Stochastic Exposure Coding for Handling Multi-ToF-Camera Interference

Jongho Lee
University of Wisconsin-Madison
jongho@cs.wisc.edu

Mohit Gupta
University of Wisconsin-Madison
mohitg@cs.wisc.edu

Abstract

As continuous-wave time-of-flight (C-ToF) cameras become popular in 3D imaging applications, they need to contend with the problem of multi-camera interference (MCI). In a multi-camera environment, a ToF camera may receive light from the sources of other cameras, resulting in large depth errors. In this paper, we propose stochastic exposure coding (SEC), a novel approach for mitigating. SEC involves dividing a camera's integration time into multiple slots, and switching the camera off and on stochastically during each slot. This approach has two benefits. First, by appropriately choosing the on probability for each slot, the camera can effectively filter out both the AC and DC components of interfering signals, thereby mitigating depth errors while also maintaining high signal-to-noise ratio. This enables high accuracy depth recovery with low power consumption. Second, this approach can be implemented without modifying the C-ToF camera's coding functions, and thus, can be used with a wide range of cameras with minimal changes. We demonstrate the performance benefits of SEC with theoretical analysis, simulations and real experiments, across a wide range of imaging scenarios.

1. Introduction

Time-of-flight (ToF) cameras are fast becoming the method of choice in various 3D imaging applications, such as 3D mapping [8, 13], human-machine interaction [5], augmented reality [11] and robot navigation [17]. ToF cameras have compact form-factors and low computational complexity, resulting in emergence of several commodity ToF cameras [2, 1]. As these cameras become ubiquitous in mobile devices and cell-phones, they will face an important problem: multi-camera interference (MCI). This is especially critical for continuous wave ToF (C-ToF) imaging, where the light source emits light continuously. When several C-ToF cameras capture the same scene concurrently, each sensor may receive light from the sources of other cameras. This interfering signal prevents correct depth estimation, resulting in potentially large, structured errors.

One way to address MCI is to use orthogonal coding functions for different C-ToF cameras, for example, sinusoids of different frequencies or phases [23, 18, 16], or pseudo-random functions [6, 7, 10]. These approaches, while theoretically capable of mitigating interference, have a practical limitation. The intensity of light emitted by a ToF camera's source is positive, with both a constant (DC) and an oscillating (AC) component; the depth information is encoded in the time-shift of the AC component. Although the orthogonal-coding approaches can remove the AC interference, the DC interference remains. The DC interference acts as additional ambient light, resulting in higher photon noise. As the number of interfering cameras increases, the signal-to-noise ratio (SNR) can degrade considerably, making it challenging to recover meaningful information.

We propose a novel MCI reduction technique with the goal of mitigating *both* DC and AC interference. Our approach is based on time-division multiple access (TDMA), a widely used scheme for facilitating multi-user access of shared communication channels. In TDMA, a single, shared communication channel is divided into multiple time slots, one slot assigned to each user [21]. In order to prevent interference, the timing across different users must be synchronized, which is done by a central authority, e.g., base stations. Applying TDMA directly for addressing MCI will require high-speed temporal synchronization of different cameras, which, unfortunately, is challenging [7].

Stochastic exposure coding: Is it possible to implement a TDMA-like approach without synchronization? Our key idea is to leverage stochasticity to avoid explicit synchronization. The proposed approach, called stochastic exposure coding (SEC), divides the total exposure time of each camera into multiple slots. In each slot, the camera and the source are turned on with a certain probability p_{ON}. By design, if a slot doesn't have a clash, i.e., only one camera is active during that slot, *both* DC and AC interference are avoided since the camera receives light only from its own source. Since the approach is stochastic, without explicit synchronization, there may still be clashes. We design a simple, light-weight clash-check algorithm to identify and discard clash-slots so they do not affect depth estimation. [1]

[1] This approach is similar to random-access protocols in communication such as ALOHA [3] and CSMA [14] in that packets are sent randomly. However, while communication protocols need to re-send packets whenever collision happens, in our case, we can simply discard clashed slots. This is because in communication, each packet has unique information, whereas in our case, all slots have the same depth information.

