A Derivation of MMSE Error (8))

$$\epsilon_{MSE} \propto \sigma^2 \mathrm{tr}\left((\mathcal{F}^* \mathbf{\Lambda}\mathbf{\Lambda}^* \mathcal{F})^{-1}\right) \tag{A.1}$$

$$= \sigma^2 \mathrm{tr}\left(\mathcal{F}|\mathbf{\Lambda}|^{-2}\mathcal{F}^*\right) \tag{A.2}$$

$$= \sigma^2 \mathrm{tr}\left(|\mathbf{\Lambda}|^{-2}\right) \tag{A.3}$$

$$= \sigma^2 \sum_i \frac{1}{|\mathbf{\Lambda}_{ii}|^2} \tag{A.4}$$

where we have used the identity $\mathrm{tr}(ABC) = \mathrm{tr}(CAB)$ and the fact that $\mathcal{F}$ is unitary.

B Experiments

B.1 Hardware Proof-of-concept

Implementation Details

Imaging Distance The system is designed such that objects as close as 1.3 m will experience identical PSFs, so we restrict the imaging distance to at least 1.3 m. Periodic Grating Phase Our SLM does not have 100% pixel fill factor, so we use the standard practice of adding a periodic grating phase to our target pupil phase function such that the desired image is separated from the pixel diffraction.

Modulation at Pupil Plane If the optical system is not properly designed, a hardware implementation will deviate from simulation. To prevent this, we use well corrected commercial lenses with the system layout optimized in Zemax OpticStudio (Fig. B.1) until the primary lens' exit pupil was relayed to the SLM plane, and the primary lens' image was relayed to the sensor. The hardware system was laid out according to the optimized design. To align the pupil plane with the SLM, a viewing card was placed at the expected SLM plane. A collimated laser was then coupled into the main lens and the exact pupil plane found by iteratively changing the beam input angle and adjusting the card's axial location until no beam translation was observed as the laser input angle varied (beam translation will be zero only at the correct SLM plane, see Fig. B.1). We then replaced the card with the SLM. To determine the exact sensor distance, we imaged a standard test target, adjusting the sensor distance until the image sharpness was maximized.

Limitations

Monochromatic Imaging Our SLM requires narrow-band illumination which limits our system to monochromatic imaging. This also reduces the system efficiency. Other programmable phase devices such as deformable mirrors, color-multiplexing SLMs, micro-mirror SLMs, or electrowetting lenses would offer programmable phase without these limitations. Our design method could be applied to these hardware systems by using pupil phase parameterization and physical optics methods that better model the specifics of different hardware.

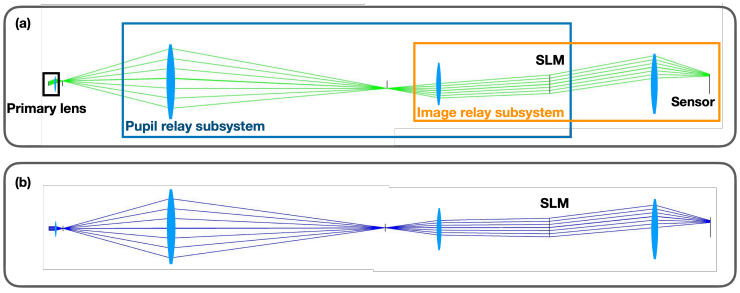

Fig. B.1: (a) First order optical layout of our hardware prototype. The primary lens forms an intermediate image. The pupil relay subsystem images the primary lens exit pupil onto the SLM plane. The Image Relay Subsystem forms the final image, with the SLM acting as the aperture stop. (b) A different input beam angle leads to zero translation of the beam footprint at the SLM plane.

Limited FOV SLM does not have 100% fill factor in each pixel or efficiency to modulate phase. this property of SLM will work as a comb function. In our system, SLM locates at pupil plane or Fourier plane. In the real scene, there will be duplicate images of original scene locates at different diffraction order. Therefore, we need to use SLM to add virtual grating upon the designed phase to move the desired image between the 0th and 1st order of the replicates of original scene images. Also, to minimize the overlapping, we add a field stop at the first image plane after the main lens to cut off the field of view. Overall, on the one hand, we loss some field of view at the imaging plane. On the other hand, we can have a ground truth image side by side the modulated image.

Practicality of SLMs The phase-only SLM used in this project may be not affordable for many price-sensitive applications. However, many controllable optics of lower complexity exist (*e.g.*, tunable lenses or MEMS-based modulators, amplitude-only LCD modulators). While we did not explore these hardware items in this paper, in principle, our method can be used to design dynamic privacy preserving systems based around other programmable hardware. We believe our method lays the groundwork for exploring this space and may enable more practical hardware realizations of dynamic privacy preservation in the future.

B.2 Qualitative Results

PSFs A set of PSFs sampled by DyPP is plotted in Table B.1.

