

# Depth Selective Camera: A Direct, On-chip, Programmable Technique for Depth Selectivity in Photography

Ryuichi Tadano  
Sony Corporation  
Tokyo, Japan

Ryuichi.Tadano@jp.sony.com

Adithya Kumar Pediredla, Ashok Veeraraghavan  
Rice University  
Houston, TX, USA

[adithya.k.pediredla, vashok] @rice.edu

## Abstract

*Time of flight (ToF) cameras use a temporally modulated light source and measure correlation between the reflected light and a sensor modulation pattern, in order to infer scene depth. In this paper, we show that such correlational sensors can also be used to selectively accept or reject light rays from certain scene depths. The basic idea is to carefully select illumination and sensor modulation patterns such that the correlation is non-zero only in the selected depth range – thus light reflected from objects outside this depth range do not affect the correlational measurements. We demonstrate a prototype depth-selective camera and highlight two potential applications: imaging through scattering media and virtual blue screening. This depth-selectivity can be used to reject back-scattering and reflection from media in front of the subjects of interest, thereby significantly enhancing the ability to image through scattering media- critical for applications such as car navigation in fog and rain. Similarly, such depth selectivity can also be utilized as a virtual blue-screen in cinematography by rejecting light reflecting from background, while selectively retaining light contributions from the foreground subject.*

## 1. Introduction

Time of flight (ToF) cameras, which acquire a three-dimensional (3D) representation of a scene based on the travel time of illuminating photons, have come of age over the last several years and are poised to revolutionize imaging capabilities. A ToF camera consists of a modulated light source and a specialized image sensor that correlates the waveform received from the scene with the modulation function in order to estimate the amount of time taken, and hence the depth, for light to travel.

ToF cameras provide unique benefits over competing 3D scene acquisition technologies. They do not require the computation of image/feature correspondences, an element of both multi-view and structured light techniques. They also do not require a complicated scanning mechanism as in LIDAR systems. Most importantly, ToF sensors can be

fabricated using traditional CMOS fabrication techniques, thereby benefiting enormously from the scaling, cost and efficiency advantages offered through the advances in semiconductor fabrication techniques.

The inexpensive ToF cameras, such as the Kinect from Microsoft, the CamBoard from PMD Technologies, and Swissranger from MESA imaging has revolutionized a range of consumer imaging applications. The 3D imaging capability offered by these sensors has enabled a new era of untethered and unconstrained game play in modern video gaming consoles [4, 23]. In addition, the environment sensing capabilities of these devices makes many head-mounted augmented, virtual and mixed reality [14] applications realistic and perceptually engaging, without the physiological discomfort associated with previous generation systems that were constrained by the effect of small errors in range and ego motion estimation. Much of the work in ToF sensors and their potential applications have focussed solely on the 3D imaging capabilities provided by these sensors.

**Correlational Sensors for 2D Imaging:** We explore an interesting alternate question. Can these modern ToF sensors that exploit active illumination and correlational sensors be used to enhance the capabilities of 2D imaging and photography? In this paper, we address two important limitations of traditional photography, (a) imaging in scattering media and (b) accurate segmentation of foreground objects for virtual blue-screening. We show that while these two limitations are seemingly unrelated, they both stem from the same fundamental handicap that plagues traditional imaging: the inability to reject contributions from photons that are not from the desired depth range. We show that careful design of the illumination and the sensor codes in correlational sensors will allow us to accept light contributions from a user-controlled, selective, programmable depth range while rejecting all other light. Such depth selectivity may potentially lead to a wide range of imaging and photography applications hitherto not possible as shown in figure 1.

### Technical Contributions:

- *Enhance 2D imaging using correlational sensors:* We

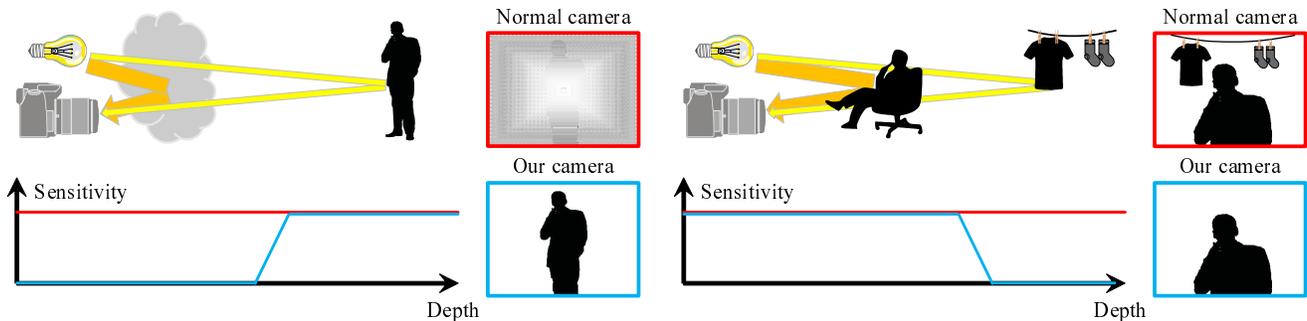


Figure 1. Two example applications of Depth Selective Camera. In our system, we can selectively image objects of interest by controlling the depth sensitivity of the camera. (Left) The sensitivity is designed to suppress the backward reflection from scattering media. (Right) Unwanted objects behind the subject are suppressed. In the plot on the bottom, blue line shows the sensitivity of our camera and red line shows the sensitivity of normal camera.

show that proper design of illumination and sensor codes in correlational sensors can overcome some limits of traditional 2D imaging.

- **Depth Selective Camera:** We show careful design of the illumination and the sensor codes in correlational sensors will allow us to accept light contributions from a user-controlled, selective, programmable depth range while rejecting light contributions from outside this range. In addition, we demonstrate techniques to sharpen the range of interest beyond the limit imposed by the modulation frequency, by using multiple image captures with different codes.
- **Back-scatter resistant photography:** We demonstrate back-scatter resistant photography using our camera and demonstrate several examples where back-scatter from occluders is rejected increasing the range of imaging in highly scattering environments.
- **Virtual blue-screening:** We demonstrate that such depth-selective cameras can be used for virtual blue-screening in teleconference and other applications, since it provides an effective means to segment the subject from its background without the need for any color or texture information.

