-

4.3. Total variation regularization via primal-dual splitting

Since the number of decision variables comes from tens of focal planes, the backlight optimization problem is ill-posed and solving it possibly yields multiple solutions. In this sense, it is reasonable and even desirable to constrain the solutions through regularization with the help of prior knowledge, from the perspective of compressive sensing [42]. We use the total variation (TV) regularization (ℓ_1 -norm of the 2D gradients) [31], which has been widely used in image processing and reconstruction. We derive the gradient operator ∇ in the discrete setting for tomographic displays, combining 2D differential operators for each layer. It is not burdensome to implement this regularization because only $\text{prox}_{\gamma F}(\mathbf{p})$ is changed when adding the TV functional:

$$\text{prox}_{\tau F}(\mathbf{p}) = \underset{\mathbf{x}}{\text{argmin}} \frac{1}{2\tau} \|\mathbf{x} - \mathbf{p}\|^2 + \frac{1}{2} \|\mathbf{A}\mathbf{U}\mathbf{x} - \mathbf{b}\|^2 + \delta_{[0,1]}(\mathbf{x}) + \lambda |\nabla \mathbf{x}|_1. \quad (19)$$

It can be minimized by applying PD-split scheme [30] to Algorithm 4, updating $\mathbf{K} = [\mathbf{A}\mathbf{U}; \nabla]$; $\mathbf{y} = [\mathbf{y}_1; \mathbf{y}_2]$; and $h(\mathbf{y}) = [h_1(\mathbf{y}_1); h_2(\mathbf{y}_2)] = [\frac{1}{2} \|\mathbf{y}_1 - \mathbf{b}\|^2; \lambda |\mathbf{y}_2|_1]$ in Eq. (18). We demonstrated the performance of the regularization in Appendix (Fig. 11).

4.4. Energy-preserving rounding scheme

Of course, our relaxed model in Eq. (15) will not deliver a binary solution while it is promising that it would results in much better performance than previous models which do not consider the binary penalty. The most straightforward solution to this problem is to conduct rounding operation at the end. In this work, we use a slightly more tailored rounding strategy, dubbed energy-preserving rounding (EPR), as we observe that the disparity of total energy at each position between before and after rounding is the main factor that degrades performance. This rounding scheme maintains the total energy of ray from each pixel to the greatest extent possible. A similar idea was used in occlusion blending [2] and we revisit and extend the idea with a simple formula here.

Specifically, we reshape the backlight vector \mathbf{x} to $N \times N \times M$ tensor \mathbf{X} , to get a rounded sum of voxels at the same position $\mathbf{s}_{ij} = [\sum_{m=1}^M \mathbf{X}_{ijm}]$. Then, the EPR function simply become the projection onto a set \mathcal{S} :

$$\text{epr}(\mathbf{X}) = \mathcal{P}_{\mathcal{S}}(\mathbf{X}), \mathcal{S} = \{\mathbf{W} | \mathbf{s}_{ij} = \sum_m \mathbf{W}_{ijm} \text{ and } \mathbf{W}_{ijm} \in \mathbb{B}, \text{ for } 1 \leq i, j \leq N, 1 \leq m \leq M\}. \quad (20)$$

It can be implemented easily by sequentially compensating the energy after rounding off. Note that this concept can be extended to include a focal stack $\mathbf{Z} = \mathbf{C} \circ \mathbf{X}$, where \mathbf{C} denotes an RGB image tensor. The coefficient \mathbf{C} can be omitted in the tomographic display setup since the RGB image is identical throughout the focal stack.

5. Results

In this section, we evaluate and compare our method with previous approaches. We report metrics such as the peak signal-to-noise ratio (PSNR), HDR-VDP-2 [43] as well as structural similarity index (SSIM) [44] for both perceptual and quantitative comparisons. Next, we validate the performance with our prototype display.

Our simulation was configured as an 80-layer focal stack within 5.5D and 0.0D to fit in with the prototype of [2]. We use our optimization framework to obtain optimal pupil-view images for four scenes with different specifications, shown in Fig. 4. For target images from 7×7 viewpoints within a $6\text{mm} \times 6\text{mm}$ exit-pupil, we rendered scenes with various depth-of-fields, field of views, and resolutions using Blender. It is reported in Table 1.

We implemented our non-convex solver based on Section 4. Binary blending and occlusion blending are implemented as described in [2] and its Supplementary Information. We are also

Table 1. Specification of rendered scenes

Scene	Field of View (°)	Resolution (pixels)	The range in which the object lies
A	10°× 10°	400 × 400	1.0D ~ 3.4D
B	20°× 20°	350 × 350	0.0D ~ 4.0D
C	30°× 30°	400 × 400	0.0D ~ 5.5D
D	30°× 30°	400 × 400	0.0D ~ 4.0D

inspired to address the artifact issue at the boundary by taking the primary idea of DART [20], reducing the number of variables by fixing non-boundary pixels. We implemented a DART-based solver and described its details in the Appendix. All codes are implemented in MATLAB and are run on a 3.5 GHz E5-1620 64-bit Intel Xeon CPU with 128GB of RAM.

In summary, all solvers can be represented as Algorithm 1. For both occlusion blending and DART-based blending, we used SART for *Update_display* and algorithms in the Appendix for *Update_backlight* step. We employed $max_iter_backlight = 90$, $max_iter = 100$, $iter_round = 30$. For the PALM-based blending, they were set to 45, 50, and 15, respectively. The non-convex solver also uses 200 iterations for the proximal splitting in each of the PALM iterations. In addition, we exploit the primal-dual residuals [45] for the stopping criterion of evaluating $prox_{\gamma F}(\mathbf{p})$ in Algorithm 4. The primal-dual algorithm is stopped when the normalized sum of the residual is less than ϵ . As calculating the residual is computationally taxing, we checked it every 10th iteration.

