Designing Phase Masks for Under-Display Cameras: supplementary material

Anqi Yang[†], Eunhee Kang[‡], Hyong-Euk Lee[‡], Aswin C. Sankaranarayanan[†]
[‡]Samsung Advanced Institute of Technology Suwon-si, South Korea
[†]Carnegie Mellon University, Pittsburgh, USA

Abstract

Section 1 describes the algorithm for phase mask optimization, and Section 2 provides additional experimental results. In Section 3, we analyze a scenario where a single phase mask is inserted a short distance behind the display, and our analysis shows that single phase masks are again largely ineffective.

1. Algorithm for Phase Mask Optimization

We first explain the physical meaning of each term in the optimization function and then describe the algorithm that solves this optimization problem.

$$\min_{\mathbf{m}} \|\mathbf{S}^\top \mathbf{V}^\top \mathbf{m} - \mathbf{1}\|_2^2 + \alpha \|\mathbf{m}\|_1 \tag{1}$$

$$\text{s.t.} \quad m_i \geq 0, \quad i = 1, \ldots, N. \tag{2}$$

#1 Achromatic. The first term minimizes the ℓ_2 difference between the invertibilities of RGB channels and an all-one vector so that the system performance is equal across RGB channels. Note that $\mathbf{S}^\top \mathbf{V}^\top \in \mathbb{R}^{3\times N}, N \gg 3$. Therefore, the first term is an under-determined system with infinitely many solutions. Without proper regularization, the magnitudes of elements in $\mathbf{m}$ can be unbounded.

#2 More invertible. The second term, ℓ_1 regularization, encourages large invertibilities. The choice of ℓ_1 norm is motivated as follows. Since $\mathbf{m}$ represents counts of microlenses, with proper normalization, it sums to the total number of microlenses in the aperture. We let the normalized counts be $\frac{\mathbf{m}L}{\|\mathbf{m}\|_1}$, and normalized invertibilities be $\mathbf{S}^\top \mathbf{V}^\top \frac{\mathbf{m}L}{\|\mathbf{m}\|_1}$. Since $\mathbf{S}^\top \mathbf{V}^\top \mathbf{m}$ is constrained to be $\mathbf{1}$ by the first term, the normalized invertibilities can be simplified as $\frac{\mathbf{1}L}{\|\mathbf{m}\|_1}$. Therefore, minimizing $\|\mathbf{m}\|_1$ is equivalent to maximizing the invertibilities of RGB channels.

We use the log-barrier approach to approximate the non-negative constraint and convert the original optimization into an unconstrained problem,

$$\min_{\mathbf{m}} \|\mathbf{S}^\top \mathbf{V}^\top \mathbf{m} - \mathbf{1}\|_2^2 + \alpha \|\mathbf{m}\|_1 - \frac{1}{t} \sum_{j=1}^{N} \log(\mathbf{z}_j^\top \mathbf{m}) \tag{3}$$

where $\mathbf{z}_j$ is a one-hot vector that is one at j-th element. We apply the barrier method to solve this problem [1]. The algorithm is summarized in Algorithm 1.

Algorithm 1 Algorithm for phase mask optimization

Input: $\alpha \leftarrow 0.1, \mu \leftarrow 2, \mathbf{m}_{\text{init}} \leftarrow \mathbf{1}, t^{(0)} \leftarrow 0.01$
Output: $\mathbf{m}^*$
$k \leftarrow 0$ ▷ iteration index
$\mathbf{m}^{(0)} \leftarrow Newton(\mathcal{L}, \mathbf{m}_{\text{init}}, t^{(0)})$
while $\mu \cdot t^{(k)} \leq 10^4$ **do**
  $k \leftarrow k + 1$
  $t^{(k)} \leftarrow \mu \cdot t^{(k-1)}$
  $\mathbf{m}^{(k)} \leftarrow Newton(\mathcal{L}, \mathbf{m}^{(k-1)}, t^{(k)})$
end while
$\mathbf{m}^* \leftarrow \mathbf{m}^{(k)}$