Black-box Attacks on DyPP More black-box inversion qualitative results are presented in Table B.2.

B.3 Ablations

Privacy Manifold Bound p The parameter p of (14) determines the privacy bound of (2). This allows control over trade-off between privacy and utility of the

Table B.1: From top to bottom: original image, DyPP image, utility task output, corresponding PSF. Best viewed with zoom.

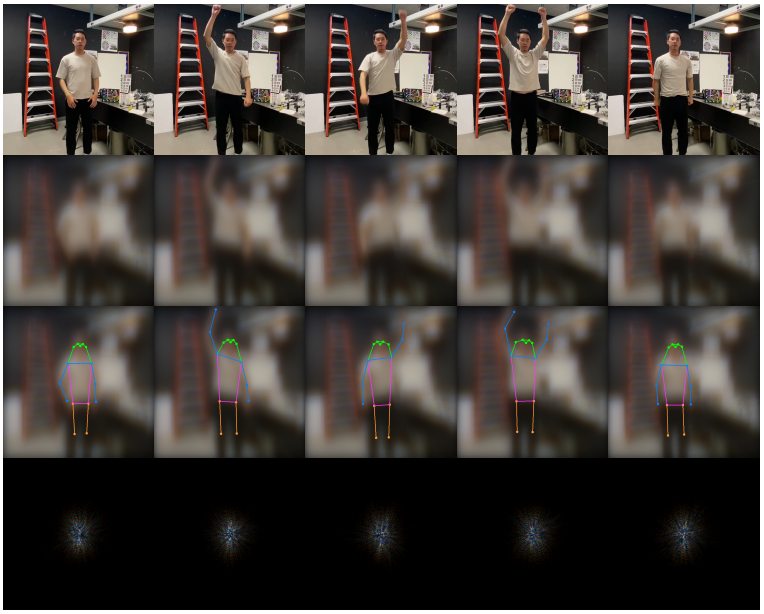

Table B.3: Effective face recognition accuracy (FRA) on PubFig test set and raw measurement's PSNR/SSIM versus p.

p	0.05	0.1	0.2
FRA	0.05	0.09	0.17
PSNR/SSIM	14.4/0.578	14.8/0.571	16.1/0.583

privacy manifold. Table B.3 illustrates this trade-off for values of p between 0.05 and 0.2. It is shown that the DyPP camera is trustworthy in that the effective face recognition accuracy meets the specified bound for different p.

Effect of the Size of $\mathcal{H}$ on Black-box Attacks Figure B.2 ablates the face recognition accuracy by black-box attacks as a function of $|\mathcal{H}|$, the number of PSFs measured by the attacker. It can be seen that the recognition accuracy increases slowly with $|\mathcal{H}|$, remaining significantly lower than that of white-box attacks even for 128 PSFs. Further, the computation cost required by the black-box attack linearly scales with $|\mathcal{H}|$ and can be prohibitive for large $|\mathcal{H}|$.

Effect of $\mathcal{L}^{\text{noninvert}}$ and $\mathcal{L}^{\text{diversity}}$ Table B.4 ablates the effects of $\mathcal{L}^{\text{noninvert}}$ and $\mathcal{L}^{\text{diversity}}$ on the robustness of DyPP to black-box attack. It is shown that both $\mathcal{L}^{\text{noninvert}}$ and $\mathcal{L}^{\text{diversity}}$ is necessary for the optimal performance of DyPP. Especially, in practice, we observe that when $\mathcal{L}^{\text{diversity}}$ is disabled, the trained network tends to generate a limited variety of PSFs (*i.e.* mode collapse) and thus the black-box attacks often succeed.

Table B.2: Some black-box inversion qualitative results. From top to bottom: original image, DyPP image, black-box inversion.

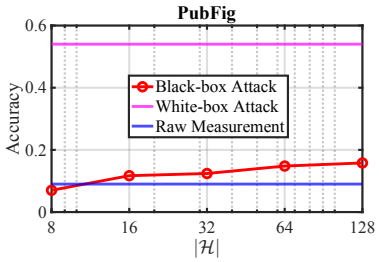

Fig. B.2: FRA versus $|\mathcal{H}|$.

Table B.4: Affects of $\mathcal{L}^{\text{noninvert}}$ and $\mathcal{L}^{\text{diversity}}$ on robustness against black-box inversion attack. ✗ indicates that the loss component is disabled.

		image reconstruction		closed-set face recognition
$\mathcal{L}^{\text{noninvert}}$	$\mathcal{L}^{\text{diversity}}$	PSNR	SSIM	PubFig Accuracy
✗	✗	20.1	0.737	0.47
✓	✗	19.5	0.717	0.42
✗	✓	18.4	0.653	0.19
✓	✓	17.3	0.638	0.14