Seeing Far in the Dark with Patterned Flash

Zhanghao Sun[1*†], Jian Wang[2*], Yicheng Wu[2], and Shree Nayar[2]

[1] Stanford University, 350 Serra Mall, Stanford, CA 94305, USA
[2] Snap Inc., 229 W 43rd St, New York, NY 10036, USA

Abstract. Flash illumination is widely used in imaging under low-light environments. However, illumination intensity falls off with propagation distance quadratically, which poses significant challenges for flash imaging at a long distance. We propose a new flash technique, named "patterned flash", for flash imaging at a long distance. Patterned flash concentrates optical power into a dot array. Compared with the conventional uniform flash where the signal is overwhelmed by the noise everywhere, patterned flash provides stronger signals at sparsely distributed points across the field of view to ensure the signals at those points stand out from the sensor noise. This enables post-processing to resolve important objects and details. Additionally, the patterned flash projects texture onto the scene, which can be treated as a structured light system for depth perception. Given the novel system, we develop a joint image reconstruction and depth estimation algorithm with a convolutional neural network. We build a hardware prototype and test the proposed flash technique on various scenes. The experimental results demonstrate that our patterned flash has significantly better performance at long distances in low-light environments. Our code and data are publicly available.[3]

Keywords: Computational Photography; Flash Imaging; Light Fall Off; Low-light Imaging; Structured Light

1   Introduction

Low-light imaging is critical in consumer photography, surveillance, robotics, and vision-based autonomous driving. Both hardware designs [58], [71], [49], [28], [18], [37], [33], [43] and computational imaging algorithms [14], [12], [10], [24], [42] have been developed for this challenging task. Among these approaches, flash illumination is widely adopted given its outstanding performance and robustness. A flash can easily boost the illumination intensity hence the image signal level by $100 \sim 1000\times$ [18], [6]. With high enough illumination level, it is sufficient to use short exposure time, casual capturing mode, and simple image
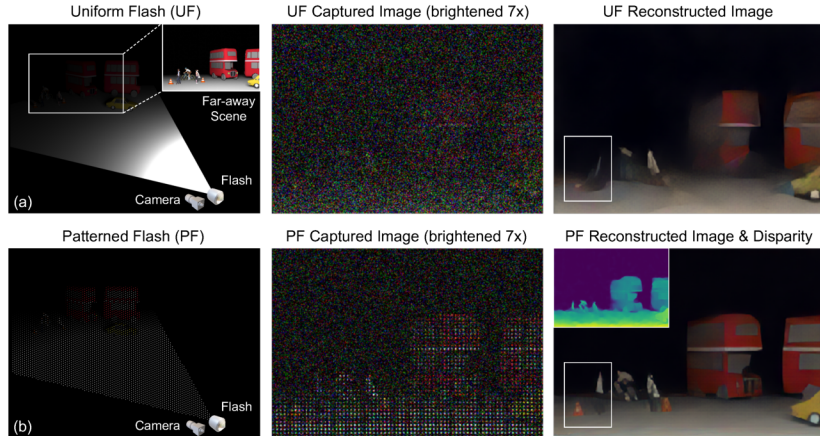

Fig. 1. (Better viewed on full screen to avoid visual artifacts) Conventional, uniform flash (UF) vs. the proposed patterned flash (PF). (a) In UF system, due to inverse square law of light, only close-by scenes are lit up, while the far-away scene is overwhelmed by sensor noise. The ground truth scene is in the inset of the left column. (b) In PF system, illumination power is concentrated into a dot array to enable long distance imaging in low-light environments. The reconstructed image contains important objects and details that are missing in UF reconstructed image (e.g., the pedestrian and the traffic cone in the white box). PF also supports depth estimation with the textured illumination pattern, as shown in the inset of the right column.

processing while still getting high-quality images. However, flash has a fundamental drawback; its strength attenuates quadratically with distance. With a limited power budget (which is the practical constraint for most mobile devices like phones, drones, and wearable devices), the conventional flash illumination system is ineffective at a long distance. This raises robustness and safety issues in surveillance and navigation (e.g., missing important objects at night time). It also leads to irreversible content degradation in photography.

In this paper, we propose patterned flash (PF) technique for low-light imaging at a long distance. Instead of distributing the flash light uniformly across the camera's field of view (FOV), we concentrate it into a dot array pattern. Considering the same power budget, in a conventional uniform flash (UF) system, flash signals from far-away objects are overwhelmed by the sensor noise or fall below the sensitivity threshold of the sensor, leading to severe information loss (as shown in Fig. 1(a)). Contrarily, in the PF system (Fig. 1(b)), the perceived dot array contains higher signals from far-away objects, despite being sparse. By leveraging the redundancies in natural images and capabilities of modern neural networks, we can reconstruct far-away scenes with higher quality. As shown in the third column of Fig. 1, PF's reconstruction of red buses is better; more importantly, the pedestrian and the traffic cone (indicated by the white box) are

completely missed in UF's result while still visible in PF's. Downstream tasks, including object detection [15], [72] and semantic segmentation [11], can also benefit from this higher quality reconstruction. Note that concentrating optical power has also been proposed to improve the sensing distance of active 3D sensors [50], [5], [62], [22], but in that context, photon noise from strong ambient light sources (e.g., sunlight), instead of sensor noise, dominates.

Apart from higher image quality at a long distance, PF introduces another advantage over UF. With a textured illumination pattern and a small baseline between the light source and the imaging device, the captured image contains depth information, as in the case of structured light (SL) 3D imaging techniques [20]. The slight differences to traditional SL are that our SL's pattern is weak in brightness, and ours has a comparatively micro baseline because of the long imaging distance. Since color and depth information highly correlate in the captured image, existing SL methods cannot be readily applied. The proposed algorithm solves this problem and can estimate sub-pixel disparity accurately.

In summary, our contributions are in three folds:

- We propose patterned flash (PF) for low-light imaging. It relieves flash's fundamental drawback, short distance.
- We develop a deep-learning-based joint image reconstruction and disparity estimation algorithm to reconstruct scene image and depth from the single captured patterned flash image.
- We evaluate our approach with simulations and real-world data captured by a hardware prototype. The results verify that the proposed system has significant improvement over the conventional uniform flash.