#### Limitations:

- **Decrease in Light efficiency:** As the proposed technique rejects light contributions outside the selected range, it suffers from a decrease in light efficiency and consequently SNR.
- **Pixel saturation:** Higher light intensities can saturate the PMD sensor resulting in wrong cross-correlation value and hence, the depth selection. However, in some specific cases especially in indoor scenes (like our experiments), the dynamic range of the PMD sensor is good enough.
- **Limited range precision:** The precision of range selection we can achieve is limited by the peak modulation frequency of the light source and PMD sensor.
- **Spatial Resolution:** The spatial resolution of the sensor we use in our experiments is  $160 \times 120$  pixels, resulting

in low resolution images. Over the last few years there is an explosion of work in ToF sensors and higher resolution ToF sensors are slowly entering the mainstream (Kinect 2.0 is a  $640 \times 480$  sensor). The current low resolution images can be improved with the use of these and other higher resolution ToF sensors.

## 2. Related work

In this section, we compare our camera with state-of-the-art techniques employed for imaging in scattering media.

**Gated imaging:** Many applications such as automobile night vision and safety [9, 11], surveillance system [22, 27, 3], long distance identification [1], and mammography [8] use gated imaging for selective range imaging.

In gated imaging system, a short pulsed laser light is emitted and a synchronized super high speed shutter in the camera is opened for a brief duration to allow light reflected back only from the range of interest. These systems suffer from poor light efficiency just like the proposed camera. However, these cameras are not robust to ambient light and also suffer from interference problems. On the contrast, the PMD sensor in our camera subtracts the ambient light making it a more robust solution compared to gated imaging. In addition, single pulsed operation requires heat management of light source as large currents are needed to create sufficient amount of light for low duty pulsed light emission.

**Adaptive illumination:** Tamburo *et al.* [24] have built a programmable automotive headlight system to increase the visibility during rain or snow. Their system first estimates the location of rain drops or snow flakes by detecting their motion. The illumination pattern of the programmable projector is adapted to carefully avoid these rain drops or snow flakes. This decreases the back scatter of the light, increasing the visibility range. The main draw back of their system is its inability to handle a homogeneous scattering medium such as fog. On the contrast, our technique is impervious to the nature of the scattering medium.

**Post processing:** To remove rain drops in a video, a comprehensive model for the visual appearance of rain was proposed by Garg and Nayar [10]. They took advantage of the

temporal information in both detecting and removing rain drops. Barnum *et al.* [2] utilized spatial-temporal frequency information to detect and remove rain and snow in captured videos. These techniques strongly rely on a specific model of rain or snow and cannot handle scenes having fog or mist.

To remove the effect of haze from captured images, Schechner *et al.* [19] introduced a method based on polarization effect of atmospheric scattering. Schwartz *et al.* [21] proposed a blind estimation method for separating the light (haze component) from measurement. He *et al.* [12] proposed dark channel prior, to recover haze-less image. None of these techniques can suppress the backward reflection from haze.

**Time of Flight Camera:** Kadambi *et al.* [15] employed pseudo random code both for illumination and reference signal of a ToF camera. Using such a setup, they were able to do transient imaging, ranging of transparent objects, looking through diffuse materials, and increasing the accuracy of depth maps. Though the application of looking through diffuse materials is similar to our technique, our method is based on straight forward cross correlation design and can be done with only one measurement.

Heide *et al.* [13] also unveiled subjects sinked in scattering medium. They proposed a modified exponential reflection model that describes the impulse response of light reflection from a scattering medium. However, their method requires hundreds of measurements at different phases/frequencies of sinusoidal wave. Thus, their method cannot be applied to real time applications.

**Primal-Dual Coding:** The goal of our technique is direct on-chip suppression of light from regions outside the selected depth range. Conceptually our work is similar to primal-dual coding for direct on-chip direct-only imaging. O'Toole *et al.* [18] present a technique that uses a spatio-temporal modulator as a projector to illuminate the scene and synchronously modulate the exposure duration of each pixel in the sensor. By carefully selecting the spatial-codes on the projector and the sensor, one can ensure that only direct paths contribute to the sensor measurements while indirect light does not affect the sensor measurements. Unlike primal dual coding which requires a spatio-temporal coding both on the illumination and on the sensor, we require only temporal modulation both on the illumination and the sensor, resulting in much simpler hardware requirements.

### 3. Temporal coding in illumination and sensor

We will explain the principles of Photon Mixer Device (PMD) sensor based ToF camera and proceed to our design of a depth selective camera. ToF camera [25, 17, 16, 20, 7] consists of a light source that emits coded illumination ( $g(t)$ ). As illustrated in figure 2, the illumination signal from the light source  $g(t)$  hits the object surface, gets scattered and reaches the sensor pixel with a delay  $\tau$  due to the finite speed of light. The received signal can be given as

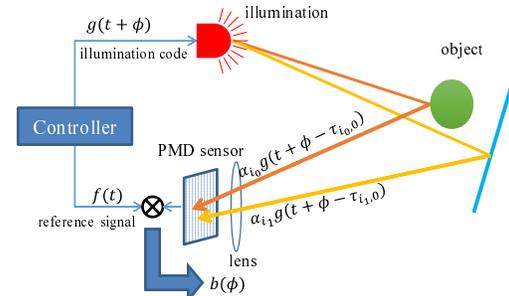


Figure 2. ToF camera system. The system controller sends two binary signals: reference signal  $f(t)$  to the PMD sensor and illumination code  $g(t)$  to the light source. For each pixel, the PMD sensor measures correlation between the reference signal and incident light coming to the sensor.

