Supplemental Material for Spatially-Varying Autofocus

Yingsi Qin, Aswin C. Sankaranarayanan, Matthew O'Toole

Carnegie Mellon University

A. Basics of Spatially-Selective Lensing

Building on the work of Qin et al. [28], we briefly discuss how to get a pixel to focus on a specific depth. Fig. 13 provides a visual illustration of the relevant variables in this section.

Suppose that we want to get the pixel at location (x_0, y_0) to focus at a depth z_{des} (with the depth measured from the main/imaging lens). This would mean that we need to bring the pixel to a distance u_{des} behind the main lens, as given by the thin-lens equation:

$$\frac{1}{u_{\text{des}}} + \frac{1}{z_{\text{des}}} = \frac{1}{f_{\text{main}}}. \tag{2}$$

When the SLM does not modulate the light, the image sensor is focused to the nominal image plane, denoted by P_5, which is a distance u_{nom} behind the main lens. Hence, to get the pixel (x_0, y_0) to focus at the depth z_{des}, we would need to axially shift its image plane by a distance $u_{\text{des}} - u_{\text{nom}}$.

We can get this axial shift by placing a (focus tunable) lens of focal length that satisfies

$$u_{\text{des}} - u_{\text{nom}} = \frac{f_0^2}{f_{\text{Loh}}}. \tag{3}$$

Since we realize this focus tunable lens with a Lohmann lens of curvature parameter κ, as defined in Equation 1 of the main paper, and refractive index η, this focal length is realized with a shift Δ_{Loh} that satisfies

$$f_{\text{Loh}} = \frac{1}{6\kappa\Delta_{\text{Loh}}(\eta - 1)}, \tag{4}$$

Finally, this shift between the cubic plates is realized in the Split-Lohmann configuration (as described in Fig. 2 of the main paper) using a phase ramp of slope v_0 such that

$$\Delta_{\text{Loh}} = v_0 f_0. \tag{5}$$

Putting it all together, we can calculate the tilt v_0 at the SLM location $(-x_0, -y_0)$ as

$$v_0(-x_0, -y_0) = \frac{1}{6\kappa f_0^3(\eta - 1)} \left(\frac{z_{\text{des}} f_{\text{main}}}{z_{\text{des}} - f_{\text{main}}} - u_{\text{nom}} \right). \tag{6}$$

Hence, given a target depth map $z_{\text{des}}(x, y)$, we can convert it to a tilt map for the SLM, which provides us with the phase pattern to display.

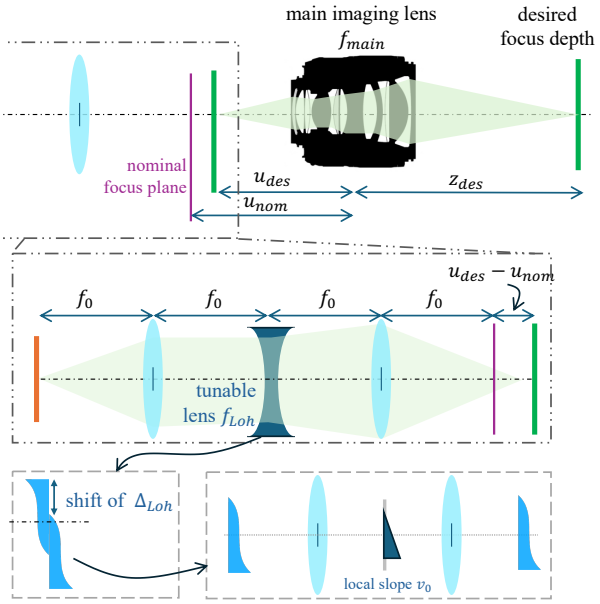


Figure 13. Guide for Eq. (2) through Eq. (6).

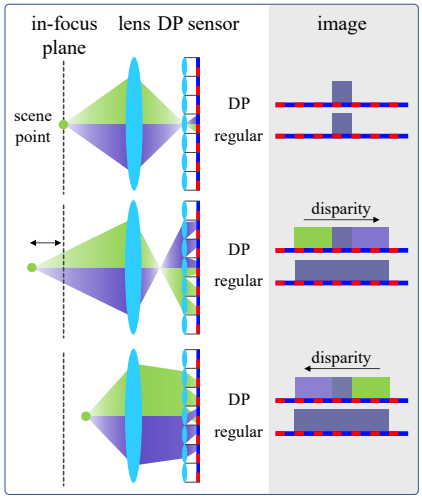


Figure 14. Dual-Pixel (DP) image formation.