In this paper, we used a single parameter setting to demonstrate our performances for the four scenes: t , M , ϵ , and λ are 24, 80, 10^{-5} , and 3, respectively. α is incremented from 0.1 to 80 quadratically during 45 iterations. γ and δ are chosen based on Eq. (21). While we have found this single parameter setting to be sufficient to obtain superior performances for the four scenes with various scene geometry, the optimal values for the parameters would be affected by many factors including the number of layers, object geometry, texture, etc. Further research is needed to investigate the optimal values for these parameters. In Section 6, we have some discussions about the choice of these parameters.

5.1. Simulation results

Table 2 summarizes the quantitative evaluations of the approaches. We measured PSNR and SSIM for reconstructed focal stacks on four scenes. Our method outperforms previous methods both in perceptual and quantitative figures. However, the increase in quality comes at a cost of much greater computation: Since the non-convex method performs hundreds of iterations of primal-dual algorithms as a sub-routine for evaluating $prox_{\gamma F}(\mathbf{p})$, it takes nearly ten times longer than previous methods. We report the runtime of each method in Table 2.

Table 2. Average PSNR (dB) and SSIM of different optimization methods for tomographic displays on four scenes. We initialized values based on the RGB-D image from the center viewpoint.

	Binary Blending	Occlusion Blending	DART-based	PALM-based
PSNR (dB)	27.4534	34.0980	33.1165	36.1777
SSIM	0.9012	0.9593	0.9552	0.9705
Runtime (s)	< 1	8.14×10^3	1.06×10^4	9.12×10^4

Yet, note that the existing method has a fundamental problem of not converging into binary values as shown in Fig. 3. This hinders them from reaching the same performance of our method. Moreover, the problem of optimizing the backlight is ill-posed because the dimension of decision variables is too large. Thus, when using the gradient-based methods, it seems to be confined in

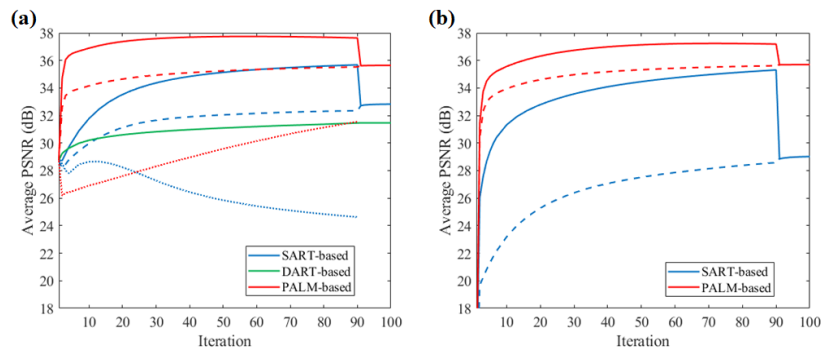


Fig. 3. Comparison on convergence rate with various blending techniques. Averaged PSNRs over 80 focal planes between 0.0D to 5.5D on *scene A* are evaluated after each iteration. We plot two graphs according to the initial condition based on (a) the RGB-D image and (b) uniform gray-level of 0.5. The continuous backlight values, which are represented as *solid* lines, are rounded using EPR after the 90th iteration. The *dashed* lines indicate PSNRs rendered with backlights after EPR at each iteration whereas the *dotted* lines represent roughly rounded ones.

the initial condition (see Figs. 12 and 13 in Appendix) [2]. As a result, our method can use at least 20 times less iteration than the previous method to achieve a similar performance.

Figure 3 demonstrates the convergence and robustness of our method. We rendered a focal stack with an intermediate display image and binary sequence after each iteration step and evaluate its average PSNRs on scene A. Note that the intermediate rounding operations are omitted here. SART-based algorithms do not lead backlight values to an integer, thus the effect of rounding operation is dominant in their performance. More importantly, they work poorly when initial conditions are roughly given as shown in Fig. 3(b). With our proposed method, the penalty caused by rounding at the 90th iteration is significantly decreased with any initial condition, leading to higher performance.

As illustrated in Fig. 4, we use the visual metric HDR-VDP-2 to compare the perceptual performance. We report the results of our methods and of three other blending methods for four scenes. For all scenes, our non-convex method gives better results than other methods, reducing boundary artifacts, and achieving higher metrics.

Figure 5 shows retinal images and insets of scene A. Our method recovers the imagery with the highest accuracy, and other methods tend to generate blurrier images as well as incorrect boundaries.

5.2. Experimental results

Figure 6 shows experimental results photographed from our prototype display used in Ref. [2] and corresponding simulation results. In addition to our simulations, it highlights the performance of our proposed method. As shown in the insets, the PALM-based method recovers the scene close to the ground truth, whereas other methods tend to generate artifacts, such as a glare around boundaries. This is because the workhorse of SART-based optimization methods is that for SLM image, not for backlights. Those gradient-based methods have limited ability to optimize backlight of 80 planes, thus it forces SLM image to have white artificial bands at boundaries to compensate the boundary artifacts (see also Figs. 12 and 13 in Appendix).