2   Related Work

2.1   Flash Imaging

The flash has long been used to enhance image signals in dim environments. It was also applied in many other computational imaging tasks, such as reflection removal [38], BRDF acquisition [31], white balance [29], illumination separation [30], image matting [59], edge detection [52] and geometric estimation [40]. In the context of using a flash to assist low-light imaging, there are three well-known problems, and some solutions have been proposed. (1) Flash images look harsh or unnatural. Flash changes the color tone of the scene, adds highlights to the human face, and makes the nearby objects much brighter than far away. Since the seminal work [51], [16], flash/no-flash techniques have been intensively discussed in literature [70], [39], [21], [67] which denoise the no-flash image with guidance from the flash image. True-tone flash has been widely used in smartphones which uses two LEDs of different colors to mimic the color tone of the scene. (2) Flash is intrusive to human eyes. To make the flash more user-friendly and social friendly, previous research focused on using flashes that operate at wavelengths less sensitive to human eyes, e.g., dark flash (NUV and NIR) [37], [64], deep-red

flash [69], or using weak flash [27]. (3) Flash has a short effective range. Current methods either increase the power or shrink the FOV, and there has been little effort to solve this problem from a computational imaging perspective. High power (10 ∼ 100 Watts) LEDs have been used to lighten large scenes at night (e.g., football field), but such systems cannot be applied to consumer devices and introduce heavy light pollution. Another approach is to concentrate light into a smaller FOV. Matrix headlight has been adopted in high-end vehicle models [3]; when necessary, it concentrates flash power into a smaller angular range to reach a longer distance. However, sacrificing FOV is a non-optimal solution.

More broadly, flash illumination or active illumination is also often used in other imaging modalities, where the environmental light level is low, including active hyper-spectral, short-wave infrared [57], middle-wave infrared [36], long-wave infrared [35], and near-ultraviolet imaging. These techniques also face the challenge of light fall off, and they may face more challenges than active, visible-light imaging in that the sensor qualities [54] and light emitters' optical efficiencies [61] are lower, or the power is constrained by safety concerns.

2.2   Low-light Imaging without Flash

Apart from flash imaging, there are other popular low-light imaging methods, which can be categorized into single image-based and burst-based methods. Usually, they require more ambient light than the flash-based methods. Efforts have been made to denoise a single image from early years of image processing to date [14], [12], [10], [65]. Chen et al. [12] first introduced a modern neural network model on RAW data for low-light imaging. Wei et al. [65] demonstrated that carefully modeling sensor noise in the training data plays an important role in improving the denoising performance. Though progress has been made, it is fundamentally difficult to recover the scene details without hallucination in single-image-based methods. Burst-based imaging has achieved more impressive results recently [24], [42], [47], [68]. By fusing a sequence of images taken with a short exposure time, SNR is greatly boosted. The major difficulty for this approach is aligning the burst frames; some algorithms are based on explicit optical flow estimation [24], [42] while others utilize implicit kernel prediction networks [47], [68]; for both approaches, the extremely low-light condition makes the image registrations unreliable, and then the reconstruction quality reduces significantly [47]. Burst-based methods also require the camera to be still and the scene to be static, which may not hold in some circumstances.

2.3   Structured Light (SL) 3D Imaging

SL is one of the most widely used techniques in 3D imaging [20]. Although there are advances which can speed up dot scanning [63], line scanning [8], and time-multiplexing gray code [60], the random dot pattern is usually adopted in consumer devices, like Microsoft Kinect [44] and Apple Face ID [45], because it supports single-shot depth estimation. Existing algorithms, like block matching or HyperDepth [17], assume the scene of interest is close, the baseline is

sufficiently large, and the dots are saturated in the image. Depth can be accurately estimated by the number of pixel shifts of pattern dots. However, these algorithms can not be directly applied in our small-baseline system, since the sub-pixel disparities mainly lead to subtle color and intensity changes. Thus, the depth and image estimations are highly correlated. Saragadam et al. [56] proposed micro-baseline SL to estimate sub-pixel disparities, but it requires a triangular wave pattern profile. Riegler et al. [53] proposed a self-supervised learning approach by modeling the joint distribution of scene albedo and depth, but it requires sufficient ambient illumination to separate scene albedo and depth. Low-light environments make this separation impractical. The proposed PF is a novel SL system where color and depth information are naturally mixed. Estimating either attribute requires knowledge about the other one, and the baseline is very small. We propose a new algorithm that can estimate both accurately.

3   Image Formation Model

In the proposed PF system, we capture one RAW image $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$, with the PF as the dominant illumination light source[4] (with H, W being spacial dimensions and C being the number of color channels). A reference pattern $\mathbf{P_r} \in \mathbb{R}^{H \times W \times C}$ is captured in the calibration stage, and the image/reference pair $\{\mathbf{I}, \mathbf{P_r}\}$ is used as the input into the reconstruction algorithm. Both $\mathbf{I}$ and $\mathbf{P_r}$ are $\in [0, 1]$. The image formation model for $\mathbf{I}$ can be expressed as:

$$\mathbf{I} = \frac{\mathcal{W}_\mathbf{d}[\mathbf{P_r}]}{\mathbf{d}^2} \odot \mathbf{A} + \mathcal{N} \tag{1}$$

In Eq. 1, $\mathcal{W}_\mathbf{d}$ is the depth-dependent warping operation. $\mathcal{W}_\mathbf{d}[\mathbf{P_r}]$ is the warped flash pattern. $\mathbf{d}$ is the scene depth, and $\mathbf{A}$ is the scene image. Here we include scene surface normal and material properties in $\mathbf{A}$. $\odot$ is the pixel-wise product. $\mathcal{N}$ includes photon noise, read noise, quantization noise, and other noise sources [65]. We leverage this image formation model in synthetic data generation, for network training and evaluation. We also implement a photometric loss based on this model for better supervision.

3.1   Signal-to-noise Ratio Analysis

Here we theoretically analyze the advantage of PF over UF and elucidate the assumptions used in the comparison. In the UF system (Fig. 2(b)), each pixel receives signal s, the signal-noise-ratio at each pixel can be expressed as:

$$SNR_{UF} = s/\sqrt{s + \sigma_r^2} \tag{2}$$

where σ_r is the standard deviation of signal-independent sensor noise (e.g., read noise), and photon shot noise of signal s has variance s.