Dual-Pixel (DP) image formation. Our PDAF algorithm makes use of a DP sensor. A DP sensor has two photodiodes

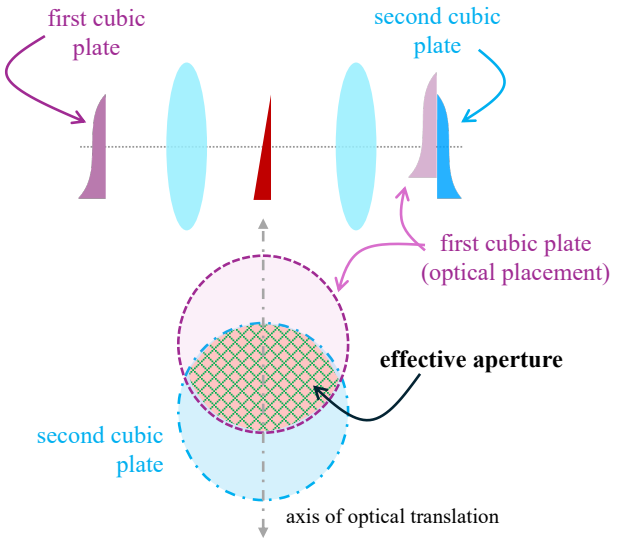


Figure 15. Shifts between the cubic plates in the Split-Lohmann system leads to different aperture shapes, which introduces the diffracted-limited resolution for the all-in-focus image.

under the microlens for each pixel, producing a stereo pair of DP images. The DP image pair provides signed disparity cues as a function of the amount of defocus, and therefore, the magnitude and direction of lens focus change needed to bring the entire field of view into all-in-focus.

B. Detailed explanation of loss of light

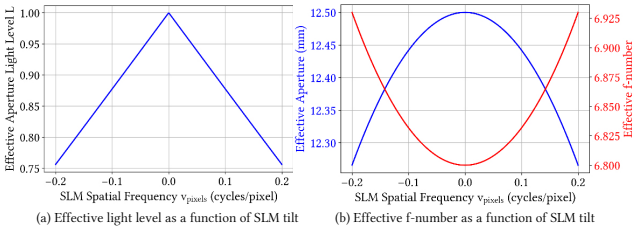
Applying a phase ramp on the SLM results in a translation of the aperture, as shown in Fig. 15. Note that we orient the cubic phase plate so that the axis of optical translation is vertical and orthogonal to the dual-pixel axis; this ensures that the DP blur kernels are decoupled from the depth-dependent ambiguity introduced by cubic phase plate translation. We plot the effective light level, aperture, and f-number as a function of the ramp's slope in Fig. 16. At an SLM tilt, the effective aperture becomes an intersection between two circles and characterized by the chord length:

$$C = 2\sqrt{r^2 - (\Delta y_{cpp}/2)^2}\,, \tag{7}$$

where r is the radius of the aperture, and Δy_{cpp} is the translation of the cubic phase plate as a function of SLM spatial frequency v_{pixels} (cycles per period):

$$\Delta y_{cpp} = \frac{\lambda f_0}{\delta_{SLM}} v_{pixels}\,. \tag{8}$$

δ_{SLM} represents the SLM's pixel pitch, and λ represents the wavelength of light. The area of the effective aperture is the



(a) Effective light level as a function of SLM tilt (b) Effective f-number as a function of SLM tilt

Figure 16. Effective f-number, aperture, and light level as a function of SLM spatial frequency.

Component	Manufacturer	Part Number
Camera sensor	Canon	EOS R10
SLM	Holoeye	GAEA-2
Cubic Phase Plate	Power Photonic	Custom
Relay Lenses	Samyang	SY85MAE-N
Objective Lens	Nikon	40mm f/2.8G
Aperture	Thorlabs	SM1D12C
Beam Splitter	Thorlabs	CCM1-BS013
Linear Polarizer	Thorlabs	LPVISE2X2

Table 2. Component list for our prototype.

following:

$$A = 2r^2 \cos^{-1}\left(\frac{\Delta y_{cpp}}{2r}\right) - \frac{\Delta y_{cpp}}{2}\sqrt{4r^2 - \Delta y_{cpp}^2}. \tag{9}$$

In our setting, at the most extreme focus settings (v_{pixels} = 0.2), the area drops to 76% of the maximum, indicating a loss of 24% of the light; see Fig. 16.