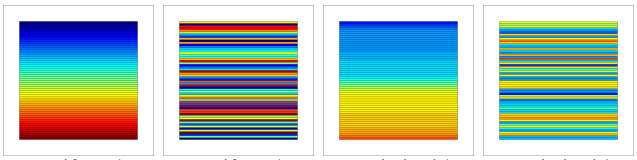
Function $Newton(\mathcal{L}, \mathbf{m}_{\text{init}}, t)$
$\mathbf{m} \leftarrow \mathbf{m}_{\text{init}}$
$\epsilon \leftarrow 1$
while $\epsilon \geq 0.001$ **do**
  $\mathbf{m}_{\text{pre}} \leftarrow \mathbf{m}$
  $\mathbf{G} \leftarrow \nabla \mathcal{L}(\mathbf{m}_{\text{pre}}; t)$ ▷ Gradient
  $\mathbf{H} \leftarrow \nabla^2 \mathcal{L}(\mathbf{m}_{\text{pre}}; t)$ ▷ Hessian
  $\mathbf{m} \leftarrow \mathbf{m}_{\text{pre}} - \mathbf{H}^{-1}\mathbf{G}$
  $\epsilon \leftarrow |\mathcal{L}(\mathbf{m}_{\text{pre}}; t) - \mathcal{L}(\mathbf{m}; t)|$
end while
Return m

During inference, we normalize and round up the optimal $\mathbf{m}$ into $\hat{\mathbf{m}}^* = \lfloor \frac{\mathbf{m}^*}{\|\mathbf{m}^*\|_1} L \rfloor$, such that $\sum_{j=1}^{N} \hat{m}_j^* = L$. The resulting phase mask contains $\hat{m}_j^*$ number of microlenses that have a maximum height of $h_j = \frac{T\lambda_j}{n-1}$.

2. Additional Experiments

Effect of Ordering of d_l. Given a set of folding heights d_ls for each microlens, we show the spatial ordering of d_ls has a

negligible effect on the imaging performance. We compare two types of d_ls — those decided by uniformly sampled λ_0s and those decided by optimization, and for each type, we compare two orderings — sorted d_ls with an ascending order and shuffled d_ls. All phase masks are controlled by $T = 1$. Figure 1 illustrates two types of d_ls together with two types of orderings. Table 1 evaluates PSNR and SSIM of each setup at a fixed light level of around 1,600 photons. Δ_{PSNR} and Δ_{SSIM} compute relative differences between the shuffled and the ordered with respect to the ordered. We can see that for each type of d_l, the ordered and shuffled have similar performance; while optimized d_ls outperform uniform d_ls. Therefore, in this paper, we only optimize for heights d_ls and adopt an ascending order after optimization.



Uniform / ordered | Uniform / shuffled | Optimized / ordered | Optimized / shuffled

Figure 1: **Different ordering of d_l.**

	PSNR	Δ_{PSNR}	SSIM	Δ_{SSIM}
uniform, *ordered*	26.69 dB	–	0.7106	-
uniform, *shuffled*	26.51 dB	−0.68 %	0.7187	1.13 %
optimized, *ordered*	27.44 dB	–	0.7368	-
optimized, *shuffled*	27.18 dB	−0.96 %	0.7396	0.38 %

Table 1: **Effect of ordering of d_l.**

Effect of Phase Masks. Figure 2 shows additional results on the validation set comparing TOLED without and with two sets of proposed phase masks that have a thickness of around 1 µm and 5 µm. For each thickness, we compare three choices of wrapping heights. Results on the validation set are similar to those on the test set, showing that the proposed setups outperform TOLED at all light levels.

Effect of Pixel Density. Figure 3(a) shows additional SSIM plots for UDCs at various pixel densities. The trends are similar to PSNR plots. At 5 µm, optimized phase masks outperform TOLED at all four pixel densities, and at 1 µm ours outperform TOLED with pixel densities larger than 300 DPI. Because microlens arrays for larger pixel pitch have larger radii, and results in phase wrapping artifacts when implemented as thin plates.