---

[4] Our method can be extended to the scenarios with stronger ambient illumination. Please refer to supplementary material for details.
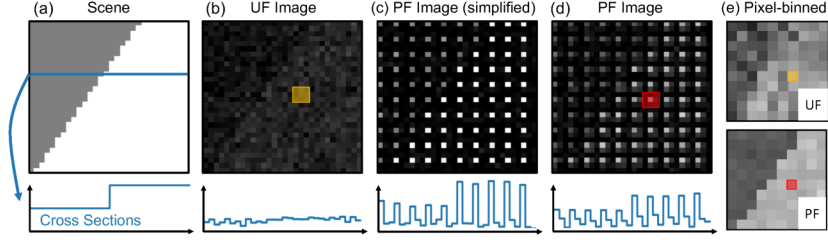
Fig. 2. Signal-noise-ratio comparisons and "pixel-binned" images. (a) Scene. (b-d) Comparisons between UF and PF captured images with cross-sections. (e) Comparisons between "pixel-binned" UF and PF images, as discussed in Sec. 3.1. The yellow and red pixels in upper and lower parts of (e) correspond to the $M \times M$ patches ($M = 3$ in this case) in (b) and (d), respectively.

In the PF system, however, suppose the entire signal in an $M \times M$ patch is concentrated in one camera pixel (Fig. 2(c), $M = 3$). The "in-pattern" pixel receives signal $M^2 s$, and the SNR for this pixel can be expressed by SNR_{PF} in Eq. 3. However, this SNR enhancement is at the expense of lower spatial resolution. For a fair comparison, we average the entire signal in the UF image patch to get the SNR at the same spatial resolution (SNR_{UFavg} in Eq. 3). We notice that when σ_r dominates, SNR_{PF} can be around M times higher than SNR_{UFavg}. Generally, this holds in low-light conditions [19], [23].

$$\begin{cases} SNR_{PF} = M^2 s / \sqrt{M^2 s + \sigma_r^2} \approx M^2 s / \sigma_r \\ SNR_{UFavg} = M^2 s / \sqrt{M^2 s + M^2 \sigma_r^2} \approx M s / \sigma_r \end{cases} \tag{3}$$

For real-world patterns, $\mathbf{P_r}$ may not be exactly "1 in $M \times M$" (Fig. 2(d)). Each pattern dot is blurred, and the overall pattern contrast is lower. We can define a similar input signal SNR comparison. For a PF image, we set a threshold to define whether an image pixel falls within a pattern dot. Then we only extract "in pattern" image pixels from each $M \times M$ pixels patch (red rectangle in Fig. 2 (d)) and combine them into a low-resolution image (as shown in Fig. 2(e), lower part). For a UF image, we average the captured image to the same spatial resolution (as shown in Fig. 2(e), upper part). In the following, we use this "pixel-binning" method to visualize the quality of image signal before processing.

We also define the "average occupancy" of $\mathbf{P_r}$ as its mean value. A pattern with average occupancy $1/\Omega$ can be regarded as an equivalence of the simplified "1 in $\sqrt{\Omega} \times \sqrt{\Omega}$" pattern, and therefore has theoretical upper bound of input SNR gain $\sqrt{\Omega} \times$ (when signal $\to 0$ and contrast $\to \infty$). With smaller average occupancy, the pattern is sparser, and the SNR gain is larger. However, recovering the image content between pattern dots becomes more difficult. Empirically, we found average occupancy $= 1/16$ achieves good balance in this trade-off. Note that higher photon shot noise and lower flash pattern contrast would reduce
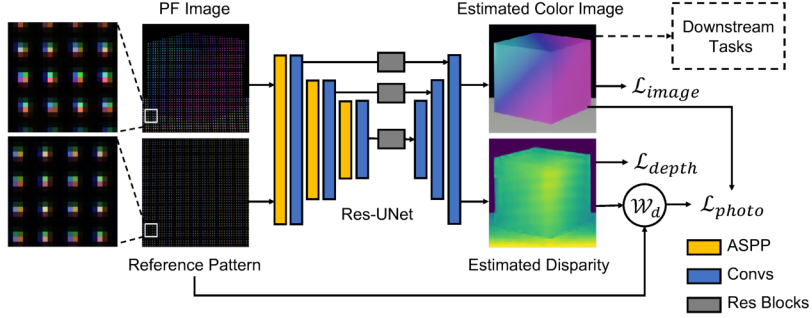
Fig. 3. Joint scene image and depth estimation. We use captured PF image and a reference pattern as input. The neural network predicts a scene image and a disparity map. We use three loss functions to supervise the network training: image loss $\mathcal{L}_{image}$, depth loss $\mathcal{L}_{depth}$, and photometric loss $\mathcal{L}_{photo}$. The reconstruction results can be further employed in downstream tasks (e.g., object detection).

the SNR gain from its theoretical limit. However, we demonstrate that, in both simulations and real-world experiments, PF improves image quality significantly.

4   Patterned Flash Processing

4.1   Network Architecture

The network architecture is shown in Fig. 3. The goal of our reconstruction algorithm is to jointly estimate a scene image $\widehat{\mathbf{A}}$ and the scene depth $\widehat{\mathbf{d}}$ from PF image $\mathbf{I}$. We use a U-Net [55] as backbone, with multiple residual blocks [26] as skip connections. Since information is distributed sparsely in $\mathbf{I}$, a big receptive field is required for feature extraction. We use an Atrous Spatial Pyramid Pooling (ASPP) module [13] to balance the receptive field and computational cost. For details of the network architecture, please refer to supplementary material.

4.2   Loss Functions

We train the network in a supervised manner, with ground truth color image $\mathbf{A}$ and depth $\mathbf{d}$. We use three types of loss functions; supervision on image reconstruction $\mathcal{L}_{image}$, supervision on depth estimation $\mathcal{L}_{depth}$, and photometric loss [53], [73] $\mathcal{L}_{photo}$. The total loss is a weighted combination.

$$\mathcal{L} = \mathcal{L}_{image} + \lambda_1 \mathcal{L}_{depth} + \lambda_2 \mathcal{L}_{photo} \tag{4}$$

The image reconstruction loss consists of three parts, an L2 loss, a perceptual loss [32] (with VGG16 as feature extractor), and an L1 gradient loss.