$\beta g(t - \tau)$ , where  $\beta$  represents the albedo of that particular scene pixel. The PMD sensor calculates the cross correlation between a reference signal ( $f(t)$ ) and received signal  $\beta g(t - \tau)$ . This process is repeated for different phase shifts  $\phi$  between the coded illumination and the reference signal, and these multiple measurements  $b_i(\phi)$ . The output of PMD sensor at each pixel is given by

$$b_i(\phi) = \int_0^\infty \alpha_i(\tau) \cdot \int_0^T \beta g(t + \phi - \tau) f(t) dt d\tau, \quad (1)$$

where  $\alpha_i(\tau) = \int_p \alpha_{i,p} \delta(|p| = \tau) dp$  is attenuation coefficient that is uniquely determined by the path traveled by light,  $T$  is exposure time, and  $\phi$  is phase offset for  $g(t)$ . The attenuation coefficient  $\alpha_i(\tau)$ , accounts for the  $d^2$  propagation loss of light intensity: without this attenuation compensation the intensity (albedo:  $\beta$ ) of objects farther away will be under-estimated, though their depth would still be correctly estimated.

In ToF camera used in Kinect to capture depth information,  $f(t)$  and  $g(t)$  are both sinusoidal waves. The frequencies of both the signals are usually the same. Three or four measurements with different amounts of phase shift are required to generate depth information. When multiple cameras are in operation, custom code such as pseudo random sequence can be utilized as in [5, 26] so as to avoid interference between the cameras.

### 4. Range selection via cross correlation design

Given that the sensor measurements can be given as the cross-correlation between illumination codes  $g(t)$  and sensor codes  $f(t)$ , one can control the depth-selectivity properties of such a measurement by appropriately designing illumination and sensor codes.

**General Problem Definition:** In this section, we propose a Depth Selectivity camera based on ToF technology. We pose it as the problem of designing the illumination signal  $g(t)$  and reference signal  $f(t)$  such that the cross-correlation provides the desired depth selectivity. In

general, let's assume that we would like to acquire an image with an arbitrary depth-dependent attenuation function  $\gamma(\tau)$ : i.e., given a scene with albedo  $\beta(x, y)$  and depth-map  $z(x, y)$ , we would like to directly acquire an image  $I(x, y) = \gamma(\tau_{z(x,y)})\beta(x, y)$ . Without loss of generality, let's drop the pixel identifier  $(x, y)$  since all operations are performed independently at each pixel. Mathematically,

find  $f$  and  $g$  such that:

$$\int_0^T g(t + \phi - \tau)f(t) dt = \gamma(\tau); \quad f(t), g(t) \in \mathbb{Z}. \quad (2)$$

i.e., find codes  $f$  and  $g$  such that their cross-correlation is the desired depth-dependent attenuation function  $\gamma$ . Rewriting the above problem statement as minimization problem,

$$\{f, g\} = \arg \min_{f, g} \left( \int_0^T g(t + \phi - \tau)f(t) dt - \gamma(\tau) \right)^2 \quad (3)$$

such that:  $f(t), g(t) \in \mathbb{Z}$ .

The discrete version of the optimization problem in equation 3 has the same constraints of the integer programming problem and further, the cost function is quadratic in both  $f$  and  $g$ . Hence, the optimization problem is NP-hard and we resort to heuristics to solve equation 3.

#### 4.1. Code design

In this section, we describe a heuristic to design illumination signal  $g$  and reference signal  $f$  that minimizes the cost function in equation 3 with out violating the design constraints. We use m-sequences [6] that have triangle shaped auto correlation as shown in figure 3 as basis for code design. The horizontal axis "shift amount" corresponds to the distance light travels. This means that when we send the same m-sequence as  $f(t)$  and  $g(t)$  in equation 1, only the objects whose distance correspond to the peaks in this cross correlation can be seen. The interval between the peaks is controlled by the bit length of the m-sequence.

We design a code with desired cross correlation function by superimposing phase shifted m-sequences. We can implement the superimposition by re-writing either  $f(t)$  or  $g(t)$  as a linear combination of shifted versions of the other code and using m-sequence for the other code. For example, if we superimpose  $f(t)$  to get  $g(t)$ , we have:

$$f(t) = \sum_i \beta_i g(t + \phi_i). \quad (4)$$

By controlling  $\beta_i$ , we can separate the scene into layers based on the depth and observe them independently. These layers allow us to recreate arbitrary range of interest just by stacking them (see figure 4).

**Range of Interest Camera:** For a range of interest camera, we would like the attenuation function  $\gamma$  to be one (no

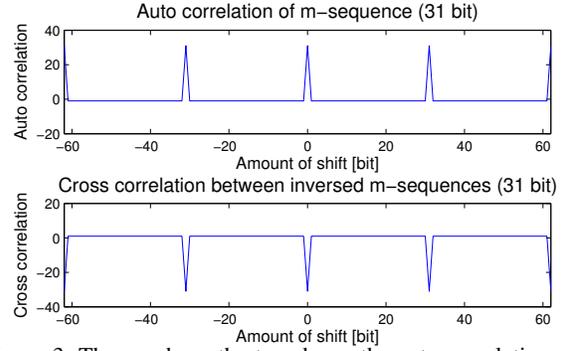


Figure 3. The graph on the top shows the auto correlation of 31 bit m-sequence. The interval between the peaks can be controlled by changing the length of the m-sequence. The bottom graph describes the cross correlation between an m-sequence and its complement.



Figure 4. Extracted layers: layers with codes of particular phases.

attenuation) within the selected depth range and zero (complete attenuation) outside the selected depth range. Hence,

$$\{f, g\} = \arg \min_{f, g} \left( \int_0^T g(t + \phi - \tau)f(t) dt - I_{(\phi_1, \phi_2)} \right)^2 \quad (5)$$

such that:  $f(t), g(t) \in \mathbb{Z}$ .

Where  $I$  is the indicator function. We next present bit shift and phase shift techniques for designing range of interest camera by appropriately choosing  $\beta_i$  and  $\phi_i$ .