C. Additional system details

Hardware. Since the phase SLM is reflective, we folded the optical path, reusing one lens and the cubic phase plate. Our fabrication of the cubic plate was done with subtractive manufacturing using laser etching. The maximum thickness of the phase plate was constrained to be $50\,\mu m$, which led us to choose a curvature parameter of $\kappa = 112$; hence, the resulting phase plate had a height $112\,(x^3+y^3)$ over a circle of diameter $12.5\,mm$ on a material with refractive index of $\eta = 1.46$. With the $f_0 = 85\,mm$ relays, this led to an effective system aperture of $f/6.8$. We detail the hardware components used for our prototype in Tab. 2.

The maximal angular tilt of the SLM is $\lambda/(2\delta_{SLM}) \approx \pm 4^\circ$ for a Holoeye GAEA-2 with pixel pitch $\delta_{SLM} = 3.74\,\mu m$ and light having wavelength $\lambda = 532\,nm$; we only used 40% of this range to reduce the effect of the phase warping artifacts (*i.e.*, $-0.2 \leq v_{\text{pixels}} \leq 0.2$). In particular, the minimum number of pixels we use for the period of the phase ramp is 5. With these parameters, the maximum

optical shift between the cubic plates we could obtain was

$$\Delta_{\text{Loh}} = \pm \frac{\lambda}{2 \times 5\delta_{\text{SLM}}} f_0 \approx \pm 1.21\,\text{mm}.$$

The resulting Lohmann lens can produce focus tunability in the range of

$$f_{\text{Loh}} = 1/(6\kappa\Delta_{\text{Loh}}(\eta - 1)) \approx \pm 2.674\,\text{m}.$$

Finally, this focus tunability results in maximal axial shift of the image plane of the main lens by

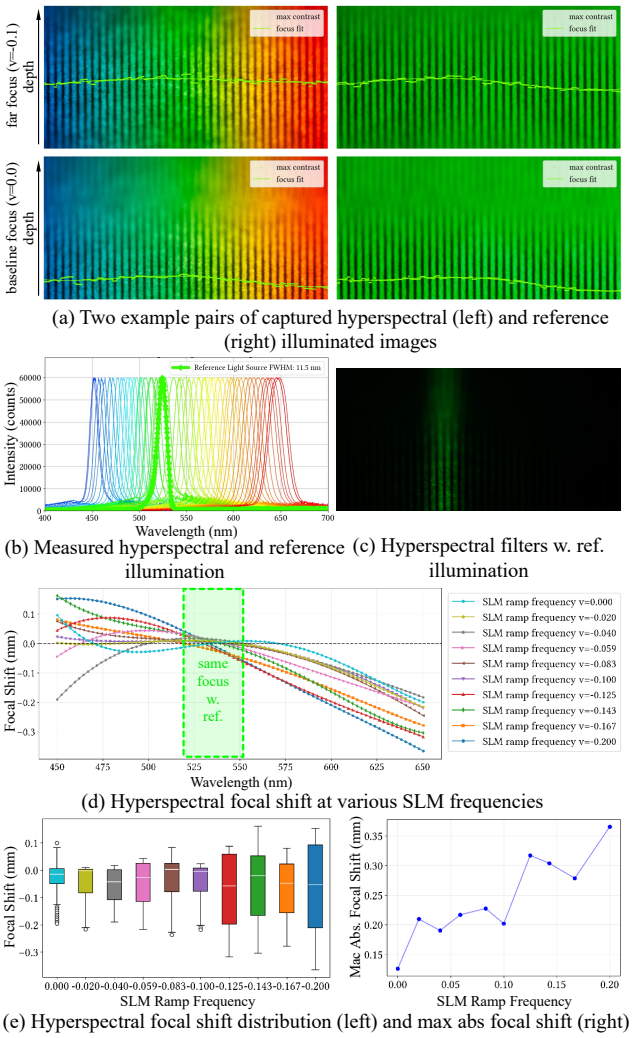
$$\Delta u_{\text{max}} = f_0^2 / f_{\text{Loh}} = \pm 2.7\,\text{mm}.$$

From this basic derivation, we can conclude that our system can move the image plane by $\pm\Delta u_{\text{max}}$ from its nominal position. This range of image plane shifts can be mapped to any depth range for focusing in the scene via judicious choice of the main imaging lens in terms of its focal length f_{main} and its position with rest to the image plane, u_{nom}.