$$\mathcal{L}_{image} = |\widehat{\mathbf{A}} - \mathbf{A}|^2 + \beta_1 |\text{VGG}(\widehat{\mathbf{A}}) - \text{VGG}(\mathbf{A})|^2 +$$
$$\beta_2 (|\nabla_x \widehat{\mathbf{A}} - \nabla_x \mathbf{A}| + |\nabla_y \widehat{\mathbf{A}} - \nabla_y \mathbf{A}|) \tag{5}$$
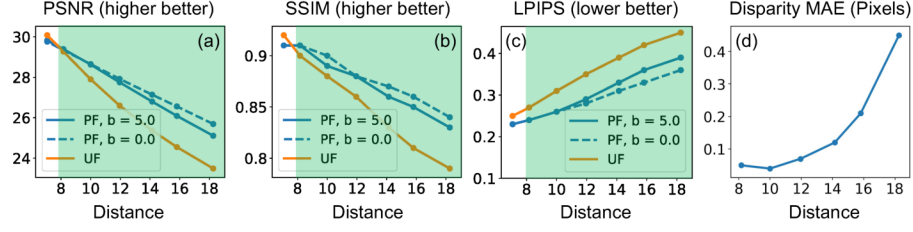
Fig. 4. Quantitative comparisons between PF and UF on synthetic data (distance unit defined in Sec. 6.1). (a, b) Image reconstruction quality from the PF system consistently outperforms that in the UF system at a long distance (green region). Also, increasing the baseline only slightly influences the image reconstruction performance. (c) Mean absolute error (MAE) of disparity estimation.

For depth estimation supervision, we use an L2 loss.

$$\mathcal{L}_{depth} = |\widehat{\mathbf{d}} - \mathbf{d}|^2 \tag{6}$$

The photometric loss utilizes the image formation model to enforce a consistent depth and image prediction.

$$\mathcal{L}_{photo} = |\mathcal{W}_{\widehat{\mathbf{d}}}[\mathbf{P_r}] \odot \widehat{\mathbf{A}} - \mathcal{W}_{\mathbf{d}}[\mathbf{P_r}] \odot \mathbf{A}|^2 \tag{7}$$

5   Implementations

We train the network on data synthesized from FlyingThings3D dataset [46], with ground truth image and depth. We use experimentally captured reference patterns for both network training and evaluations. For each scene, we set a distance $\in [8, 16]$. Compared with imaging an object at distance $= 1$, this corresponds to an attenuation of the color image (normalized to $[0, 1]$) by $64\times \sim 256\times$. Note that we use a uniform attenuation factor across one scene in training data synthesis instead of a depth-dependent attenuation $1/\mathbf{d}^2$. This is to not weigh the farther objects less in loss function $\mathcal{L}_{image}$. We assume a heteroscedastic Gaussian noise model to account for both read and shot sensor noise. For an attenuated image $\mathbf{I_0}$, the noise variance is defined as $\sigma_r^2 + \sigma_s^2\mathbf{I_0}$. For each scene, we randomly select $\sigma_r \in [0.002, 0.005]$ and $\sigma_s \in [0.015, 0.04]$. We also add a small "row noise", with $\sigma_{row} = 0.0005$, to be consistent with our CMOS sensor used in real-world experiments [65]. For each scene, we randomly choose the sign of disparities and randomly select a baseline between $0.5 \sim 5.0$, consistent with the unambiguous range defined by the reference pattern.

   We train our model using the Adam optimizer [34], with learning rate 2×10^{-4}, for a total of roughly 200k iterations. The training process takes around 1.5 days on a single Nvidia Tesla-V100 GPU.

6   Simulation Results

6.1   Joint Image and Depth Estimation

We first evaluate the image reconstruction quality at discrete distances. We use an experimentally captured pattern with average occupancy $\sim 1/16$ (as discussed in Sec. 3.1) , which is a real-world equivalence of "1 in 4×4" pattern. We fix the noise settings with $\sigma_r = 0.004$ and $\sigma_s = 0.02$. We set the distance values at $\{8, 10, 12, 14, 16, 18\}$ (the distances are defined such that no image attenuation at distance $= 1$), and assume the reference pattern is at distance $= 8$ in calibration. We use a baseline $= 5.0$, which leads to disparities $= \{0, 1.0, 1.7, 2.1, 2.5, 2.8\}$ pixels at the evaluation distances.[5] The evaluation is conducted on a test set with 128 scenes. The quantitative comparisons between these two approaches are shown in Fig. 4 (a, b, c). It can be seen that PF consistently outperforms UF at long distances (green region). When flash power is more attenuated (with a farther imaging distance), the difference between the two approaches is larger. Note that at closer range (e.g., < 8), PF has lower reconstruction quality compared with the UF. We attribute it to the fact that the PF system discards part of the high-frequency information. This is the major limitation of PF technique. For more discussions, please refer to Sec. 8.

We further evaluate the influence of the baseline. We evaluate PF image reconstruction with a zero baseline (disparity is always zero). As shown by the blue solid lines and dashed lines in Fig. 4 (a, b, c), the discrepancies in image reconstruction with and without baseline are generally small. When distance increases, performance of the system with baseline drops slightly faster due to more inaccurate depth estimation with lower signal levels. We also evaluate the mean absolute error (MAE) of disparity estimation (Fig. 4 (d)). The proposed system achieves sub-pixel disparity estimations in the distance range of interest, which provides a rough depth estimation.

As discussed in Sec. 3.1, we can directly compare the *captured* image quality through a "pixel-binning" method. The theoretical upper bound of input SNR gain, with the specific pattern being used, is $4\times$ (12dB). With the existence of photon shot noise and lower pattern contrast, we get $\sim$ 6dB input SNR gain across the distance range. A qualitative comparison of the "pixel-binned" input images is shown in the second column of Fig. 5. PF images contain most edges and textures in the scene, while some of these details are invisible in UF images, even after averaging. In this way, PF enables much more detailed reconstructions for far-away scenes (as shown in the third column of Fig. 5). We also show the disparity estimation results from PF images (thr fourth column of Fig. 5). Note that the algorithm is unable to predict a decent depth map from UF images.