**Bit shift technique for high SNR:** One straight forward way to design ROI camera is to choose  $\beta_i$  and  $\phi_i$  as:

$$\begin{aligned} \beta_i &\in B_{\text{int}} = \{1, 1, 1, \dots, 1\}, \\ \phi_i &\in \Phi_{\text{int}} = \{\phi_0, \phi_0 + \Delta\phi, \\ &\quad \phi_0 + 2\Delta\phi, \dots, \phi_0 + (n-1)\Delta\phi\} \end{aligned} \quad (6)$$

where  $\phi_0$  is the phase that corresponds to the starting point of the ROI,  $\Delta\phi$  denotes one bit shift and  $n$  is number of m-sequences required to sweep the ROI. The cross correlation of  $f(t)$  and  $g(t)$  will be in a trapezoidal shape having non-zero value at the ROI distance as shown in figure 5. Each small dotted triangle shape indicates the plot of  $f(t) * \beta_i f(t + \phi_i)$ , where  $*$  denotes cross correlation of two discrete sequences. The slopes at both ends of ROI are undesirable, and will contribute to the optimization cost of equation 3.

**Phase shift technique for sharper side slopes:** The bit shift technique suffers from long slopes. The corresponding distance to this slope is about 3 m when we use 50 MHz as modulation frequency. This value is too large for applications like virtual blue screen.

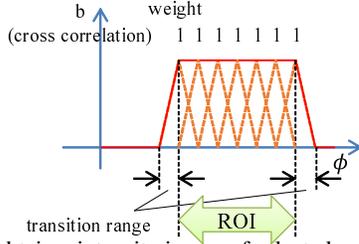


Figure 5. To obtain a intensity image of selected range, we design a trapezoidal shaped cross correlation. The non-zero cross correlation corresponds to ROI and the slopes on both sides are transition ranges.

The VCO (Voltage controlled oscillator) frequency that determines the minimum phase shift amount is very large ( $\sim$ GHz) compared to modulation frequency. Hence, we use the phase shift amount for decreasing the slope width by choosing  $\beta_i$  and  $\phi_i$  as:

$$\begin{aligned} \beta_i \in B_{\text{sharp}} &= \{1, -1\} \\ \phi_i \in \Phi_{\text{sharp}} &= \{\phi_0, \phi_0 + \epsilon\} \end{aligned} \quad (7)$$

where  $\epsilon$  is the width of the desired transition range and must be a positive integral multiple of 96 ps. To realize negative coefficient in  $\beta$ , we use bit inverted version of m-sequence for light modulation code  $g$ . The resulting cross correlation function is shown in figure 6. Although we will see negative sensitivity in the right hand of the graph, we can treat this as 0 by clipping. To extend the width of ROI we can use lower modulation frequency or add further coefficients as follows:

$$\begin{aligned} \beta_i \in B_{\text{sharp}} &= \{1, -1, 2, -2, \dots, n, -n\} \\ \phi_i \in \Phi_{\text{sharp}} &= \{\phi_0, \phi_0 + \epsilon, \\ &\quad \phi_0 + \Delta\phi, \phi_0 + \Delta\phi + \epsilon, \\ &\quad \phi_0 + (n-1)\Delta\phi, \phi_0 + (n-1)\Delta\phi + \epsilon\} \end{aligned}$$

The sharper slope technique has smaller optimization cost compared to former technique, but it suffers from decreased light throughput, and hence lesser SNR.

#### Why is this technique still single image technique?

Though the way we have described our measurements seem to indicate that our depth selective camera is a linear combination of several appropriately shifted m-sequences, note that this linear multiplexing can be achieved directly within a single acquired image. This is because, the modulation frequency is very high (say 60 MHz), compared to the frame rate (say 30 Hz): therefore, 2 million periods of the modulation frequency are being averaged within each measurement. By changing the m-sequence patterns within the exposure duration of a single sensor image, one can capture the linear multiplexed measurements directly within a single captured image. For example, in the sharper side slope technique we are averaging 2 different m-sequences. This can be accomplished by alternating these two m-sequences

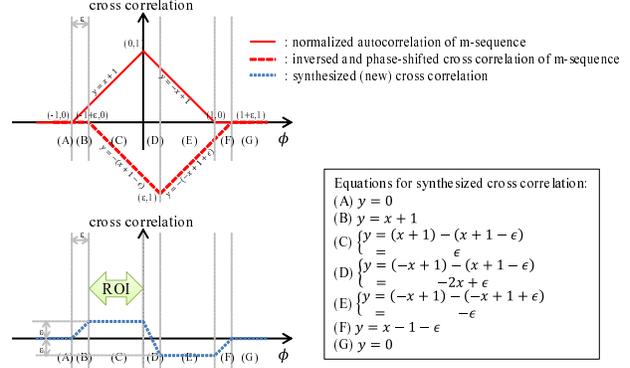


Figure 6. For smaller ROI design, we will need to make a narrower transition range, which is determined by the modulation frequency. To mitigate this problem, we subtract a slightly phase shifted version of cross correlation from the original one. The resulting cross correlation is shown in the second row.

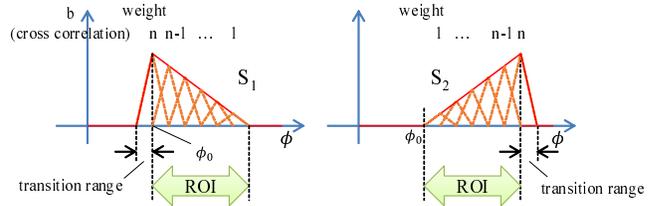


Figure 7. To generate depth map we just need two frames. The cross correlation of each frame is designed to encode depth differently. Using the information from both the frames, we recover the depth map.

at the illumination and the sensor within the exposure duration of the sensor.