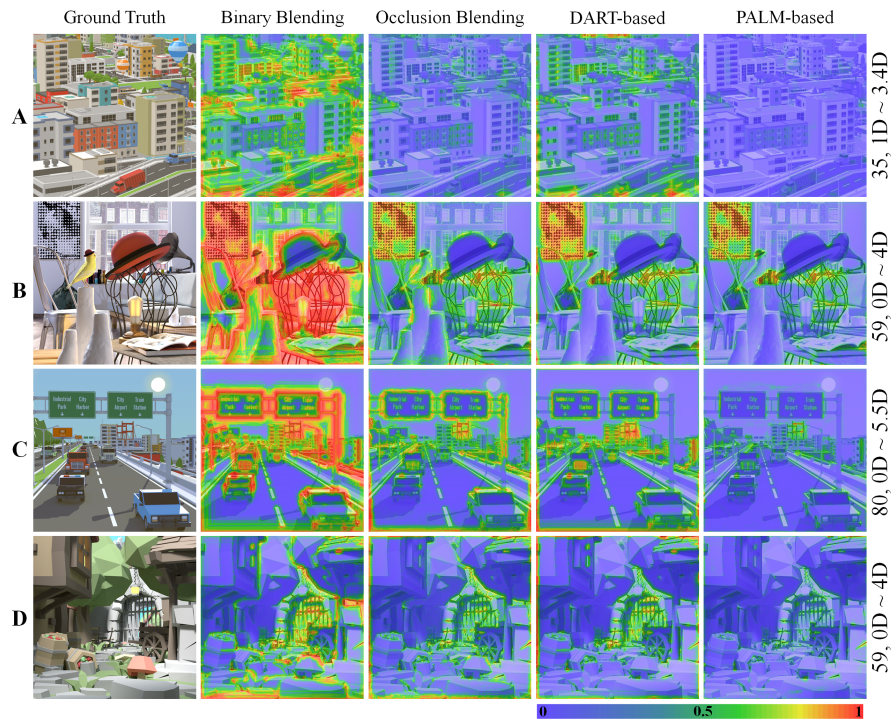


Fig. 4. Comparison of the HDR-VDP-2 metric for different blending methods applied to various scenes. We average the metric over the depth range on the right side. The colormaps correspond to the probability of detection of differences between reconstructed and original ground truth focal stacks. The boundary artifacts are significantly decreased in PALM-based method. The numbers on the right side denote the number and range of focal slices used for each scene, according to Table 1. (Source image courtesy: "Interior Scene" and "SimplePoly Urban", www.cgtrader.com)

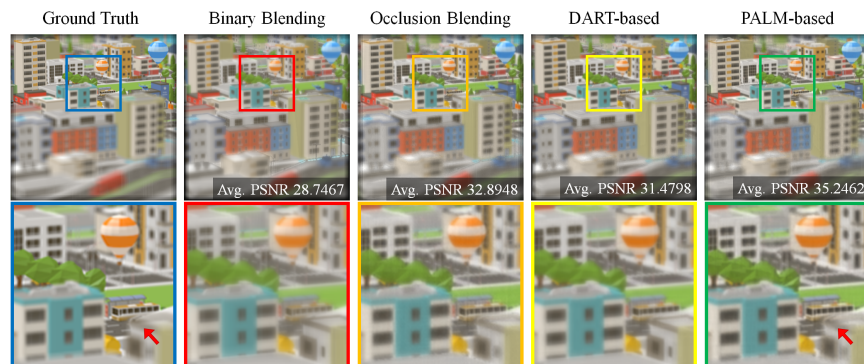


Fig. 5. A reconstructed focal plane of *scene A* at 1.81D (*top*) and insets (*bottom*). The average PSNR is calculated from 80 synthesized retinal images within 5.5D and 0.0D. Insets highlight that our PALM-based algorithm recovers the 3D scene with the highest accuracy and correct occlusion boundaries (*red arrows*).

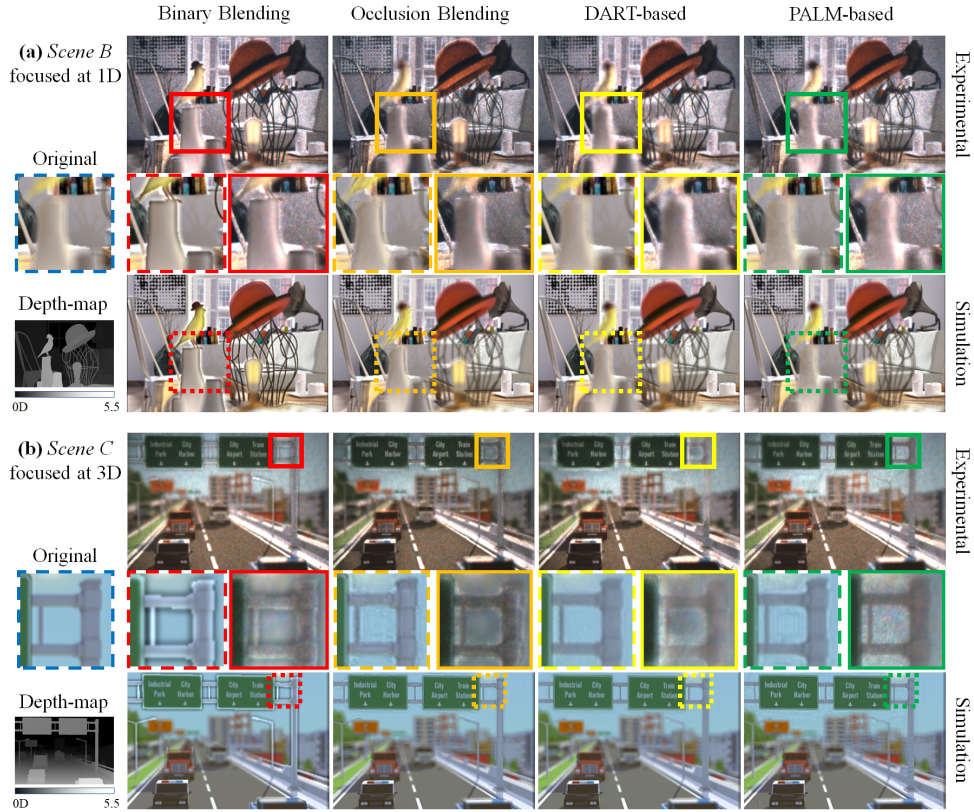


Fig. 6. Comparison of experimental and simulation results from our prototype display with various blending methods on (a) *scene B* and (b) *scene C*. We captured the scenes with a CCD camera, of which focal distances are denoted at the top left.