6.2   Illumination Pattern Design

Designing optimal patterns in structured light has long been investigated [48], [66]. Since for natural scenes, important details appear with uniform proba-

---

[5] The disparity estimation is not influenced by the reference's distance as long as the disparity is within reasonable range.
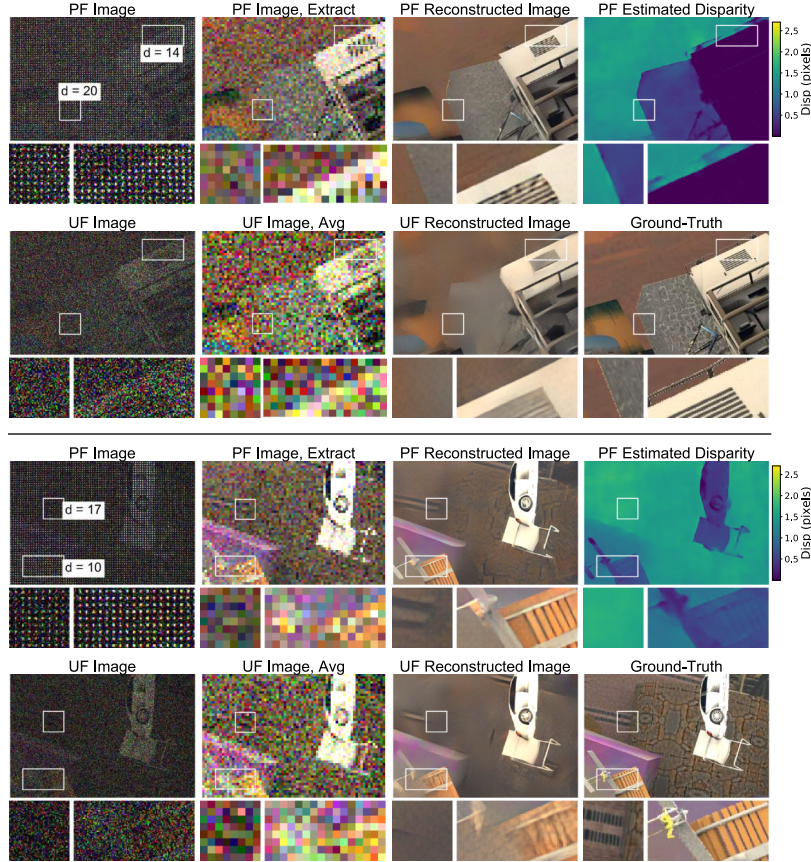
Fig. 5. Qualitative comparisons between patterned flash (PF) and uniform flash (UF) on synthetic data. PF preserves more details in the captured image (first and second columns), and supports higher quality reconstruction (third column). PF also provides disparity estimation, which is non-trivial with UF.

bilities across the FOV, we also consider patterns with uniformly distributed sampling dots. Current structured light systems utilize random patterns due to less ambiguity in stereo matching. We compare a "regular" pattern with multiple "pseudo-random" patterns generated from jittering each dot in the regular pattern by $\{-1, 0, 1\}$ pixels. Two examples of the pseudo-random pattern are shown in Fig. 6(a, b). We empirically found that using pseudo-random pattern results in a slightly worse image reconstruction quality (Fig. 6(c)). This might be due to unavoidable "gap" regions in pseudo-random patterns, where information is missing and difficult to be completed (marked by red boxes in Fig. 6(a, b)). Also, in the proposed system and applications, the baseline is much smaller than
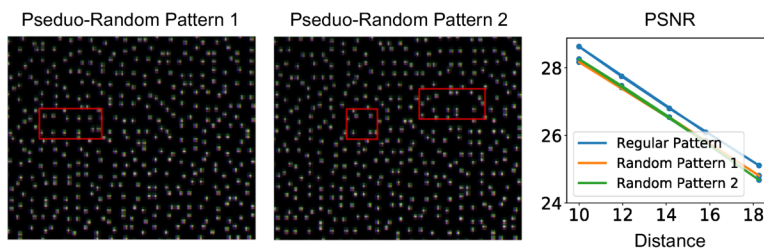
Fig. 6. Pattern design:"pseudo-random" pattern vs. "regular" pattern. (a, b) Part of example pseudo-random patterns generated from the regular pattern. Red boxes show big "gaps" in the patterns. (c) Quantitative comparison on image reconstruction quality between regular pattern and pseudo-random patterns.
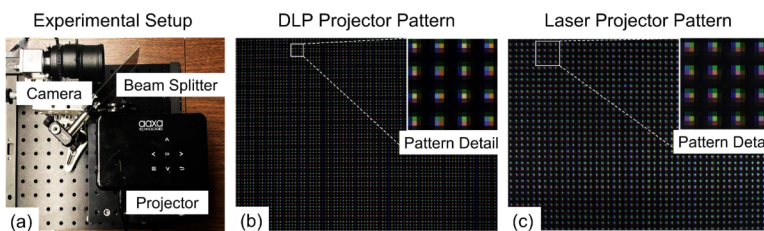

Fig. 7. (a) Photo of the hardware prototype, with a camera, a beam splitter (to enable tunable baseline between camera and projector from $0 \sim 2$cm), and a projector. (b) DLP projector pattern (with zoomed in details). (c) Laser projector pattern (with zoomed in details).

the scene distance, and disparities are smaller than the unambiguous range in regular patterns. As an example, a 90° FOV camera, with 2000×2000 pixels, and a baseline = 3cm between camera and flash light source, would perceive 2.25 pixels pattern dot shift when imaging an object at 20m distance (suppose reference pattern is calibrated at 8m). Therefore, we choose the regular pattern in our setup. More advanced pattern designs (e.g. Poisson disk sampling [9]) or even an end-to-end optimization on the illumination pattern [66, 7] could be left for a future research work.