**Depth Camera:** Interestingly, we can design  $\beta_i$  and  $\phi_i$  to measure the depth of the scene with just two measurements by designing  $\beta_i$  and  $\phi_i$  as follows:

$$\begin{aligned} \beta_{i,1} \in B_{\text{depth},1} &= \{n, n-1, \dots, 1, 0\} \\ \beta_{i,2} \in B_{\text{depth},2} &= \{0, 1, \dots, n-1, n\} \\ \phi_i \in \Phi_{\text{depth}} &= \{\phi_0, \phi_0 + \Delta\phi, \\ &\quad \phi_0 + 2\Delta\phi, \dots, \phi_0 + n\Delta\phi\} \end{aligned} \quad (8)$$

Using these coefficients, we can get two measurements  $S_1$  and  $S_2$  for  $\beta_{i,1}$  and  $\beta_{i,2}$  respectively. The shape of obtained cross correlation function is shown in figure 7. After getting  $S_1$  and  $S_2$ , we can calculate the depth utilizing the relationship between depth  $z$  and ratio of observed value.

$$z - \phi_0 \propto \frac{S_2}{S_1 + S_2} \quad (9)$$

Similar to the ROI camera for intensity, a transition range will remain at the end of ROI.

## 5. Experiments

In this section, we explain our implementation and present results on several applications for demonstrating direct, programmable, depth-selective photography.

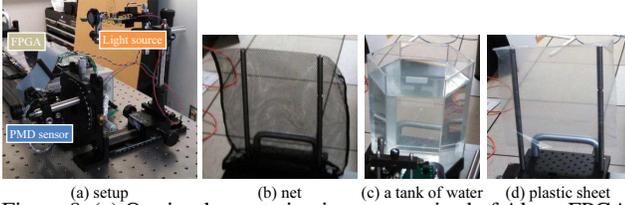


Figure 8. (a) Our implementation is comprised of Altera FPGA development kit DE2-115, infrared laser diode, and PMD 19k-S3. Obstacles used in our experiments are: (b) Net. (c) Milk dropped in a tank of water. (d) Plastic sheet.

### 5.1. Implementation

Our system consists of a PMD sensor, laser diode, and Altera FPGA development kit DE2-11. Figure 8 (a) shows the picture of our setup. FPGA controls various functions of PMD sensor including the reference code  $f(t)$ . Captured image is read out via FPGA and finally stored in a PC. FPGA also controls laser diode driving board by sending illumination code  $g(t)$ . This ensures the frequency and phase synchronization between the light source and sensor. The infrared laser diode is driven by iC-HG from iC-Haus. We employ a 31-bit pseudo random sequence as illumination code  $g(t)$ . The PLL circuit in the FPGA allows us to control the phase of the illumination signal. In our configuration, we can control  $\phi$  in multiples of 96 ps, which corresponds to a light travel distance of about 2.8 cm. Most of the hardware and software design of our system is based on the work by Kadambi *et al.* [15]

### 5.2. Suppressing back-scattering and reflection

In several application scenarios, such as imaging through scattering media (fog etc), imaging through grid/net like occluders, and imaging through glass and windows, the back-scattering and reflections from the front surfaces and media overwhelm and corrupt the sensor measurements. If the depth of the offending occluding materials is different from that of the subjects of interest, then we could use our direct, programmable technique to suppress all the back-scattering and reflection from these occluders, while retaining light contributions from the subjects of interest. We conduct a few experiments to demonstrate our method. Note that traditional computational means to enhance contrast of images in such scenarios (such as dehazing and contrast stretching) can be additionally employed to the captured depth selective camera images to further enhance the performance. Here, we demonstrate the direct improvement in acquired images to illustrate the potential benefits of our method alone.

As described in figure 9, we put obstacles such as a net, a tank of water with slightly dropped milk, and a plastic sheet. Refer figure 8 for details. The presence of these occluders and scattering media significantly affect image contrast as seen in figure 10 and 11.

**Net:** Figure 10 shows the result. In the top most row (intensity images), we can see that the contrast of the image is degraded by back-scattering and reflection from the net. On

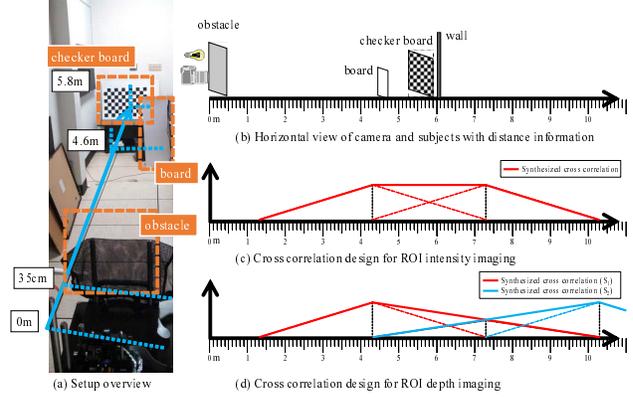


Figure 9. Setup and cross correlation design for suppressing back-scattering and reflection: (a) We placed our camera, an obstacle, a white board, and several subjects including charts linearly. (b) shows their relative placement. (c) We designed a cross correlation function such that the white board and the charts are included in the ROI and the obstacle is excluded. Two phase shifted m-sequences are utilized to synthesize this cross correlation. (d) To obtain depth information, we designed two cross correlation  $sS_1$  and  $S_2$  that have values proportional to the distance inside the ROI, but different slopes and phases.

the other hand, our camera (second row) preserves the contrast. Besides, in the third row, we applied simple contrast enhancement to demonstrate that additional computational techniques can be used to further improve the directly acquired images. Note that the PMD sensor we used is low resolution ( $160 \times 120$ ), even compared to the Kinect sensor ( $640 \times 480$ )– but we still chose to use it because of its programmability. Using higher-resolution sensors will allow us to improve the spatial resolution of all our results.

As shown in the bottom row (plot of depth on the straight lines marked), the depth estimation of normal camera totally fails due to the mixed contributions from objects at different positions. On the contrast, our camera can distinguish the distance of the white board placed on the near side and that of charts at the wall.