Analysis of chromatic aberrations. Consider a point light source emitting polychromatic light. When the point is in focus, the polychromatic light coming from the point spanning the visible spectrum should resolve to the same minimum radius of blur. This is the ideal scenario, where polychromatic light would focus on the same depth plane. However, like most refractive imaging systems, chromatic aberrations are also inherent in our system—different wavelengths of light would focus on a slightly different depth plane. The chromatic aberration in our system is limited in twofold.

The first limit is from the use of the SLM. The spatially-selective focusing modulation is achieved by displaying a pattern comprised of local phase ramps on the SLM. Depending on the wavelength that is used to compute the phase ramps, light at this particular wavelength would be focused to the desired depth, while other wavelengths would deviate slightly to different depths. The range of max deviation, in our system, is in the order of $0.1\,\text{mm}$. However, since the deviation range depends on the frequency of the ramp, it is worth mentioning that one can reduce this number by reducing the maximum frequency needed on the SLM. This can be achieved by increasing the steepness of the cubic phase plate or by using shorter focal length relay lenses so that the same tilt on the SLM plane will induce a larger change in focus.

The second limit is a compound of the chromatic aberrations that already exist in the optical components used in our system. Refractive lenses incur phase delay of light using the thickness of materials that have higher refractive index than air, and thus by nature, they exhibit different focusing powers for different wavelengths of light.



(a) Two example pairs of captured hyperspectral (left) and reference (right) illuminated images

(b) Measured hyperspectral and reference illumination

(c) Hyperspectral filters w. ref. illumination

(d) Hyperspectral focal shift at various SLM frequencies

(e) Hyperspectral focal shift distribution (left) and max abs focal shift (right)

Figure 17. Chromatic focal shift under different focus settings.

To evaluate the focal shift at various SLM frequencies, we design an experiment that allows us to cover depth across the entire visible spectrum in each capture. As shown in Fig. 17(a), a plane is tilted vertically at $55°$ with the hyperspectral spectrum spanned horizontally. We achieve this by using a pair of linearly-varying shortpass and longpass filters displaced $3\,\text{mm}$ apart, and placed under a full spectrum light source. Moving from left to right, each column is effectively illuminated by a bandpass light source. We capture an image for each focus setting and compare it to a bandpass green light ($530\pm11\,\text{nm}$) shown in Fig. 17(a)(right). Both hyperspectral and bandpass green light sources are measured and shown in Fig. 17(b). The focus is identified using the sum of squared Laplacian, and the difference between the two at each column is converted to focal shift in units of the sensor coordinates. Fig. 17(d) shows a plot of the hyperspectral focal shift at various SLM frequencies, including the largest absolute SLM spatial fre-

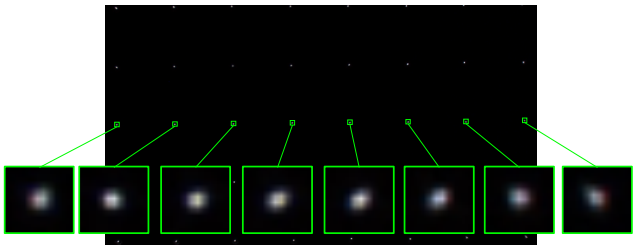


Figure 18. Image captured of a grid of pinholes spanning the full horizontal FOV of the SLM, focused using SLM spatial frequency $v_{pixels} = 0.09$. Insets magnify the image by a factor of $20\times$.

quency we had set for our system, $v_{pixels} = 0.2$.

We observe two findings from our experiment. First, the maximum absolute chromatic focal shift increases with increasing SLM spatial frequencies, as shown in Fig. 17(e)(right). Second, our maximum absolute focal shift is measured to be $0.365\,\mathrm{mm}$ and happens at the extreme SLM phase ramp frequency at the extreme wavelength of $650\,\mathrm{nm}$. This number is slightly larger than that of a single lens, which alone is simulated to have a maximum absolute focal shift of $0.233\,\mathrm{mm}$ at $450\,\mathrm{nm}$ [30]. Therefore, we conclude that chromatic focal shifts is small, and is not significantly impacting the results.

Artifacts caused by the SLM. Aside from chromatic focal shift, the use of an LCOS-SLM also introduces a slight remnant of the unmodulated light in the captured images, due to the 90% fill factor of the SLM and field fringing effects at high frequencies of the SLM. In the case of an all-in-focus image, high spatial frequencies dominate in energy, and the remnant becomes indiscernable.