7   Experimental Results

We implement a patterned flash hardware prototype to evaluate our approach in a real-world environment, as shown in Fig. 7. The system consists of a machine-vision camera (Basler acA2040-120uc), a projector as the PF light source, and a beam splitter. The beamsplitter is used to remove the physical constraints from the bulky camera lens and the projector case, and enable tuning the baseline
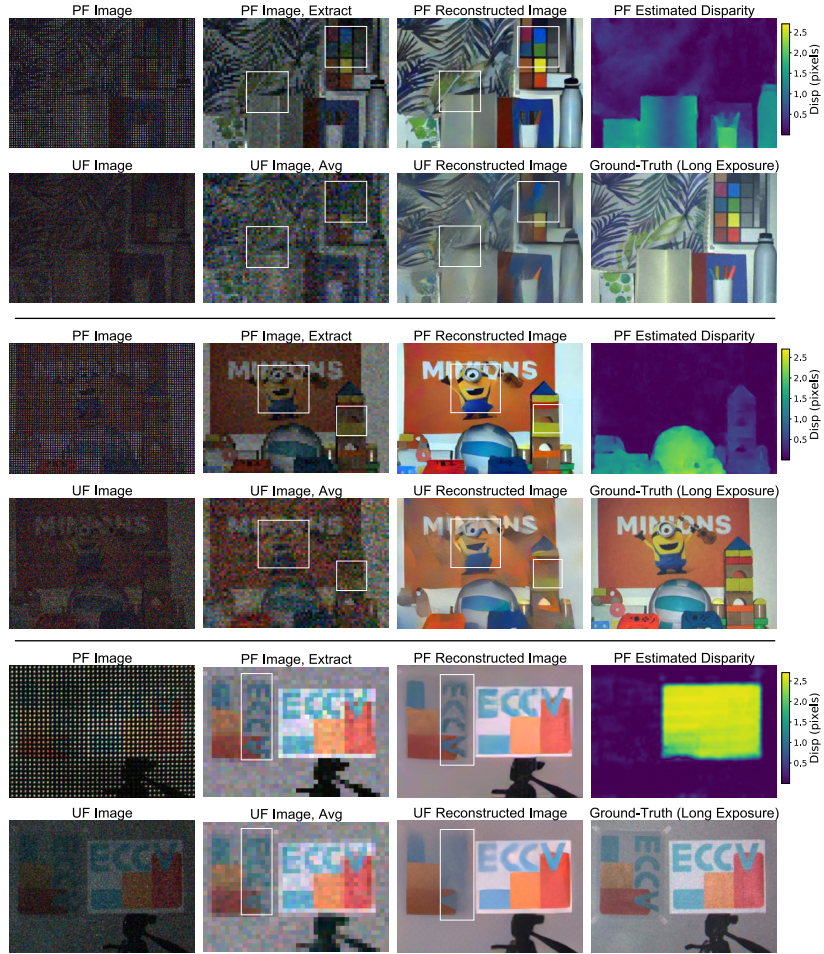
Fig. 8. Experimental results of PF vs. UF on real-world data with a DLP projector (first two rows) and a laser projector (third row). Notice that PF resolves more details, especially in the regions indicated by the white boxes.

between the camera and the projector from 0 to 2cm. We tested our system with two projectors: a DLP projector (Aaxa P300 Pico Projector) and a laser projector (AnyBeam Laser Projector). The DLP projector has a limited depth of focus but with a sharper projected pattern and higher power. The laser projector has a large depth of focus, but has less contrast and lower power. We set the projector display content as a regular pattern for PF, and as a uniform pattern for the UF. The two patterns are both projected onto a white wall for calibration. We adjust the projector brightness to make the two calibration patterns have

PF, Reconstructed Image PF, Estimated Disparity   PF, Reconstructed Image PF, Estimated Disparity

Baseline = 1.5cm   Baseline = 2.0cm

Fig. 9. Experimental image reconstruction and disparity estimation with different baselines, with the same camera and DLP projector illumination settings as that in Fig. 8.

the same per-channel average intensities for a fair comparison. We compare both the captured image and the reconstruction qualities in Fig. 8. In the first setup (Fig. 8 first two rows), the DLP projector projects a pattern with 57×80 dots, and the scene is set to be ~ 3m away (limited by lab space), with a depth range ~ 50cm (limited by projector depth of focus). In the second setup (Fig. 8(b)), the laser projector projects a pattern with 28×46 dots. The scene is set to be ~ 3m aways, with depth range ~ 1.5m.

In Fig. 8, it can be seen that for both setups, PF out-performs UF, with higher visibility of objects under low flash illumination power, as highlighted by the white boxes. This can be observed in both "pixel-binned" images and reconstruction results. Specifically, UF fails to distinguish the outlines of objects (e.g., curtain textures in the first row and toy building blocks in the second row) while PF can faithfully reconstruct the fine textures. In the third row, we placed two boards at different distances. At close range, both PF and UF recovers the board content well enough, due to sufficiently high signal level. At a long distance, UF signals are overwhelmed by the sensor noise, and the reconstructed object is heavily blurred out. The PF system suffers from aliasing artifacts. However, this is a minor effect compared with the artifacts in reconstructed images from the UF. We also demonstrate depth estimation with the experimental setup. As shown in Fig. 8, we are able to estimate disparities in both regions with rich textures (e.g., curtain, stripes on the baseball) and non-textured regions (e.g., white wall, white notebook). Also, as shown in Fig. 9, we tune the baseline between the camera and DLP projector to generate different disparities for the same scene. When the baseline changes in a reasonable range, image reconstruction quality is not influenced.

Object Detection. We further demonstrate the advantage of PF, by applying face detection on the real-world reconstruction images. We use Google MediaPipe [1] as the detector. As shown in Fig. 10, the detector is able to both detect the face and the key points from the PF reconstructed image while it fails with the UF reconstructed image, especially for the mannequin with darker skin color. For more discussions and examples of applying PF reconstructed images in downstream tasks, please refer to the supplementary material.
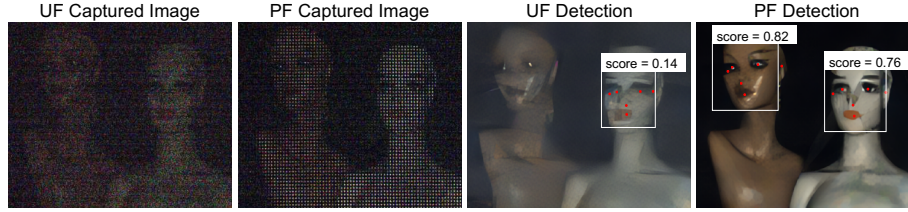
| UF Captured Image | PF Captured Image | UF Detection | PF Detection |

Fig. 10. Real-world face detection with UF and PF. The detector fails on UF reconstructed image but succeeds in PF reconstructed image, especially for the mannequin with darker skin color.

8   Discussions & Conclusions

In this paper, we propose patterned flash for low-light imaging at a long distance. From a noisy patterned flash image, we jointly reconstruct a color image and estimate depth. We build a hardware prototype and demonstrate compelling performance under low-light conditions compared to traditional uniform flash. The proposed technique has the potential to be implemented as an energy-efficient low-light imaging module, which improves visual quality and robustness in applications across disciplines, including photography, navigation, and surveillance.