6. Discussion

6.1. Parameters

One of the notable points about the PD algorithm is that it can be accelerated under certain circumstances and parameters [32]. Basically, we can accelerate it by updating the parameters every iteration, yet it needs to calculate the norm of the linear operator precisely and efficiently. Alternatively, we use the diagonal preconditioning technique to avoid slow convergence speed [46], by setting γ and δ as follows:

$$\gamma = [\gamma_1, \dots, \gamma_{MN^2}]^T, \delta = [\delta_1, \dots, \delta_{k^2N^2}]^T, \text{ for } \gamma_j = \left(\sum_{i=1}^{k^2N^2} \mathbf{K}_{ij} \right)^{-1}, \delta_i = \left(\sum_{j=1}^{MN^2} \mathbf{K}_{ij} \right)^{-1}. \quad (21)$$

We use these parameters in a vector with element-wise multiplication. Note that this kind of step-size setting is similar to the tactic used in SART [11].

Determining the coefficient of the coupling term $H(\mathbf{x}, \mathbf{Y})$ in Eq. (15) plays a significant role in yielding optimal results for our method since it determines a step-size as well as a tendency of the result towards binary values. When α is sufficiently small, our method would converge into an optimal solution quickly with enough step-size, yet it would not lead them to binary values. On the other hand, with large constant α , it is at risk of being stuck in local minimum while

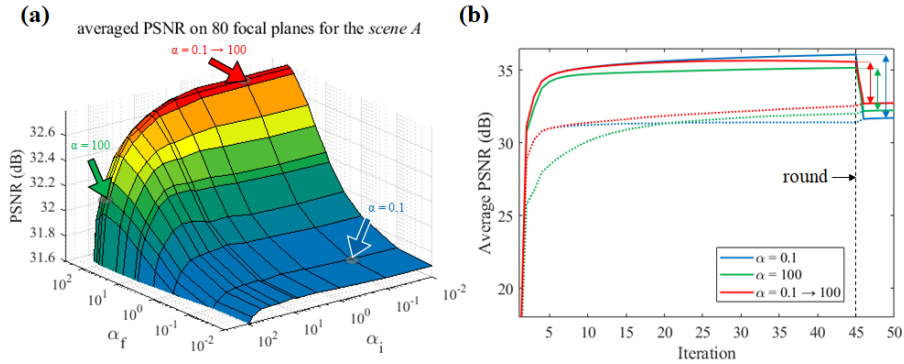


Fig. 7. Optimizing the coefficient of $H(\mathbf{x}, \mathbf{Y})$. (a) Averaged PSNR over 80 focal planes between 0.0D and 5.5D is evaluated, as a function of initial and final values of α (α_i and α_f , respectively), for *scene A*. (b) For the points that arrows are pointing, we plotted a convergence graph. The *dotted* lines denote the metrics of focal planes rendered with rounded binary backlight each iteration using EPR, thus can be interpreted as the effective PSNR during optimization. Here, we used resized images of 128×128 resolution with a simple GPU-based implementation and without intermediate rounding operations.

more enforcing binary values. Thus, it is intuitive to gradually increase α during the iteration, intending to make the coupling term active through an increasing sequence of α .

In Fig. 7(a), we report the performance of our method as a function of the initial value and the final value of α on *scene A*. As expected, increasing α results in the best performance. Also, we note that a similar analysis was performed in binary tomography regarding an upper bound and the varying sequence of the coefficient [22]. While our method differ slightly from theirs, we observe that there is the upper bound around $\alpha = 120$. Figure 7(b) demonstrates that the growing sequence of α is beneficial by taking each advantage at the beginning and the end.

6.2. Dependence of layers and brightness

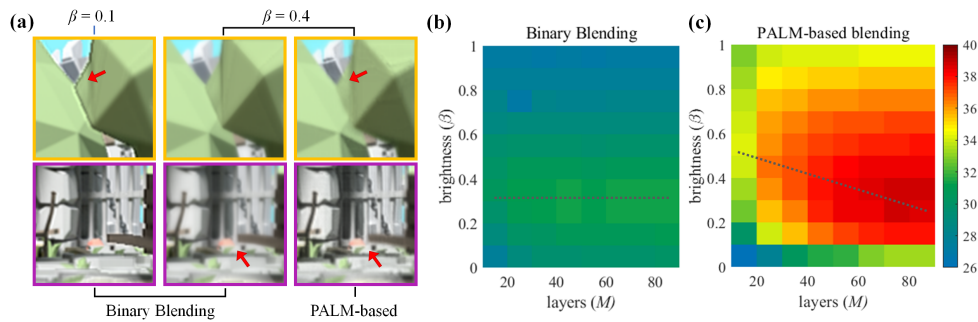


Fig. 8. (a) Insets of *scene D* at 0.35 diopters, using binary blending method and PALM-based blending with $\beta = 0.1, 0.4$ and $M = 80$. In binary blending, by turning on pixels on adjacent planes, the boundary artifacts are mitigated (*orange box*) whereas details are blurred (*purple box*). (b,c) Trade-off relationship in designing tomographic displays with the two methods. averaged PSNR over 80 focal planes is evaluated as a function of the number of layers M and brightness β for *scene D* and the parameters in Section 5.