Effect of phase mask quantization. Figure 3(b) evaluates phase masks without and with a quantization of a 200 nm step in height using SSIM. Phase masks with quantization perform similarly to ones before quantization.

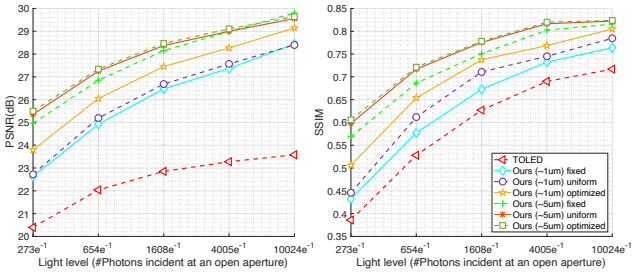
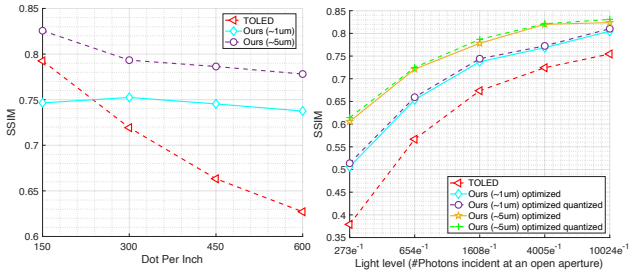


Figure 2: **Comparison of our setups with a traditional UDC with TOLED on validation set.**



(a) Effect of pixel densities (b) Effect of phase mask quantization

Figure 3: **Effect of (a) setups with varying display pixel densities and (b) quantization of phase masks.**

Comparisons with Other OLED Displays. We compare our design with TOLED, POLED [5, 4], and two displays layouts designed specifically for UDCs [3, 2]. Figure 4 shows qualitative results for scenes with an indoor light level of around 650 photons. Our design falls into the category of requiring no change to the display openings. Compared to two other high-quality displays in this category, TOLED and POLED, ours produces significantly fewer artifacts. POLED has an LTR of around 8 % and produces photographs with the most noise. Displays designed specifically for UDCs, including Yang et al. and ZTE, have better performance than ours. However, these modifications also degrade the display quality. For example, the random tiling proposed by Yang et al. produces non-negligible visual artifacts for the display, and the ZTE display trades off pixel densities for larger transparent regions.

Figure 5 shows the performance of all displays for scenes ranging from indoor to outdoor light levels. Note that the performance of POLED increases fast as the light level of the scene increases, however, at a light level of 250 photons, POLED is worse than ours by around 10 dB in PSNR. This is due to its low LTR, which becomes a pronounced issue when capturing photographs of indoor scenes.

SOTA Restoration. We compare the quantitative results from an iterative solver and from a cutting-edge deep neu-