**Milk dropped in a tank of water:** This scene consists of two posters placed behind a tank containing diluted milk. The back-scattering from milk reduces image contrast in a normal camera, but our proposed scheme is able to directly obtain higher contrast images (Figure 11). Unfortunately, the fixed pattern noise of the prototype sensor we use causes artifacts in the captured images.

**Plastic sheet:** This is one of the most difficult cases for any method designed to reject back-scattered light as the amount of back-scattering is an order of magnitude more than the direct contribution. Unfortunately, as shown in figure 12, this causes saturation in the sensor response resulting in failure to suppress the out-of-region light paths. One must note that the PMD sensor we used is noisy and has lower dynamic range, even compared to the Kinect sensor – but we still chose to use it because of the programmability it allowed. Using sensors with better SNR and dynamic

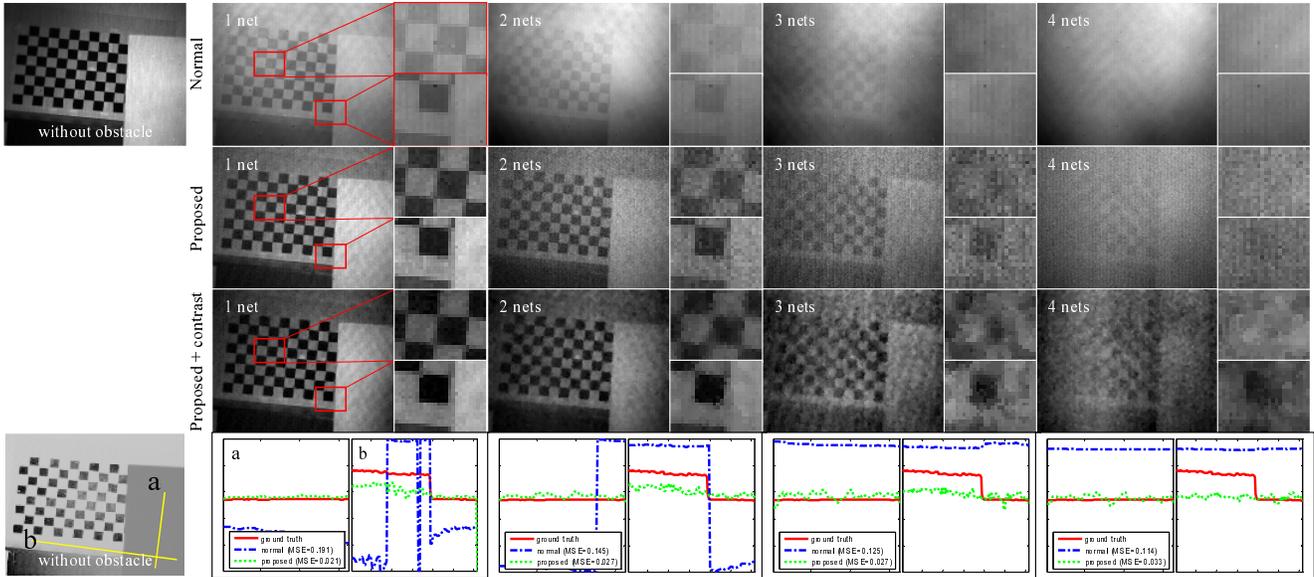


Figure 10. Results for the net scene. The images in the four rows correspond to intensity images with normal camera, intensity images with our camera, contrast enhanced images of our camera, comparison of depth images with ground truth, and our camera respectively. The left most pictures are taken without any obstacles (ground truth). The number of nets increase from left to right. The depth profile of lines ‘a’ and ‘b’ is shown in the bottom most row. The depth value is normalized to 0–1 when calculating MSE (mean squared error) value.

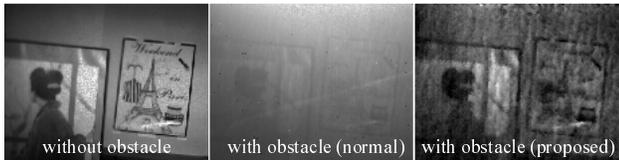


Figure 11. **Imaging through milk:** (Left) Scene containing two posters is shown. (Middle) Normal camera image has low contrast due to scattering caused by diluted milk in front of subject. (Right) Proposed camera directly acquires better contrast image. The fixed pattern noise of the sensor, unfortunately causes artifacts in our method.

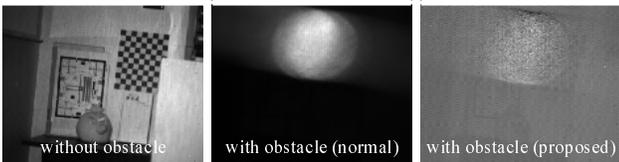


Figure 12. **Failure Case:** The scene contains backward-reflection from a plastic sheet that is an order of magnitude greater than the light from depth range of interest and this results in sensor saturation and consequently a failure of our method to suppress the backward reflections.

range will perhaps allow us to image even such challenging scenarios.

### 5.3. Illustration of phase shift technique

As illustrated in figure 13, we put an obstacle and subject in a closer place. The distance between them is around 1.5 m. Recall that the code modulation frequency of our setup is 50 MHz which turns into translation range of 3 m (figure 5, 7). This means that we cannot separate the responses from the obstacle and the subject completely. However, by designing a sharper slope as described in the bottom chart of

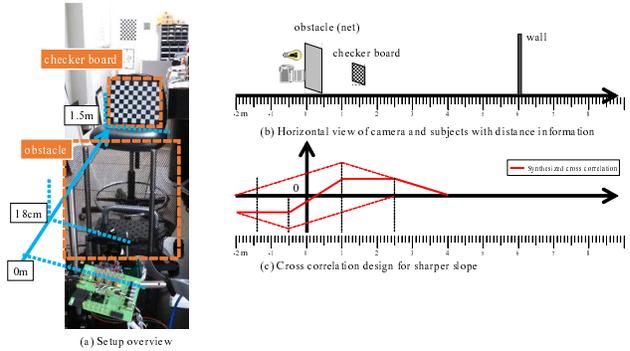


Figure 13. Setup and cross correlation design for sharper slope: (a) Our camera detects the checker board at a distance of 1.5 m. Between them, a net is placed as an obstacle. 1.5 m is smaller than the translation range of 3 m which determined by the current modulation frequency, 50 MHz. (b) The relationship of their positions. (c) We designed a cross correlation which include sharper slope.

figure 13, we can remove the effect of backward reflection. Figure 14 shows the results for various scenarios.