One of the main benefits of our algorithm is to give a new perspective on the brightness of the

display. We can obtain the brightness scale $\beta \in [0, 1]$ by dividing the illumination time t , which is applied to the target light field, by the number of focal layers M . Tomographic displays show a distinct feature in terms of brightness due to the discontinuous nature in its backlight. While conventional multilayer displays yield higher-quality reconstructions with lower brightness [12], in tomographic displays, the optimal point of design space is shifted to higher brightness because longer illumination gives more leeway to rendering algorithms with respect to the arrangement of binary voxels. When it comes to the binary blending, if we turn on adjacent planes, we can mitigate the artifacts at boundaries, resulting in higher PSNR to some extent, but blurring details as shown in Fig. 8(a).

In Figs. 8(b) and 8(c), we explore the trade space of tomographic displays. To do so, we evaluate average PSNR of the reconstructed focal stack as a function of the number of layers M and brightness β . As can be seen, rendering algorithms alter the design trade space. In general, increasing the number of layers yields higher performance for the same brightness while the optimal brightness of our algorithm gets lower as the number of layers increases (dotted line).

6.3. Depth error analysis

Optimized backlights may float some pixels on wrong depths which can cause depth error. Even though this kind of error can occur deliberately by the algorithm to reduce other artifacts and to improve image fidelity, we demonstrate the depth error is negligible. Human visual system perceives the depth of objects as where the peak contrast of frequency components appears [47]. To find the peak, we cropped patches from a focal stack and collected the signals of 4 to 8 cycles per degree as a function of accommodation distance from Fourier transformed images. The signals are then normalized by the components obtained from the ground truth images. In Fig. 9, we plot the retinal contrast ratio for a range of distances. For the patch with flat depths, our algorithm shows robustness for presenting accurate depth information. Moreover, even in other fragments, which include very far objects and occlusion boundaries, the contrast produced by our algorithm behaves very close to that of the ground truth and thus enables the viewer to drive the correct accommodation response as well.

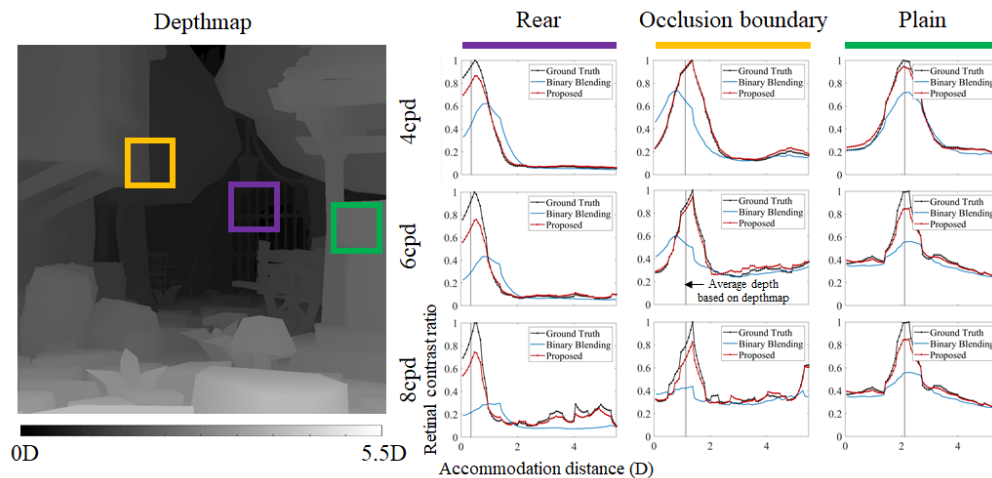


Fig. 9. Retinal contrast as a function of accommodation distance. Normalized contrast ratio is obtained for spatial frequency of 4, 6, and 8 cycles per degree (cpd) with samples collected within 1 cpd interval. The vertical dotted line indicates average depth of the patch based on depth-map.

6.4. Application on other displays: voxel-oriented decomposition

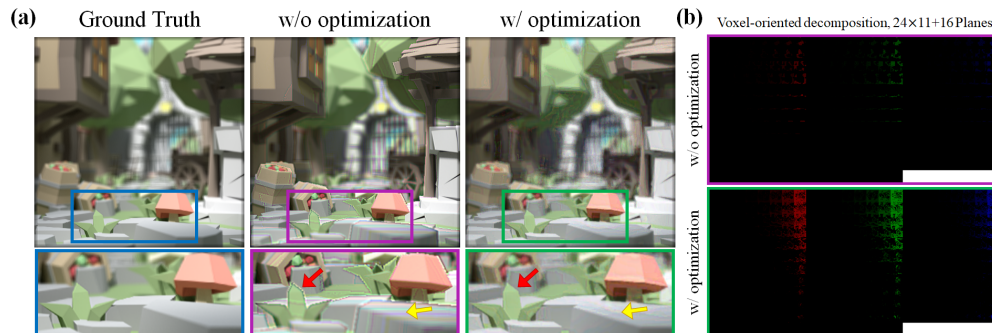


Fig. 10. Application of our non-convex method on the display proposed by Rathinavel et al. [3] for *scene D*. (a) Insets demonstrate that the PALM based method can correct boundary artifacts as well as color distortions. (b) Voxel-oriented decomposed images. We show images for their system without and with optimization. Note that higher brightness ($t = 5$) can be achieved with our method.

Our binary optimization solver can be applied to other systems that harness hardware with binary nature, such as DMD or binary masks. For example, recently, Rathinavel et al. [3] have proposed volumetric near-eye display with an extended depth-of-field using HDR LEDs. They change HDR RGB image sequentially, to support the *direct digital synthesis* (DDS). Their system can be optimized easily with our framework: We manipulate a fixed translation matrix \mathbf{T}_{RGB} with the fixed \mathbf{u} for their DDS decomposition for each RGB color. Then, the optimization problem reduces to the much easier problem of optimizing for binary pattern \mathbf{x} . Namely, its optimization problem is the same as the problem in section 4.2 when we replace \mathbf{U} in Eq. (6) with \mathbf{T}_{RGB} .