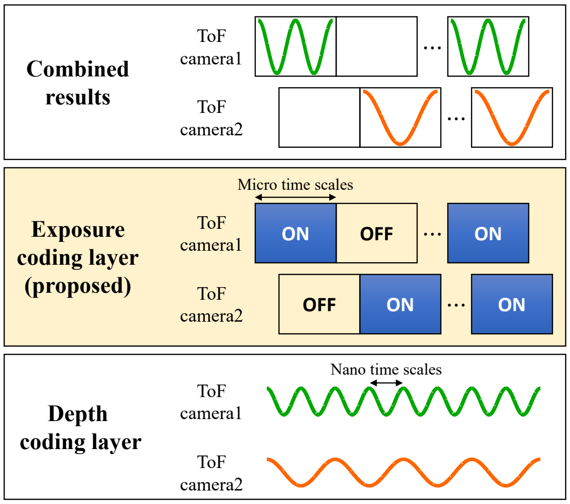


Figure 1. **Layered C-ToF coding.** The proposed approach operates in the exposure coding layer, where the camera and the source are modulated at micro/millisecond time scales. In contrast, existing MCI reduction approaches operate in the lower depth coding layer, where modulation is performed at nanosecond time scales.

What is the optimal p_{ON}? This is a critical question that must be addressed for the proposed approach to be successful. A high p_{ON} will increase the likelihood of clashes (multiple simultaneously active cameras), resulting in interference and depth errors. On the other hand, if p_{ON} is too low, although the clashes are avoided, the cameras are inactive during most of the integration time, and thus, don't receive sufficient signal. We perform a detailed theoretical analysis, and determine the optimal p_{ON}, given system constraints and the number of interfering cameras. This enables each source to send light sufficiently sparsely to mitigate interference without synchronization, while maintaining a high SNR, for a fixed time and power budget.

Layered view of C-ToF coding: A key benefit of the proposed SEC approach is that it does not need to modify the C-ToF camera's coding functions, and thus, can be implemented without extensive hardware modifications. SEC can be implemented by rapidly switching the camera off and on during the integration time, in a way reminiscent of temporal exposure coding for motion deblurring [19]. This creates a layered view of C-ToF camera coding, as shown in Figure 1. Existing approaches for MCI reduction operate in the *depth coding layer* since they change the camera's coding functions at nanosecond time scales. In contrast, SEC operates at a higher *exposure coding layer* by modulating the camera and source at micro/millisecond scales.

Practical implications: SEC and existing MCI reduction approaches can be used in a complementary manner because they operate in different layers. We show, via theoretical analysis, simulations and hardware experiments that such combined multi-layer coding approaches significantly outperform existing methods. The proposed approaches reduce both DC and AC interference, making it possible to achieve high SNR while consuming low power. Because they require minimal modifications to existing C-ToF systems, these approaches are broadly applicable for 3D imaging in low-complexity, power-constrained mobile devices.

2. Related Work

Most existing approaches for MCI reduction rely on orthogonal functions, such as sinusoids of different modulation frequencies for different cameras [20], and pseudo-noise (PN) sequences [6, 7]. Other approaches divide the total integration time into multiple time slots and randomly assign one of predetermined phases to each slot [23, 18, 16]. While all these approaches reduce only AC interference, our goal is to design methods that mitigate both AC and DC interference. Another recent approach for handling MCI is to project light only along a planar sheet which is scanned over the scene. Since only a portion of the scene is illuminated at a time, the chance of interference by other cameras is reduced [4]. Although this approach can also reduce DC interference, it requires mechanical scanning. In contrast, our approach can be implemented without moving parts.

3. Mathematical Preliminaries

C-ToF Image Formation Model: A C-ToF camera consists of a (typically co-located) camera and a light source [15]. The intensity of the light source is temporally modulated as a periodic function $M(t)$, ($M(t) \geq 0$) with period T_0. The light emitted by the source travels to the scene of interest, and is reflected back toward the camera. The radiance of the reflected light incident on a sensor pixel $\mathbf{p}$ is a time-shifted and scaled version of $M(t)$:

$$R(\mathbf{p}; t) = \alpha P_s M\left(t - \frac{2d}{c}\right), \tag{1}$$

where d is the distance between the camera and the scene point imaged at $\mathbf{p}$, c is the speed of light. P_s is average power of the light source with an assumption of $\frac{1}{T_0}\int_{T_0} M(t)\,dt = 1$. α is a scene-dependent scale factor that contains scene albedo, reflectance properties and light fall-off. The camera then electronically computes the correlation between $R(\mathbf{p}; t)$ and a periodic demodulation function $D(t)$ $(0 \leq D(t) \leq 1)$[2] with the same frequency as $M(t)$. The intensity value $C(\mathbf{p}; d)$ measured at pixel $\mathbf{p}$ is given as the correlation between $R(\mathbf{p}; t)$ and $D(t)$:

$$C(\mathbf{p}; d) = s \int_T \left(R(t; d) + P_a\right) D(t)\,dt, \tag{2}$$

[2] Several C-ToF camera architectures [15, 6] use a bipolar demodulation functions ($-1 \leq D(t) \leq 1$). For ease of analysis, we consider unipolar $D(t)$ $(0 \leq D(t) \leq 1)$. All the results and analysis in the paper can be generalized to bipolar $D(t)$.

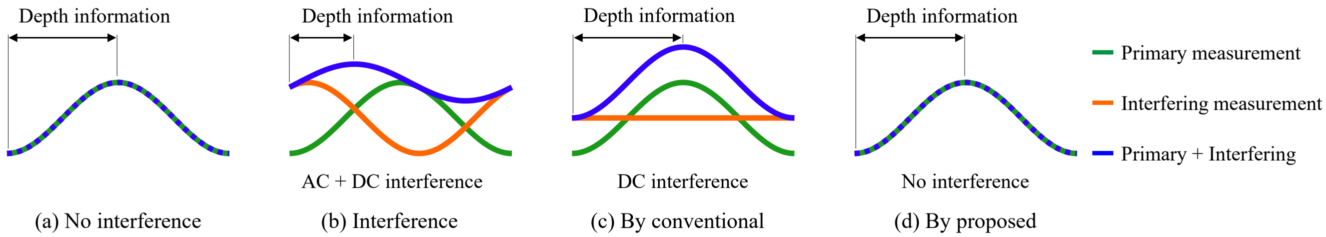


Figure 2. **Multi-camera interference and interference reduction in C-ToF imaging.** (a) In C-ToF imaging, depths are recovered from the phases of the measured waveforms. (b) If there are multiple cameras, interfering sources corrupt the measured waveforms, resulting in systematic depth errors. (c) Conventional MCI reduction approaches reduce systematic errors by removing AC interference, but DC interference remains, resulting in lower SNR and random depth errors due to higher photon noise. (d) Our approach mitigates both AC and DC interference, thus reducing both systematic and random depth errors.

where s is a camera-dependent scale factor encapsulating sensor gain and sensitivity, T is the total integration time, and P_a is average power of ambient light incident on the scene (e.g., due to sunlight in outdoor operation). In order to estimate the scene depths, several (≥ 3) different $C(\mathbf{p}; d)$ values are measured, by using different pairs of modulation and demodulation functions [15].

3.1. Multi-Camera Interference in C-ToF Imaging

Consider a scenario where multiple C-ToF cameras are simultaneously illuminating and imaging a scene point. The total intensity measured by one of the cameras (referred to as the primary camera) is given by:

$$C_{mult}(d) = C(d) + \underbrace{\sum_{n=1}^{N} C_n(d)}_{\text{multi-camera interference}}, \tag{3}$$

where N is the number of interfering cameras, $C(d)$ is the intensity measured by the primary camera due to its own source (Eq. 2), and $C_n(d) = s \int_T R_n(t) D(t) dt$ is the measured intensity due to the n^{th} source. $R_n(t)$ is the radiance received by the primary camera due to light emitted by the n^{th} source. We drop the argument $\mathbf{p}$ for brevity. The summation term in Eq. 3 corrupts the true correlation value $C(d)$, thus resulting in erroneous depth estimates.