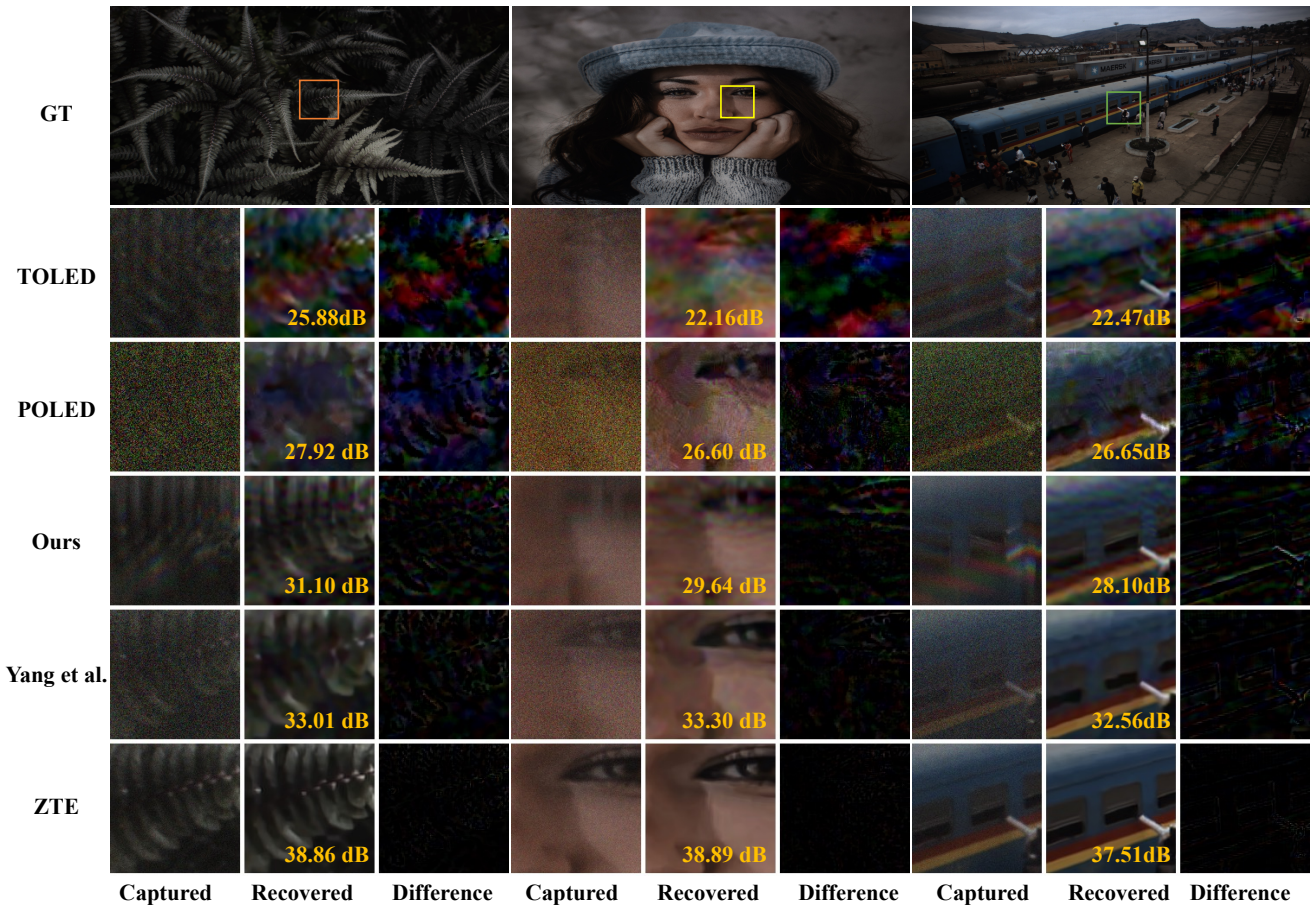


Figure 4: **Qualitative results comparing ours with common display layouts.** All displays are 600 DPI except for the ZTE display. ZTE display is modified to have low pixel densities to accommodate UDCs. All scenes are at an indoor light level.

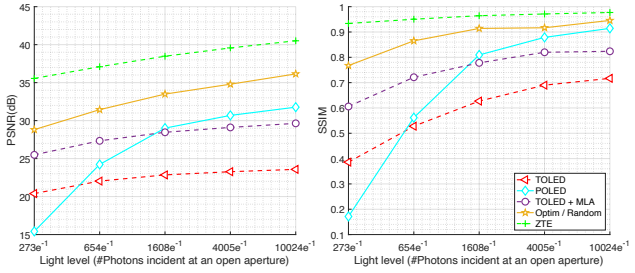


Figure 5: **Comparisons with other OLED displays.**

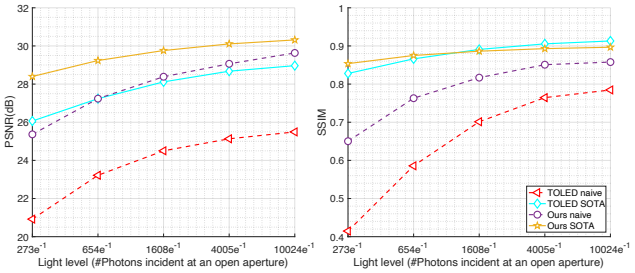


Figure 6: **Deblurring using an iterative solver versus using a SOTA deep neural network.**