To demonstrate our framework, we configured 280 planes between 0.15D and 6.7D and implemented their rendering pipeline as described in Ref. [3]: Color voxels are decomposed to binary voxels based on its depth-map. In Fig. 10, we compared it with our optimized result. For our solver, we run 10 iterations with the extended EPR, and use α incremented from 10^{-4} to 0.5, $t = 5$, and $\lambda = 0.3$. The raw voxel-oriented decomposition also shows boundary artifacts, which can be corrected with our framework as shown in Fig. 10.

7. Conclusion

With the presented work, we take the first steps towards obtaining an optimal binary pattern in tomographic displays. We have introduced a non-convex optimization method with the relevant algorithmic framework including proximal splitting, TV regularization, and a tailored rounding technique. We have demonstrated high-fidelity reconstructions both in simulation and with an experimental prototype display. Without requiring fine-tuned initial conditions for optimization, it also yields stable and versatile solutions for other displays with binary representation. However, we still have a long way to go real-time application for these displays. We hope that our formulation will inspire future approaches such as learning-based algorithms toward real-time applications.

Appendix

Partial gradients: Eq. (22) shows partial gradients of the coupling term $H(\mathbf{x}, \mathbf{Y})$ in the Algorithm 3.

$$(\nabla_{\mathbf{x}} H(\mathbf{x}_n, \mathbf{Y}_n))_i = \alpha \sum_{k=1}^2 \mathbf{Y}_{ik}^2 (\mathbf{x}_i - \mathbb{B}_k), \quad (\nabla_{\mathbf{Y}} H(\mathbf{x}_n, \mathbf{Y}_n))_{ik} = \alpha \mathbf{Y}_{ik} (\mathbf{x}_i - \mathbb{B}_k)^2. \quad (22)$$

Blended binary backlight sequences and RGB image: In Figs. 12 and 13, we show blended RGB image and binary backlight sequences on *scenes B* and *C*, used for the comparison.

Algorithms for updating backlight: Algorithms 5 and 6 describe the backlight update scheme used in occlusion blending and the DART-based solver. For DART, we use $p = 0.5$ and a Gaussian smoothing filter of radius 0.1, while more tailored parameter engineering can be adopted. See Refs. [12, 20, 35] for more details.

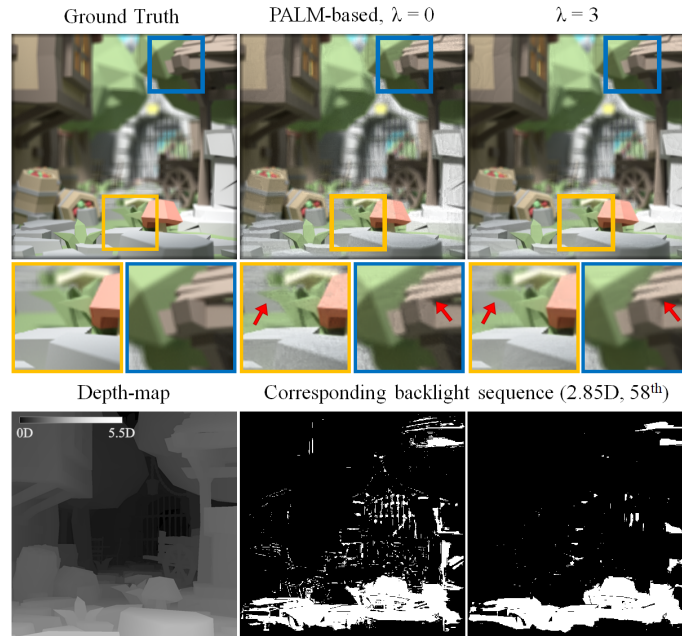


Fig. 11. Rendered focal slices at 2.85D using the PALM-based method with $\lambda = 0$ and 3, on *scene D*. It demonstrates the effect of the TV regularization. By penalizing 2D gradient of backlight image, sparse backlights are obtained (*bottom row*) and the noises are suppressed (*arrows*).

Algorithm 5: Backlight update using SART [12, 27]

- 1 **function** Update_backlight_using_SART($\mathbf{x}_n, \mathbf{u}_n$)
 - 2 - calculate \mathbf{K} with \mathbf{A} and \mathbf{u}_n using Eqs. (6) and (12);
 - 3 - $\gamma_j \leftarrow (\sum_{i=1}^{k^2 N^2} \mathbf{K}_{ij})^{-1}$ for $j = 1, \dots, MN^2$;
 - 4 - $\delta_i \leftarrow (\sum_{j=1}^{MN^2} \mathbf{K}_{ij})^{-1}$ for $i = 1, \dots, k^2 N^2$;
 - 5 - $\mathbf{x}_{n+1} \leftarrow \mathbf{x}_n - \gamma \circ (\mathbf{K}^T (\delta \circ (\mathbf{K}\mathbf{x}_n - \mathbf{b})))$;
 - 6 **return** \mathbf{x}_{n+1} ;
-