   Limitations and Envisioning. (1) In reality, the proposed system can be implemented with more stable and more compact devices. Diffractive optical element (DOE) [4], Vertical-cavity surface-emitting laser (VCSEL) arrays [2] or Micro-electromechanical systems (MEMS) mirror can be used. (2) As mentioned in Sec. 6, at close range, patterned flash has lower image reconstruction quality compared with the uniform flash. This originates from the sparse nature of captured patterned flash signals. We envision two variants of the patterned flash system to solve this problem: (i) a combination of uniform and patterned flashes. Power can be distributed between the two light sources according to the required imaging distance and the relative importance of close or far-away objects. (ii) a lens-based system with a large aperture that focuses at infinity (or the maximum imaging distance of interest), to create patterned flash far away and uniform flash (blurred) nearby. (3) The sparse patterned flash signal might also lead to aliasing in the reconstructed image. However, this is a secondary effect when imaging at long distance, where the uniform flash fails to reconstruct even big structures. Also, when there is certain amount of ambient light or the flash has some uniform flash component, this problem can be relieved. Besides, image inpainting and super-resolution algorithms [25], [41] can be applied.

9   Acknowledgement

References

1. Google mediapipe face detector. https://google.github.io/mediapipe/ (2020), [Online; accessed 1-March-2022]
2. Vcsel array applications. http://vixarinc.com/wp-content/uploads/2020/07/ (2020), [Online; accessed 1-March-2022]
3. Audi led matrix headlights. https://media.audiusa.com/en-us/releases/444 (2021), [Online; accessed 1-March-2022]
4. Nil technology introduces flat, multifunctional optics platform for 3d sensing and lidar applications. https://www.nilt.com/nilt-introduces-flat-optics-for-3d-sensing-and-lidar/ (2021), [Online; accessed 1-March-2022]
5. Achar, S., Bartels, J.R., Whittaker, W.L.R., Kutulakos, K.N., Narasimhan, S.G.: Epipolar time-of-flight imaging. ACM Transactions on Graphics (TOG) **36**(4) (July 2017)
6. Altenburg, M.: Understanding flash photography: How to shoot great photographs using electronic flash. PSA Journal **78**(7), 8–9 (2012)
7. Baek, S.H., Heide, F.: Polka lines: Learning structured illumination and reconstruction for active stereo. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5757–5767 (2021)
8. Bartels, J.R., Wang, J., Whittaker, W., Narasimhan, S.G., et al.: Agile depth sensing using triangulation light curtains. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 7900–7908 (2019)
9. Bridson, R.: Fast poisson disk sampling in arbitrary dimensions. SIGGRAPH sketches **10**(1), 1 (2007)
10. Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., Barron, J.T.: Unprocessing images for learned raw denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11036–11045 (2019)
11. Buchholz, T.O., Prakash, M., Schmidt, D., Krull, A., Jug, F.: Denoiseg: joint denoising and segmentation. In: European Conference on Computer Vision (ECCV). pp. 324–337. Springer (2020)
12. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3291–3300 (2018)
13. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **40**(4), 834–848 (2017)
14. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on Image Processing (TIP) **16**(8), 2080–2095 (2007)
15. Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., Tian, Q.: Centernet: Keypoint triplets for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6569–6578 (2019)
16. Eisemann, E., Durand, F.: Flash photography enhancement via intrinsic relighting. ACM Transactions on Graphics (TOG) **23**(3), 673–678 (2004)
17. Fanello, S.R., Rhemann, C., Tankovich, V., Kowdle, A., Escolano, S.O., Kim, D., Izadi, S.: Hyperdepth: Learning depth from structured light without matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5441–5450 (2016)

18. Flint, K.: Flash!: Photography, Writing, and Surprising Illumination. Oxford University Press (2017)
19. Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K.: Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. IEEE Transactions on Image Processing (TIP) **17**(10), 1737–1754 (2008)
20. Geng, J.: Structured-light 3d surface imaging: a tutorial. Advances in Optics and Photonics **3**(2), 128–160 (2011)
21. Guo, X., Li, Y., Ma, J., Ling, H.: Mutually guided image filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **42**(3), 694–707 (2018)
22. Gupta, M., Yin, Q., Nayar, S.K.: Structured light in sunlight. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 545–552 (2013)
23. Hasinoff, S.W.: Photon, poisson noise. (2014)
24. Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., Chen, J., Levoy, M.: Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Transactions on Graphics (TOG) **35**(6), 1–12 (2016)
25. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. arXiv preprint arXiv:2111.06377 (2021)
26. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)
27. Huawei: Huawei x2391-20-t 20t 9mp low-light its ai bullet camera. http://www.tepuhui.com/huawei-x2391-20-t-en (2022), [Online; accessed 1-March-2022]
28. Hubel, P.M., Liu, J., Guttosch, R.J.: Spatial frequency response of color image sensors: Bayer color filters and foveon x3. In: Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V. vol. 5301, pp. 402–407. SPIE (2004)
29. Hui, Z., Sankaranarayanan, A.C., Sunkavalli, K., Hadap, S.: White balance under mixed illumination using flash photography. In: 2016 IEEE International Conference on Computational Photography (ICCP). pp. 1–10. IEEE (2016)
30. Hui, Z., Sunkavalli, K., Hadap, S., Sankaranarayanan, A.C.: Illuminant spectra-based source separation using flash photography. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6209–6218 (2018)
31. Hui, Z., Sunkavalli, K., Lee, J.Y., Hadap, S., Wang, J., Sankaranarayanan, A.C.: Reflectance capture using univariate sampling of brdfs. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 5362–5370 (2017)
32. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision (ECCV). pp. 694–711. Springer (2016)
33. Jung, Y.J.: Enhancement of low light level images using color-plus-mono dual camera. Optics express **25**(10), 12029–12051 (2017)
34. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
35. Koerperick, E.J., Norton, D.T., Olesberg, J.T., Olson, B.V., Prineas, J.P., Boggess, T.F.: Cascaded superlattice inas/gasb light-emitting diodes for operation in the long-wave infrared. IEEE Journal of Quantum Electronics **47**(1), 50–54 (2010)
36. Koerperick, E.J., Olesberg, J.T., Hicks, J.L., Prineas, J.P., Boggess, T.F.: High-power mwir cascaded inas–gasb superlattice leds. IEEE journal of quantum electronics **45**(7), 849–853 (2009)