ral network-based method for TOLED and our setup. Our setup refers to an optimized phase mask with a 5 µm thickness. First, results from SOTA deep neural network are significantly better than those from the iterative solver. For TOLED, SOTA method outperforms the iterative solver by around 4 dB; and for ours, by around 2 dB. Second, when comparing SOTA restorations for both setups, TOLED performs similarly to ours in SSIM, while worse than ours by around 2 dB in PSNR. This is because, although DISCNet recovers many sharp details for TOLED, it fails in removing widespread ringing artifacts caused by the ill-conditioning of the PSF of TOLED. The resulting visual artifacts are ghosting effects and repetitive copies that are unfaithful to the ground-truth scene, and therefore, ours produce much more visually appealing results than TOLED.

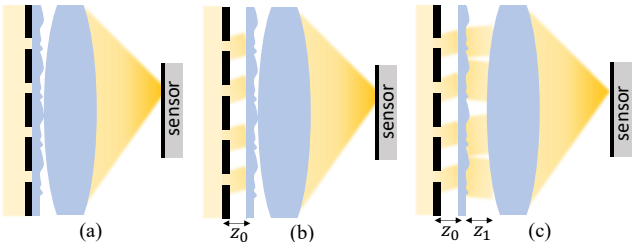


Figure 7: **Three scenarios where a single phase mask is inserted behind the display in UDCs.**

3. Inadequacy of Single Phase Masks: Additional Analysis

In the main paper, we consider the scenario where a single phase mask is inserted tightly against the display and prove its inadequacy in improving the image quality of UDCs, as show in in Figure 7(a). In this section, we consider two additional scenarios shown in Figure 7(b)(c). We move the phase mask away from the display panel by a short distance z_0. Note that after introducing the distance z_0, plane waves that are incident on the display from different directions produce different PSFs. These spatially-varying PSFs break the convolutional imaging model, and thus prevent us from analyzing system invertibility as in Section 3.1 in the main paper, i.e., using the MTF as a tool for analysis. Instead, we solve for a phase mask that minimizing the difference between wavefronts observed in a UDC and a camera with fully open aperture, and examine the resulting analytical solution.

Given a UDC, we define the display as the aperture plane, and a phase mask and the camera lens are at a plane parallel to and z_0 distance away from this aperture plane. We assume wave propagation for z_0 can be well approximated by Fresnel diffraction. Now consider a plane wave incident on the display/aperture at an angle $\theta_i \in [\theta_{\text{min}}, \theta_{\text{max}}]$, where the bounds denote the field of view of a conventional smartphone camera. The wavefront after propagation to the phase mask, i.e., free-space propagation by a distance z_0, is denoted as u_{θ_i}; the effect of the phase mask can be denoted as a pointwise multiplication with a unit-norm phasor, and so the wavefront after the phase mask is denoted as $\phi[m]u_i[m]$, where m is a spatial index. We repeat this for a number of different incident angles $\{\theta_i, i = 1, \ldots, N\}$. Now, consider an ideal alternative, where the display (and its aperture) is not present, and we simply have the main lens z_0 distance away from the aperture plane. This ideal system provides us with a target set of wavefronts, one for each incident angle, that we denote as $T = [t_{\theta_1}, ..., t_{\theta_i}, ..., t_{\theta_N}]$.