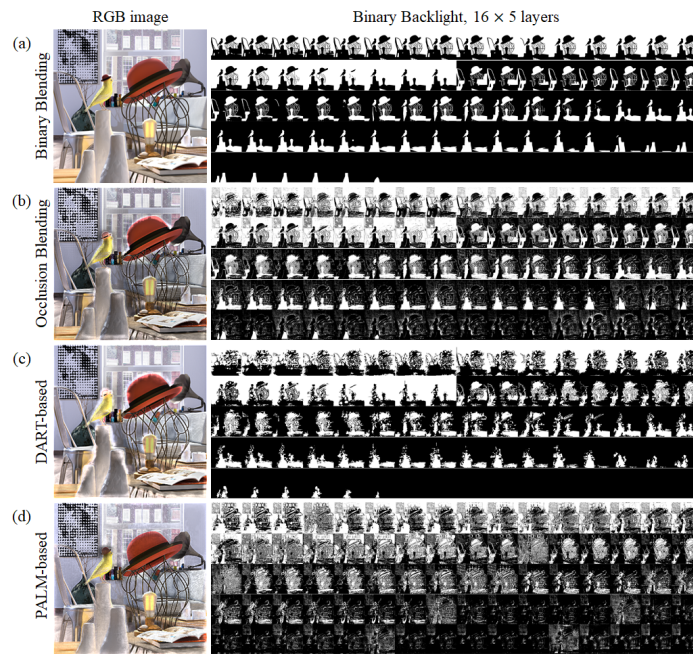


Fig. 12. (a) Binary, (b) occlusion, (c) DART-based, and (d) PALM-based blended RGB image and binary backlight sequences for tomographic displays on *scene B*

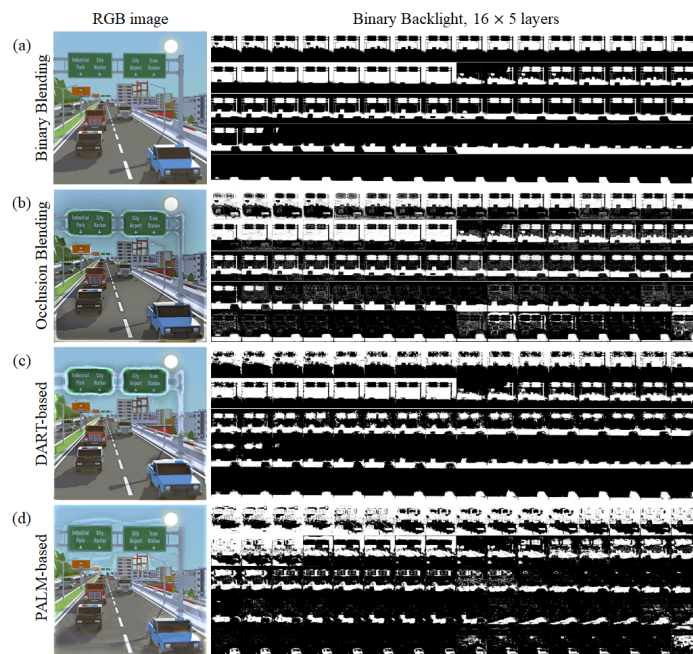


Fig. 13. (a) Binary, (b) occlusion, (c) DART-based, and (d) PALM-based blended RGB image and binary backlight sequences for tomographic displays on *scene C*

Algorithm 6: Backlight update using DART [20]

```

1 function Update_backlight_using_DART( $\mathbf{x}_n, \mathbf{u}_n$ )
2   -  $\mathbf{x}_n \leftarrow$  Update_backlight_using_SART( $\mathbf{x}_n, \mathbf{u}_n$ );
3   - obtain segmented image  $\mathbf{s}_n = \text{EPR}(\mathbf{x}_n)$ ; (Section 4.4)
4   - get free pixels  $\mathbf{f}_n$  to update from  $(p, \mathbf{s}_n)$ ;
5   -  $\mathbf{y}_n \leftarrow \mathbf{f}_n \circ \mathbf{x}_n + (1 - \mathbf{f}_n) \circ \mathbf{s}_n$ ;
6   for  $i = 1, 2, 3$  do
7     | - update  $\mathbf{y}_n$  using SART, while keeping the pixels not in  $\mathbf{f}_n$  fixed;
8   end
9   - apply a smoothing operation to the pixels of  $\mathbf{y}_n$  in  $\mathbf{f}_n$ ;
10  return  $\mathbf{x}_{n+1} = \mathbf{y}_n$ ;

```

Funding

Korea Government (MSIT) (Grant 2017-0-00787)

Acknowledgments

This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korea Government (MSIT No. 2017-0-00787, Development of vision assistant HMD and contents for the legally blind and low vision).

References

1. D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis.* **8**(3), 33–33 (2008).
2. S. Lee, Y. Jo, D. Yoo, J. Cho, D. Lee, and B. Lee, "Tomographic near-eye displays," *Nat. Commun.* **10**, 2497 (2019).
3. K. Rathinavel, H. Wang, A. Blate, and H. Fuchs, "An extended depth-of-field volumetric near-eye augmented reality display," *IEEE Trans. Vis. Comput. Graph.* **24**(11), 2857–2866 (2018).
4. J.-H. R. Chang, B. V. K. V. Kumar, and A. C. Sankaranarayanan, "Towards multifocal displays with dense focal stacks," *ACM Trans. Graph.* **37**(6), 198 (2018).
5. K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," *ACM Trans. Graph.* **23**(3), 804–813 (2004).
6. R. Narain, R. A. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O'Brien, "Optimal presentation of imagery with focus cues on multi-plane displays," *ACM Trans. Graph.* **34**(4), 59 (2015).
7. O. Mercier, Y. Sulai, K. Mackenzie, M. Zannoli, J. Hillis, D. Nowrouzezahrai, and D. Lanman, "Fast gaze-contingent optimal decompositions for multifocal displays," *ACM Trans. Graph.* **36**(6), 237 (2017).
8. L. Xiao, A. Kaplanyan, A. Fix, M. Chapman, and D. Lanman, "Deepfocus: learned image synthesis for computational displays," in *SIGGRAPH Asia 2018 Technical Papers*, (ACM, 2018), p. 200.
9. F.-C. Huang, K. Chen, and G. Wetzstein, "The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues," *ACM Trans. Graph.* **34**(4), 60 (2015).
10. S. Lee, C. Jang, S. Moon, J. Cho, and B. Lee, "Additive light field displays: realization of augmented reality with holographic optical elements," *ACM Trans. Graph.* **35**(4), 60 (2016).
11. G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar, "Layered 3d: Tomographic image synthesis for attenuation-based light field and high dynamic range displays," *ACM Trans. Graph.* **30**(4), 95 (2011).
12. D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar, "Polarization fields: Dynamic light field display using multi-layer lcds," *ACM Trans. Graph.* **30**(6), 186 (2011).
13. G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting," *ACM Trans. Graph.* **31**(4), 80 (2012).
14. D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Trans. Graph.* **32**(6), 220 (2013).
15. C.-K. Lee, S. Moon, S. Lee, D. Yoo, J.-Y. Hong, and B. Lee, "Compact three-dimensional head-mounted display system with savart plate," *Opt. Express* **24**(17), 19531–19544 (2016).
16. D. Kim, S. Lee, S. Moon, J. Cho, Y. Jo, and B. Lee, "Hybrid multi-layer displays providing accommodation cues," *Opt. Express* **26**(13), 17170–17184 (2018).
17. G. Yuan and B. Ghanem, "Binary optimization via mathematical programming with equilibrium constraints," arXiv preprint arXiv:1608.04425 (2016).
18. G. T. Herman and A. Kuba, *Discrete Tomography: Foundations, Algorithms, and Applications* (Springer Science & Business Media, 2012).