Off-axis aberration. Our technique of spatially-varying autofocus can correct for defocus in a spatially-varying way. However, other types of aberration that are a compound effect of the optical elements used in the system are present. To achieve good image quality, we used compound camera lenses (3 Samyang 85mm lenses) designed for full frame sensors to maintain good performance over a large area. Our quantitative analysis also aggregates performance across the entire sensor, including off-axis regions. Fig. 11 calculates PSNR and SSIM across the entire sensing region. The MTF chart in Fig. 12(b) averages contrast calculated from images of 3 resolution charts, placed at different depths **and** spatial positions (see Fig. 12(a)).

Moreover, Fig. 18 shows our device focusing on a grid of small pinholes punctured on a black sheet of paper with a needle, back-illuminated with full-spectrum white light; the insets span the full horizontal FOV of the SLM.

System form factor. The proposed imaging system uses three relay lenses and one objective lens, limiting its form factor similar to that of a traditional microscope. While imaging systems can stay stationary and form factor might not matter for a number of use cases, the portability needs to be improved to be suitable for a wider range of use cases, such as everyday use and autonomous driving. This can potentially be achieved in multiple ways. For weight, each lens can be interchanged with an optimized hybrid refractive-diffractive lens [43] so to reduce the total number of elements in all lenses and thus the weight while remaining aberration-free. For size, the proposed 2f-based equivalent setup in Qin et al. [28] can be combined with the end-to-end redesign of the lenses [32, 34] to reduce the length by half with half focal lengths. When the lenses are end-to-end redesigned, the housing of the elements will be more compact since the mountings become custom-made.

Improving light levels. Our current prototype system transmits at most 12.5% of the incident light to the sensor: 75% of light is lost due to the beamsplitter, and another 50% of light is lost due to a polarizer. However, there are potential optical designs that address this issue. It may be possible to remove the beamplitter entirely, either by using the reflective SLM in an off-axis configuration or by using a transmissive SLM. Our LCOS-based SLM requires the incident light to be linearly polarized; other types of SLMs (e.g., PLMs [6]) do not require light to be polarized, which would remove the need for polarization.

D. Pseudocode for autofocusing algorithms

We show the pseudocode for spatially-varying CDAF in Algorithm 1 and spatially-varying PDAF in Algorithm 2. The PDAF algorithm includes SegmentAnything [14] and a layered horizontal-only optical flow algorithm, for which we also show the pseudocode in Algorithm 3. Algorithm 3 uses a modification of the optical flow algorithm by Liu [20] as the backbone. It takes a DP image pair and a layer label map, and computes the horizontal optical flow only for each layer. The modification was implemented and recompiled via the Python wrapper developed by Pathak et al. [25].

E. Additional results

Flexible depth-of-field photography We show in Fig. 19 an additional example of Scheimpflug-style focusing.

All-in-focus images We provide additional qualitative comparisons for AIF imaging in Fig. 20 and Fig. 21, as well as interactive and video results in the supplement webpage.

ALGORITHM 1: Spatially-varying CDAF algorithm.

Result: optical AIF image $I_*(x,y)$, depth map $d(x,y)$
$v_*(x,y) \leftarrow 0$ # current phase gradient map
$l(x,y) \leftarrow abs(v_{SLM_{max}})$ # current search range
$v_-(x,y) \leftarrow v_*(x,y) - l(x,y)/2$ # left gradient map
$v_+(x,y) \leftarrow v_*(x,y) + l(x,y)/2$ # right gradient map
$k \leftarrow \inf$ # initialize stopping criteria
$I_*(x,y) \leftarrow$ Capture() # current optical AIF image