Lemma (Inadequacy of a single phase mask behind the display)**.** *Following the setup for a UDC described above, inserting a phase mask a distance away from the display panel can not decrease the Frobenius norm between the set of wavefront in the ideal camera and that in the UDC, $\|T - diag(\phi)U\|_F^2 \geq \|T - U\|_F^2$, where ϕ is the phase and amplitude modulation introduced by the phase mask.*

Proof. We solve the modulation of phase mask such that the Forbenius norm between the modulated wavefront $\text{diag}(\phi)U$ and the target wavefront T is minimized. By taking the derivative of the objective function with respect to phase modulation $\phi[m_k]$ at each location m_k and set the derivative to zero, we obtain

$$\phi[m_k] = \frac{\sum_\theta u_\theta^*[m_k] t_\theta[m_k]}{\sum_\theta u_\theta^*[m_k] u_\theta[m_k]}. \tag{4}$$

We can substitute wavefront incident from direction θ with that from normal direction, $u_\theta[m_k] = e^{j\frac{2\pi}{\lambda}(\theta(m_k - \frac{1}{2}\theta z))} u_0[m_k - \theta z]$ and $t_\theta[m_k] = e^{j\frac{2\pi}{\lambda}(\theta(m_k - \frac{1}{2}\theta z))} t_0[m_k - \theta z]$. The wavefront $t_0(\cdot)$ is propagated from the fully-open aperture, $t_0(x) = \frac{e^{j\lambda z}}{j\lambda z} \int_{-\infty}^{+\infty} 1 \cdot e^{j\frac{k}{2z}(x-\xi)^2} d\xi = c_0$, and is thus a constant. We can simplify the expression for $\phi[m_k]$ as

$$\phi[m_k] = \frac{c_0 \sum_\theta u_0^*[m_k - \theta z]}{\|u_0[m_k - \theta z]\|_2^2}. \tag{5}$$

The numerator of $\phi[m]$ can be recognized as a convolution between $u_0^*[m]$ and a rectangular window running from $\theta_{\text{min}} z$ to $\theta_{\text{max}} z$, and the denominator is a normalization term. The periodic pixel tiling on a display panel produces a peroidic wavefront $u_0^*[m]$ with a period of pixel pitch p. When the rectangular window is significantly larger than the period of $u_o^*[m]$, $(\theta_{\text{max}} - \theta_{\text{min}})z = \tilde{\theta} z + np$, where $n \in \mathbb{N}$ and $\tilde{\theta} z < p$,

$$\sum_{\theta=\theta_{\text{min}}}^{\theta_{\text{max}}} u_0^*[m_k - \theta z] = \sum_{\theta=\theta_{\text{min}}}^{\theta_{\text{min}}+\tilde{\theta}} u_0^*[m_k - \theta z] + nc_p, \tag{6}$$

where $c_p = \sum_{m=0}^{p} u_0^*[m]$ is the summation of u_0^* over one period. Since $nc_p \gg \sum_{\theta=\theta_{\text{min}}}^{\theta_{\text{min}}+\tilde{\theta}} u_0^*[m_k - \theta z]$, (6) is dominated by a constant term. Thus, $\phi[m]$ is approximately a constant function.

Further, let us introduce a distance z_1 between the phase mask and lens, as is shown in scenario (c) in Figure 7. It is easy to see that varying distance z_1 has no effect on the PSFs, and thus the same conclusion holds. ■

Implication. The optimal phase mask that can be inserted a short distance away from the display panel is approximately a constant. Any other phase masks can only deviate wavefronts from that of an ideal camera, making the PSFs formed on the sensor less desired.

References

[1] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[2] Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Chen Change Loy, and Jinwei Gu. Removing diffraction image artifacts in under-display camera via dynamic skip connection network. In *CVPR*, 2021.

[3] Anqi Yang and Aswin C Sankaranarayanan. Designing display pixel layouts for under-panel cameras. *IEEE TPAMI*, 43(7):2245–2256, 2021.

[4] Yuqian Zhou, Michael Kwan, Kyle Tolentino, Neil Emerton, Sehoon Lim, Tim Large, Lijiang Fu, Zhihong Pan, Baopu Li, Qirui Yang, et al. Udc 2020 challenge on image restoration of under-display camera: Methods and results. In *ECCV*, 2020.

[5] Yuqian Zhou, David Ren, Neil Emerton, Sehoon Lim, and Timothy Large. Image restoration for under-display camera. In *CVPR*, 2021.