Example with sinusoid coding: In a C-ToF camera with sinusoid coding, both modulation $M(t)$ and demodulation $D(t)$ functions are sinusoids of the same frequency (homodyne). The camera takes $K \geq 3$ intensity measurements (Eq. 2). Each measurement $C^k(d)$, $k \in \{1, \ldots, K\}$ is taken by shifting the demodulation function $D(t)$ by a different amount ψ_k, while $M(t)$ remains fixed. For example, if $K = 4$, $[\psi_1, \psi_2, \psi_3, \psi_4] = [0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}]$. The set of measurements $\left\{C^k(d)\right\}$, $k \in \{1, \ldots, K\}$ is defined as the measurement waveform. For sinusoid coding, the measurement waveform is a sinusoid as a function of the shift ψ_k, as shown in Fig. 2 (a). Let ϕ be the phase of the measurements waveform sinusoid. Scene depth d is proportional to ϕ, and can be recovered by simple, analytic expressions [12].

If multiple cameras simultaneously image a scene point, a camera receives light from the interfering sources as well as its own source. Assuming all the sources use sinusoids of the same frequency, the intensities $\left\{C_n^k\right\}$, $k \in \{1, \ldots, K\}$ measured by the camera due to the n^{th} source *also form a sinusoid*. The total measurement $\left\{C_{mult}^k\right\}$, $k \in \{1, \ldots, K\}$ (Eq. 3) is the sum of these individual sinusoids, and thus, also forms a sinusoid. This is shown in Fig. 2 (b). However, since the phases ϕ_n of the individual sinusoids (one due to each interfering source) may be different, the phase of the total measurement waveform may differ from the true phase, resulting in *systematic*, potentially large depth errors.

3.2. Orthogonal Coding for Mitigating Interference

One way to mitigate multi-camera interference (MCI) is to ensure that the intensities $\left\{C_n^k\right\}$, $k \in \{1, \ldots, K\}$ due to an interfering source form a constant waveform, i.e., $C_n^k = C_n$, $\forall k$. For example, in sinusoid coding, this can be achieved by assigning a different modulation frequency to each camera [20]. [3] As a result, the total measurement waveform $\left\{C_{mult}^k\right\}$, $k \in \{1, \ldots, K\}$ *has the same phase* as the sinusoid due to the primary source. This is because the interfering components are constant waveforms, and thus do not alter the phase, thereby preventing systematic depth errors. This is shown in Figure 2 (c).

We call this *AC-Orthogonal (ACO)* approach, since it reduces the interference to constant waveforms by removing the AC component. However, the offset (DC-component) of the total waveform still increases, as shown in Figure 2 (c). The extra offset acts as additional ambient light, and thus lowers the SNR of the estimated depths due to increased shot noise [23]. [4] For example, the depth standard deviation for a 4-tap sinusoid-based ACO method is given as:

$$\sigma_{ACO} = \frac{c}{2\sqrt{2}\pi f_0 \sqrt{T}} \frac{\sqrt{e_s + e_a + N e_i}}{e_s}, \tag{4}$$

where f_0 is the modulation frequency, T is the total cap-

[3] Sinusoids of different frequencies are orthogonal functions, i.e., their correlation is zero, or a constant if the sinusoids have a non-zero DC offset.

[4] With bipolar demodulation functions, although the DC-offset is removed, the shot noise still increases. See technical report for a discussion.

ture time for each measurement, and c is the light speed. $e_s = s\alpha P_s$, $e_i = s\alpha_i P_s$ and $e_a = sP_a$ are the average number of signal photons (due to the primary camera's own source), interfering photons (due to an interfering source), ambient photons (due to ambient source), respectively, incident on the pixel per unit time. Without loss of generality, we assume that e_i is the same for all interfering cameras. See technical report for derivation of Eq. 4.

Although an ACO approach prevents systematic errors due to MCI, random errors due to photon noise increase as the number of interfering cameras increases (Eq. 4). This is because each interfering source has a non-zero DC component, contributing additional photon noise to the intensity measurements. Is it possible to design a *DC-Orthogonal* (DCO) approach, that removes *both* the AC and DC components of the interference, as shown in Figure 2 (d)?

4. Stochastic Exposure Coding

In this section, we describe the proposed stochastic exposure coding (SEC) technique. SEC is a DC-orthogonal approach since it can mitigate both DC and AC interference. SEC is based on the principle of time-division multiple access (TDMA) used in communication networks to facilitate simultaneous multi-user access to a shared channel. Consider a scenario where multiple ToF cameras are simultaneously imaging the same scene. One way to prevent interference is to divide the capture time into multiple slots, and ensure that exactly one camera (and its source) is on during any given slot. However, assigning cameras to slots deterministically requires temporal synchronization, which may be challenging, perhaps even infeasible, especially in uncontrolled consumer applications.