while $k \geq$ *threshold* **do**
  $l(x,y) \leftarrow l(x,y)/2$
  #— Acquire search images —#
  $\phi_-(x,y) \leftarrow 2\pi v_-(x,y)(x+y)$ # left phase map
  $\phi_+(x,y) \leftarrow 2\pi v_+(x,y)(x+y)$ # right phase map
  $I_-(x,y) \leftarrow$ Capture$(\phi_-(x,y))$
  $I_+(x,y) \leftarrow$ Capture$(\phi_+(x,y))$
  #— Get contrast and gradient stack —#
  $b(x,y) \leftarrow$ SuperpixelLabels$(I_*(x,y))$
  $C(x,y,i \in \{-,*,+\}) \leftarrow$ Contrast(I_-, I_*, I_+, b)
  $V(x,y,i) \leftarrow [v_-(x,y), v_*(x,y), v_+(x,y)]$
  #— Update gradient maps for the best contrast —#
  $i_*(x,y) \leftarrow \underset{i}{\operatorname{argmax}}\, C(x,y,i)$
  $v_*(x,y) \leftarrow V(x,y,i_*(x,y)]$
  $v_-(x,y) \leftarrow v_*(x,y) - l(x,y)/2$
  $v_+(x,y) \leftarrow v_*(x,y) + l(x,y)/2$
  #— Update optical AIF and search range —#
  $\phi_*(x,y) \leftarrow 2\pi v_*(x,y)(x+y)$; # optical AIF phase
  $I_{current^*}(x,y) \leftarrow$ Capture$(\phi_*(x,y))$ # optical AIF
  $k \leftarrow$ CompareContrast$(I_*(x,y), I_{current^*}(x,y))$
  $I_*(x,y) = I_{current^*}(x,y)$
end
$d(x,y) = 1/\big(a * (v_*(x,y) - abs(v_{SLM_{max}}))\big)$ # depth map

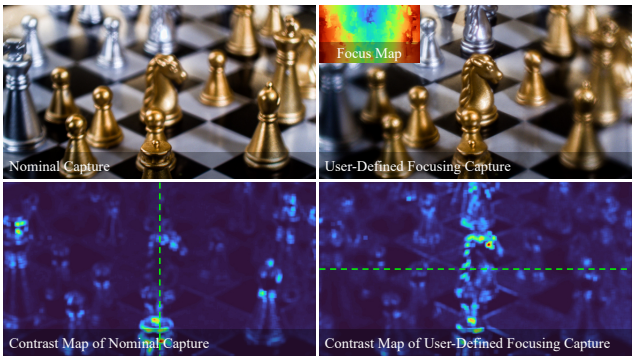


Figure 19. An additional result of a Scheimpflug depth-of-field example. Since our device allows spatially varying focusing, we show an example of Scheimpflug-focusing a scene, where the scene depth varies vertically (top left) but the blur scales horizontally (top right).

ALGORITHM 2: Spatially-varying PDAF algorithm.

Result: optical AIF image $I_*(x,y)$, depth map $d(x,y)$
$I_{left^*}, I_{right^*} \leftarrow$ Capture()
$I_* = (I_{left^*} + I_{right^*})/2$ # current optical AIF image
$b(x,y) = 0$ # layer label map
$v_*(x,y) \leftarrow$ LayeredFlow$(I_{l^*}, I_{r^*}, b(x,y))$ # initialization
$v_*(x,y) \leftarrow$ Superpixelate$(I_*, v_*(x,y))$
$k \leftarrow \inf$ # initialize stopping criteria

while $k \geq$ *threshold* **do**
  #— Capture optical AIF image —#
  $I_{left^*}, I_{right^*} \leftarrow$ Capture$(v_*(x,y))$
  $I_* = (I_{left^*} + I_{right^*})/2$
  #— Perform layered optical flow —#
  $b(x,y) \leftarrow$ SegmentAnything(I_*)
  $v_{current}(x,y) \leftarrow$ LayeredFlow$(I_{left^*}, I_{right^*}, b(x,y))$
  #— Update gradient and depth maps —#
  $v_*(x,y) \leftarrow v_*(x,y) + v_{current}(x,y)$
  $v_*(x,y) \leftarrow$ Superpixelate$(I_*, v_*(x,y))$
  $d(x,y) \leftarrow 1/\big(a * (v_*(x,y) - abs(v_{SLM_{max}}))\big)$
end

ALGORITHM 3: LayeredFlow horizontal optical flow. It computes the horizontal optical flow for each object layer in the DP image pair. This function uses OpticalFlow $[20]^*_x$ as the backbone, which is the modified version of Liu et al. [20]; it keeps only x-direction regularization and smoothing paramters, and compute the flow and regularize only using the valid regions indicated by an input binary mask $m_i(x,y)$. All y direction components and invalid regions are zeroed out. Edge-repeat boundary padding was used during image resizing to construct the Gaussian pyramid.