### 5.4. Virtual blue screening

To provide a solution to applications that require background suppression, we propose virtual blue screen that finds applications in teleconferencing. This technique is intended to hide objects behind the subject, for example, a messy room while talking over Skype. The overview of the setup and the cross correlation design are shown in figure 15. Note that we used the phase shift technique here again, because the range we would like to control in the indoor is usually less than 3 m determined by the modulation frequency.

Figure 16 demonstrates the virtual blue screen effect. We can notice that the background objects including two

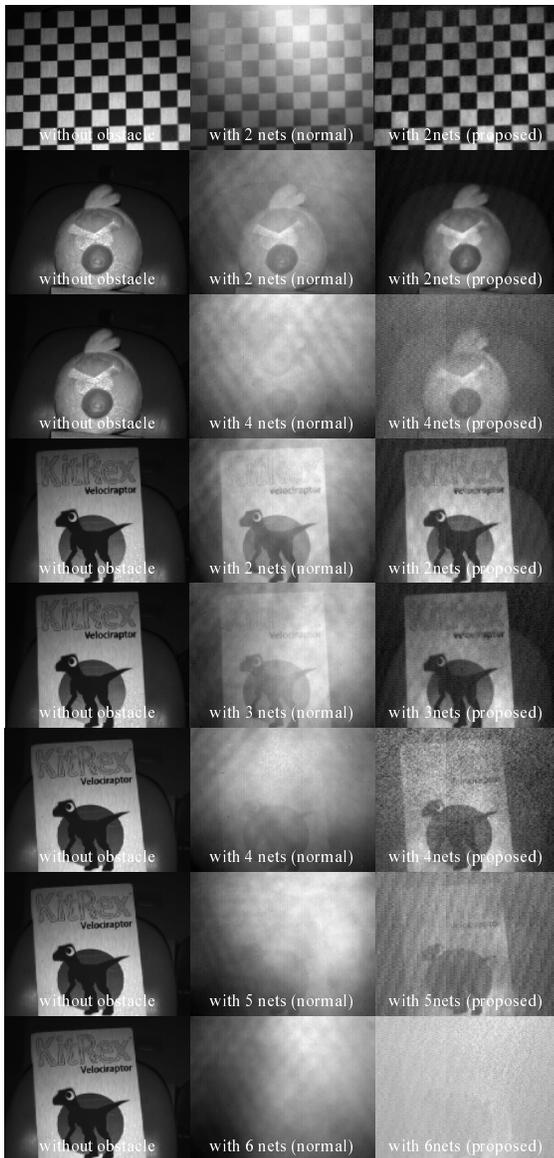


Figure 14. Results for sharp slope technique. From left to right, the pictures are image taken by normal camera without any obstacle (ground truth), normal camera image with obstacle, and backward reflection suppressing result by our camera.

checker boards on the wall are disappeared. This background can be potentially replaced with visually appealing background.

### 5.5. Artifacts

Many of the real experimental results shown in this paper have artifacts and are noisy. The primary reason is the low SNR and dynamic range of the PMD sensor we used in our implementation, primarily because it provided us with the programmability that we needed unlike other higher resolution, higher quality sensors such as Kinect 2.0 etc. In particular, we noticed that there was significant fixed pattern noise on the sensor and this fixed pattern noise is clearly visible in many of our real results. Simple post-processing

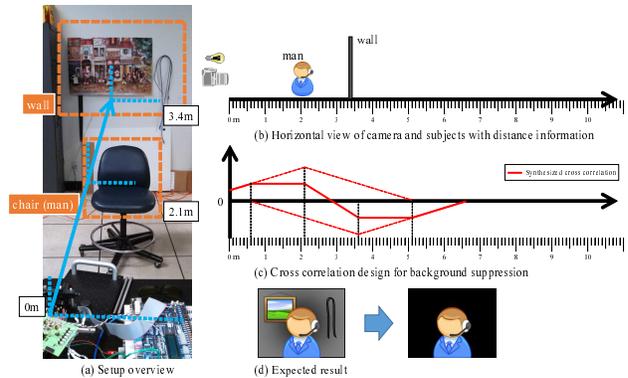


Figure 15. Setup and cross correlation design for background suppression: (a) We put a random picture and wire codes hanging on the wall as an example of unwanted object to be seen behind the subject. (b) The relationship of their positions. (c) In designing cross correlation, the sharper slope technique is used to suppress the sensitivity at the distance of the wall. (d) Expected result is illustrated. Note that the results are grayscale because our sensor is a monochrome sensor.

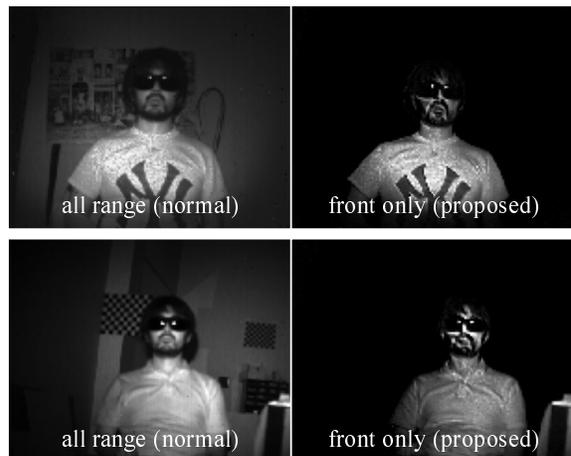


Figure 16. Virtual blue screening result. The left picture is captured by normal intensity camera. The right picture is the result of our background suppression. The background objects including random things on the wall are gone.

such as median filtering significantly reduces the effect of this fixed pattern noise – but ultimately, the best solution to avoid these artifacts is indeed to implement the technique on higher resolution and higher quality sensors. We plan to do this as soon as API interfaces to control the illumination and sensor codes for these higher resolution sensors become available.