19. S. Roux, H. Leclerc, and F. Hild, "Efficient binary tomographic reconstruction," *J. Math. Imaging Vis.* **49**(2), 335–351 (2014).
20. K. J. Batenburg and J. Sijbers, "Dart: a practical reconstruction algorithm for discrete tomography," *IEEE Trans. Image Process.* **20**(9), 2542–2553 (2011).
21. M. Zisler, J. H. Kappes, C. Schnörr, S. Petra, and C. Schnörr, "Non-binary discrete tomography by continuous non-convex optimization," *IEEE Trans. Comput. Imaging* **2**(3), 335–347 (2016).
22. T. Schüle, C. Schnörr, S. Weber, and J. Hornegger, "Discrete tomography by convex–concave regularization and dc programming," *Discret. Appl. Math.* **151**(1-3), 229–243 (2005).
23. T. Lukić and P. Balázs, "Binary tomography reconstruction based on shape orientation," *Pattern Recognit. Lett.* **79**, 18–24 (2016).
24. J. H. Kappes, S. Petra, C. Schnörr, and M. Zisler, "Tomogc: binary tomography by constrained graphcuts," in *German Conf. Pattern Recognit. (Springer, 2015)*, pp. 262–273.
25. K. J. Batenburg, "A network flow algorithm for reconstructing binary images from continuous x-rays," *J. Math. Imaging Vis.* **30**(3), 231–248 (2008).
26. L. Wang, B. Sixou, and F. Peyrin, "Binary tomography reconstructions with stochastic level-set methods," *IEEE Signal Process. Lett.* **22**(7), 920–924 (2014).
27. A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm," *Ultrason. Imaging* **6**(1), 81–94 (1984).
28. M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Trans. Graph.* **25**(3), 924–934 (2006).
29. G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections* (Springer Science & Business Media, 2009).
30. A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numer.* **25**, 161–319 (2016).
31. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D: Nonlinear Phenom.* **60**(1-4), 259–268 (1992).
32. A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.* **40**(1), 120–145 (2011).
33. L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.* **158**(2), 460–479 (2013).
34. D. Davis and W. Yin, "A three-operator splitting scheme and its optimization applications," *Set-valued variational analysis* **25**(4), 829–858 (2017).
35. M. Yan, "A new primal–dual algorithm for minimizing the sum of three functions with a linear operator," *J. Sci. Comput.* **76**(3), 1698–1717 (2018).
36. M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.* **26**(1), 337–364 (2016).
37. J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.* **146**(1-2), 459–494 (2014).
38. T. Pock and S. Sabach, "Inertial proximal alternating linearized minimization (ipalm) for nonconvex and nonsmooth problems," *SIAM J. Imaging Sci.* **9**(4), 1756–1787 (2016).
39. H. Yu, M. Bemana, M. Wernikowski, M. Chwesiuk, O. T. Tursun, G. Singh, K. Myszkowski, R. Mantiuk, H.-P. Seidel, and P. Didyk, "A perception-driven hybrid decomposition for multi-layer accommodative displays," *IEEE Trans. Vis. Comput. Graph.* **25**(5), 1940–1950 (2019).
40. P. Wang, C. Shen, A. van den Hengel, and P. H. Torr, "Large-scale binary quadratic optimization using semidefinite relaxation and applications," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(3), 470–485 (2016).
41. Y. Chen and X. Ye, "Projection onto a simplex," *arXiv preprint arXiv:1101.6081* (2011).
42. Y. Lou, T. Zeng, S. Osher, and J. Xin, "A weighted difference of anisotropic and isotropic total variation model for image processing," *SIAM J. Imaging Sci.* **8**(3), 1798–1823 (2015).
43. R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.* **30**(4), 40 (2011).
44. 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.* **13**(4), 600–612 (2004).
45. T. Goldstein, M. Li, X. Yuan, E. Esser, and R. Baraniuk, "Adaptive primal-dual hybrid gradient methods for saddle-point problems," *arXiv preprint arXiv:1305.0546* (2013).
46. T. Pock and A. Chambolle, "Diagonal preconditioning for first order primal-dual algorithms in convex optimization," in *Proceedings of IEEE Conference on Computer Vision (IEEE, 2011)*, pp. 1762–1769.
47. K. J. MacKenzie, D. M. Hoffman, and S. J. Watt, "Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control," *J. Vis.* **10**(8), 22–22 (2010).