Input: DP image pair $I_{left}(x,y)$ and $I_{right}(x,y)$, and layer label map $b(x,y)$
Output: optical flow image $v(x,y)$
$n \leftarrow len(\text{unique}(b(x,y)))$ # number of layers
if $n = 1$ **then**
  $v(x,y) \leftarrow$ OpticalFlow $[20]^*_x(I_{left}, I_{right}, \mathbf{1}(x,y))$
  return $v_(x,y)$
end
$v(x,y) \leftarrow 0$
for $i = 0$ **to** $n-1$ **do**
  $m_i(x,y) \leftarrow [b(x,y) == i]$ # layer binary mask
  $I_1(x,y) \leftarrow I_{left}(x,y) * m_i(x,y)$ # layer left image
  $I_2(x,y) \leftarrow I_{right}(x,y) * m_i(x,y)$ # layer right image
  $v_i(x,y) \leftarrow$
    OpticalFlow $[20]^*_x(I_1(x,y), I_2(x,y), m_i(x,y))$
  $v(x,y) \leftarrow v(x,y) + v_i(x,y)$
end
return $v(x,y)$

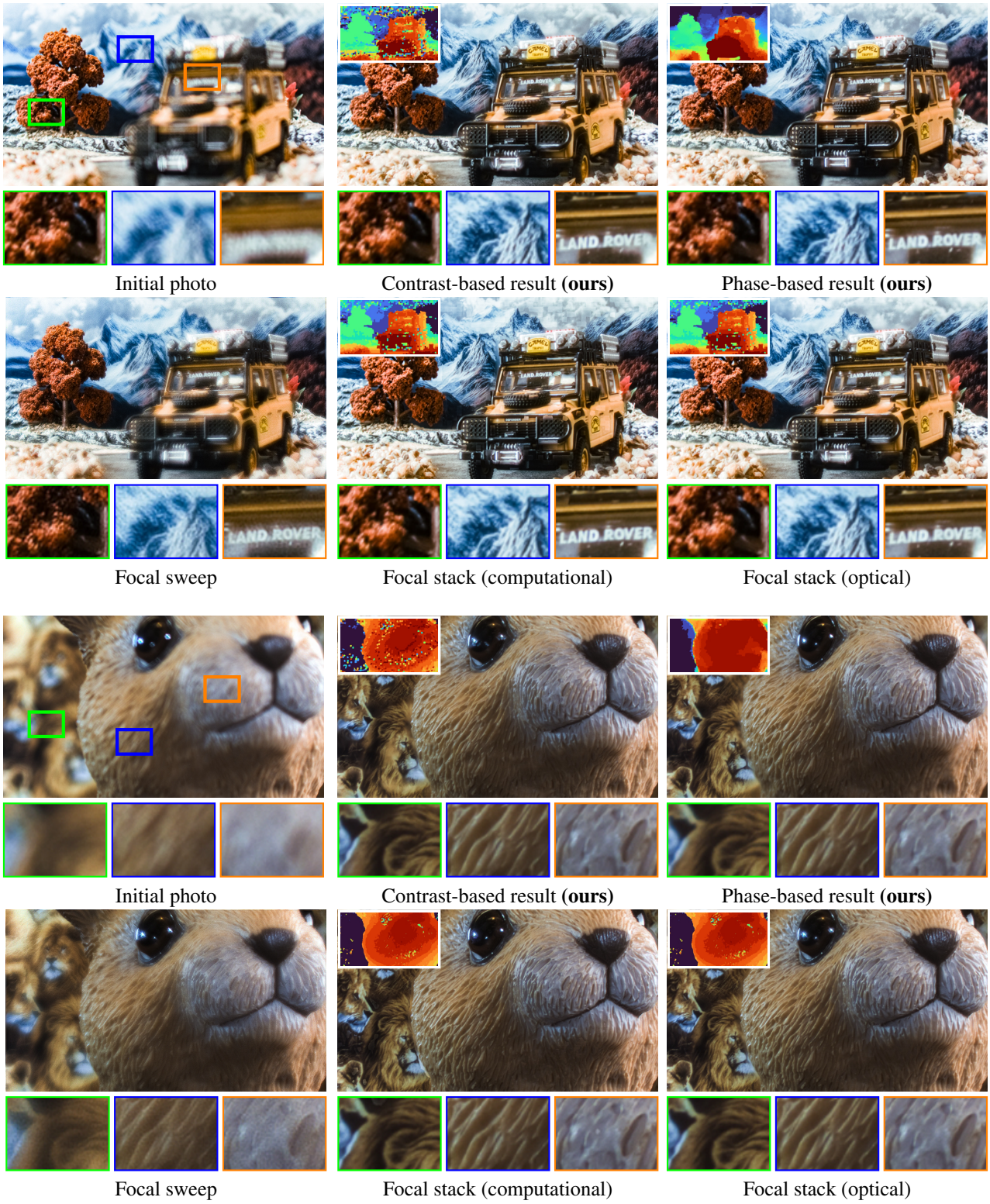


Figure 20. Qualitative comparisons for the *Adventure* and *Bunny* scenes. The recovered depth map is shown in the top-left corner of the corresponding image; this excludes the focal sweep method, which does not output depth.