37. Krishnan, D., Fergus, R.: Dark flash photography. ACM Transactions on Graphics (TOG) **28**(3), 96 (2009)
38. Lei, C., Chen, Q.: Robust reflection removal with reflection-free flash-only cues. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14811–14820 (2021)
39. Li, Y., Huang, J.B., Ahuja, N., Yang, M.H.: Deep joint image filtering. In: European Conference on Computer Vision (ECCV). pp. 154–169. Springer (2016)
40. Li, Z., Xu, Z., Ramamoorthi, R., Sunkavalli, K., Chandraker, M.: Learning to reconstruct shape and spatially-varying reflectance from a single image. ACM Transactions on Graphics (TOG) **37**(6), 1–11 (2018)
41. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE International Conference on Computer Vision (CVPR). pp. 1833–1844 (2021)
42. Liba, O., Murthy, K., Tsai, Y.T., Brooks, T., Xue, T., Karnad, N., He, Q., Barron, J.T., Sharlet, D., Geiss, R., et al.: Handheld mobile photography in very low light. ACM Transactions on Graphics (TOG) **38**(6), 1–16 (2019)
43. Ma, S., Gupta, S., Ulku, A.C., Bruschini, C., Charbon, E., Gupta, M.: Quanta burst photography. ACM Transactions on Graphics (TOG) **39**(4), 79–1 (2020)
44. MacCormick, J.: How does the kinect work. Presentert ved Dickinson College **6** (2011)
45. Mainenti, D.: User perceptions of apple's face id. Information Science, Human Computer Interaction (DIS805) (2017)
46. Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., Brox, T.: A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4040–4048 (2016)
47. Mildenhall, B., Barron, J.T., Chen, J., Sharlet, D., Ng, R., Carroll, R.: Burst denoising with kernel prediction networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2502–2510 (2018)
48. Mirdehghan, P., Chen, W., Kutulakos, K.N.: Optimal structured light a la carte. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6248–6257 (2018)
49. O'Connor, M.: Tested: Huawei p30 pro. Australian Photography (Jun 2019), 66–68 (2019)
50. O'Toole, M., Achar, S., Narasimhan, S.G., Kutulakos, K.N.: Homogeneous codes for energy-efficient illumination and imaging. ACM Transactions on Graphics (TOG) **34**(4), 1–13 (2015)
51. Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., Toyama, K.: Digital photography with flash and no-flash image pairs. ACM Transactions on Graphics (TOG) **23**(3), 664–672 (2004)
52. Raskar, R., Tan, K.H., Feris, R., Yu, J., Turk, M.: Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. ACM Transactions on Graphics (TOG) **23**(3), 679–688 (2004)
53. Riegler, G., Liao, Y., Donne, S., Koltun, V., Geiger, A.: Connecting the dots: Learning representations for active monocular depth estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7624–7633 (2019)
54. Rogalski, A.: Infrared and terahertz detectors. CRC Press (2019)
55. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

56. Saragadam, V., Wang, J., Gupta, M., Nayar, S.: Micro-baseline structured light. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 4049–4058 (2019)
57. Steiner, H., Sporrer, S., Kolb, A., Jung, N.: Design of an active multispectral swir camera system for skin detection and face verification. Journal of Sensors (2016)
58. Sukegawa, S., Umebayashi, T., Nakajima, T., Kawanobe, H., Koseki, K., Hirota, I., Haruta, T., Kasai, M., Fukumoto, K., Wakano, T., et al.: A 1/4-inch 8mpixel back-illuminated stacked cmos image sensor. In: 2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC). pp. 484–485. IEEE (2013)
59. Sun, J., Li, Y., Kang, S.B., Shum, H.Y.: Flash matting. In: ACM Transactions on Graphics (TOG), pp. 772–778 (2006)
60. Sun, Z., Zhang, Y., Wu, Y., Huo, D., Qian, Y., Wang, J.: Structured light with redundancy codes. arXiv preprint arXiv:2206.09243 (2022)
61. Tan, M.C., Connolly, J., Riman, R.E.: Optical efficiency of short wave infrared emitting phosphors. The Journal of Physical Chemistry C **115**(36), 17952–17957 (2011)
62. Wang, J., Bartels, J., Whittaker, W., Sankaranarayanan, A.C., Narasimhan, S.G.: Programmable triangulation light curtains. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 19–34 (2018)
63. Wang, J., Sankaranarayanan, A.C., Gupta, M., Narasimhan, S.G.: Dual structured light 3d using a 1d sensor. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 383–398. Springer (2016)
64. Wang, J., Xue, T., Barron, J.T., Chen, J.: Stereoscopic dark flash for low-light photography. In: 2019 IEEE International Conference on Computational Photography (ICCP). pp. 1–10. IEEE (2019)
65. Wei, K., Fu, Y., Zheng, Y., Yang, J.: Physics-based noise modeling for extreme low-light photography. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) (2021)
66. Wu, Y., Boominathan, V., Zhao, X., Robinson, J.T., Kawasaki, H., Sankaranarayanan, A., Veeraraghavan, A.: Freecam3d: Snapshot structured light 3d with freely-moving cameras. In: European Conference on Computer Vision (ECCV). pp. 309–325. Springer (2020)
67. Xia, Z., Gharbi, M., Perazzi, F., Sunkavalli, K., Chakrabarti, A.: Deep denoising of flash and no-flash pairs for photography in low-light environments. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2063–2072 (2021)
68. Xia, Z., Perazzi, F., Gharbi, M., Sunkavalli, K., Chakrabarti, A.: Basis prediction networks for effective burst denoising with large kernels. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11844–11853 (2020)
69. Xiong, J., Wang, J., Heidrich, W., Nayar, S.: Seeing in extra darkness using a deep-red flash. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10000–10009 (2021)
70. Yan, Q., Shen, X., Xu, L., Zhuo, S., Zhang, X., Shen, L., Jia, J.: Cross-field joint image restoration via scale map. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 1537–1544 (2013)
71. Yang, H., Chao, C.K., Wei, M.K., Lin, C.P.: High fill-factor microlens array mold insert fabrication using a thermal reflow process. Journal of Micromechanics and Microengineering **14**(8), 1197 (2004)

72. Yu, K., Li, Z., Peng, Y., Loy, C.C., Gu, J.: Reconfigisp: Reconfigurable camera image processing pipeline. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4248–4257 (2021)
73. Zhang, Y., Khamis, S., Rhemann, C., Valentin, J., Kowdle, A., Tankovich, V., Schoenberg, M., Izadi, S., Funkhouser, T., Fanello, S.: Activestereonet: End-to-end self-supervised learning for active stereo systems. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 784–801 (2018)