## 6. Conclusions

In this paper, we proposed a technique to selectively image specified range of distance by designing cross correlation of ToF camera. We formulated the problem mathematically and showed that it is NP hard. We have provided heuristics to solve the problem and showed the effectiveness of our solution on several applications.

## References

- [1] I. M. Baker, S. S. Duncan, and J. W. Copley. A low-noise laser-gated imaging system for long-range target identification. In *Proc. SPIE 5406, Infrared Technology and Applications XXX*, volume 5406, pages 133–144, 2004. [2](#)
- [2] P. Barnum, T. Kanade, and S. Narasimhan. Spatio-Temporal Frequency Analysis for Removing Rain and Snow from Videos. *Proceedings of the First International Workshop on Photometric Analysis For Computer Vision - PACV 2007*, 2007. [3](#)
- [3] D. Bonnier and V. Larochelle. A range-gated active imaging system for search-and-rescue and surveillance operations. In *Proc. SPIE 2744, Infrared Technology and Applications XXII*, volume 2744, pages 134–145, 1995. [2](#)
- [4] T. F. Braeunig. Body-mounted video game exercise device, 1990. [1](#)
- [5] B. Büttgen and P. Seitz. Robust optical time-of-flight range imaging based on smart pixel structures. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 55(6):1512–1525, 2008. [3](#)
- [6] M. Cohn and A. Lempel. On fast M-sequence transforms (Corresp.). *IEEE Transactions on Information Theory*, 23(1):135–137, Jan. 1977. [4](#)
- [7] R. M. Conroy, A. a. Dorrington, R. Künnemeyer, and M. J. Cree. Range imager performance comparison in homodyne and heterodyne operating modes. *Proc. SPIE 7239, Three-Dimensional Imaging Metrology*, 723905 (January 19, 2009), 7239:723905–723905–10, 2009. [3](#)
- [8] B. B. Das, K. M. Yoo, and R. R. Alfano. Ultrafast time-gated imaging in thick tissues: a step toward optical mammography. *Optics letters*, 18(13):1092, 1993. [2](#)
- [9] O. David, N. S. Kopeika, and B. Weizer. Range gated active night vision system for automobiles. *Applied optics*, 45(28):7248–7254, 2006. [2](#)
- [10] K. Garg and S. Nayar. Detection and removal of rain from videos. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 1, 2004. [2](#)
- [11] Y. Grauer and E. Sonn. Active Gated Imaging for Automotive Safety Applications. In *Proc. SPIE 9407, Video Surveillance and Transportation Imaging Applications 2015*, volume 9407, pages 1–18, 2015. [2](#)
- [12] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2011. [3](#)
- [13] F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Optics Express*, 22(21):26338, 2014. [3](#)
- [14] S. Izadi, A. Davison, A. Fitzgibbon, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, and D. Freeman. Kinect Fusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*, page 559, 2011. [1](#)
- [15] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded Time of Flight Cameras : Sparse Deconvolution to Address Multipath Interference and Recover Time Profiles. *ACM Transactions on Graphics*, 32(6):1–10, 2013. [3](#), [6](#)
- [16] R. Lange and P. Seitz. Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*, 37(3):390–397, 2001. [3](#)
- [17] R. Lange, P. Seitz, A. Biber, and S. Lauxtermann. Demodulation pixels in CCD and CMOS technologies for time-of-flight ranging. *Proceedings of SPIE*, 3965:177–188, 2000. [3](#)
- [18] M. O’Toole, R. Raskar, and K. N. Kutulakos. Primal-dual coding to probe light transport, 2012. [3](#)
- [19] Y. Schechner, S. Narasimhan, and S. Nayar. Instant dehazing of images using polarization. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 1:325–332, 2001. [3](#)
- [20] R. Schwarte. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD). *Proceedings of SPIE*, 3100:245–253, 1997. [3](#)
- [21] S. Shwartz, E. Namer, and Y. Y. Schechner. Blind haze separation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:1984–1991, 2006. [3](#)
- [22] K. J. Snell, A. Parent, M. Levesque, and P. Galarneau. Active range-gated near-IR TV system for all-weather surveillance. In *Proc. SPIE 2935, Surveillance and Assessment Technologies for Law Enforcement*, volume 2935, pages 171–181, 1997. [2](#)
- [23] J. Suarez and R. R. Murphy. Hand gesture recognition with depth images: A review. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 411–417, 2012. [1](#)
- [24] R. Tamburo, a. Chugh, E. Nurvithadhi, M. Chen, a. Rowe, T. Kanade, and S. G. Narasimhan. Programmable Automotive Headlights. *European Conference on Computer Vision (submitted)*, pages 1–16, 2014. [2](#)
- [25] M. Tobias, K. Holger, F. Jochen, A. Martin, and L. Robert. Robust 3D Measurement with PMD Sensors. *Range Imaging Day, Zürich*, 7:8, 2005. [3](#)
- [26] R. Z. Whyte, A. D. Payne, A. a. Dorrington, and M. J. Cree. Multiple range imaging camera operation with minimal performance impact. *Image (Rochester, N.Y.)*, 7538:753801–753801–10, 2010. [3](#)
- [27] W. Xin-Wei, Z. Yan, F. Song-Tao, H. Jun, and L. Yu-Liang. Range-Gated Laser Stroboscopic Imaging for Night Remote Surveillance. *Chinese Physics Letters*, 27(9):094203, 2010. [2](#)