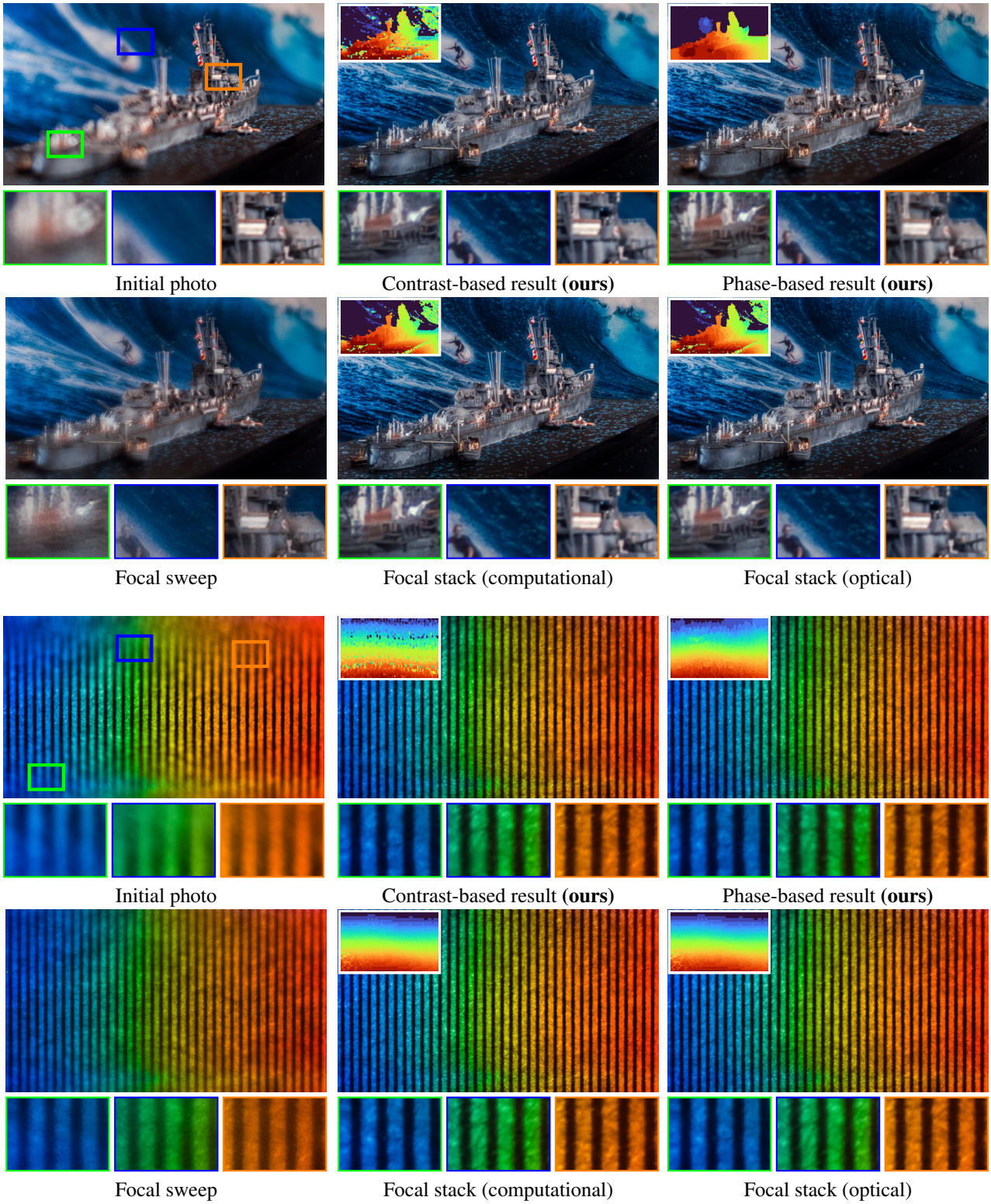


Figure 21. Qualitative comparison for the *Ship* and *Rainbow* scenes.