The *key idea* behind the SEC is that by performing the slot assignment *stochastically*, interference can be prevented without synchronization. SEC can be considered a stochastic version of the TDMA described above, where in each slot, every camera is turned on with a probability p. The on-off decision is made *independently* for each slot, for every camera, without synchronization. If a slot doesn't produce a clash, both DC and AC interference are avoided since the camera receives light only from its own source, as shown in Figure 3. Since the approach is stochastic, a slot may have clashes, which can be identified and discarded with a simple clash-check algorithm (Section 4.2).

4.1. Optimal Slot ON Probability

The performance of the SEC is determined by the slot ON probability p (we will use p instead of p_{ON} for brevity). If p is high, each camera utilizes a larger fraction of the capture time, but may lead to more clashes. On the other hand, for a low p, clashes may be minimized, but the cameras incur a longer 'dead time' during which they are neither emitting light, nor capturing measurements. Thus, a natural question is: What is the optimal p? To address this, we express the depth standard deviation of the SEC in terms of p.

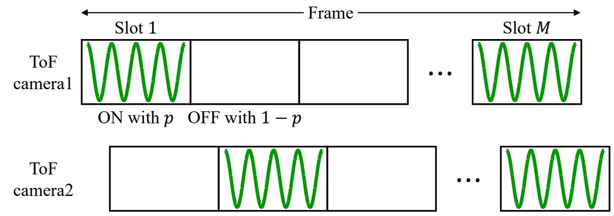


Figure 3. **Concept of SEC.** A frame, the most basic unit to estimate the depth, is divided into M number of slots. Each slot is activated with a probability p. A depth value is estimated from non-clashed ON (activated) slots.

Depth standard deviation of SEC: Consider a scene being imaged by $N+1$ C-ToF cameras. For ease of analysis, we assume the cameras are identical. The capture time of each camera is divided into slots of the same duration. For each camera, it is turned on with a probability p in every slot. In general, the boundaries of the slots may not be aligned across cameras. Therefore, any given slot of a camera will overlap with two slots of another camera. Thus, the probability p_{noclsh} that a given slot does not produce a clash, i.e., only one camera is active during that slot, is:

$$p_{noclsh} = p\,(1-p)^{2N}\,. \tag{5}$$

Assuming we can identify all the non-clash slots, the effective exposure time for each camera, on an average, is $T\,p_{noclsh}$, where T is the total capture time. In order to compensate for the reduced exposure time, we assume that the peak power of the source can be amplified. Let A be the source peak power amplification. Theoretically, A should be $1/p$, so the total energy used during the capture time remains constant. Practically, however, A is limited by device constraints. Thus, $A = \min(1/p, A_0)$, where A_0 is the upper bound of A determined by physical constraints.

Given the effective exposure time $T\,p_{noclsh}$ and source power amplification A, the depth standard deviation of SEC can be derived from Eq. 4:

$$\sigma_{SEC} = \frac{c}{2\sqrt{2}\pi f_0 \sqrt{T p_{noclsh}}} \frac{\sqrt{Ae_s + e_a}}{Ae_s}, \tag{6}$$

where $A = \min(1/p, A_0)$ and $p_{noclsh} = p\,(1-p)^{2N}$.[5] The optimal ON probability for SEC p_{SEC} is defined as:

$$\boxed{p_{SEC} = \arg\min_p \sigma_{SEC} = \min\left(\frac{1}{2N+1}, \frac{1}{A_0}\right).} \tag{7}$$

See technical report for a derivation. As the number of interfering cameras N increases, the optimal ON probability

[5] Strictly speaking, randomness due to slot ON probability can influence the depth standard deviations. However, in practice, the effect of randomness is relatively small if sufficient number of slots are used.

decreases so that the number of clashes remains low. If p is too small or large, the optimal SNR cannot be achieved since the effective integration time is reduced.

4.2. Clash Check and Depth Estimation in SEC

Since SEC is a stochastic, asynchronized approach, a fraction of the slots in each frame may still have clashes. These clash slots need to be identified and discarded so that they do not affect the depth computations. Our clash check algorithm is based on the following, simple intuition: In a clashed slot, the camera receives light from multiple sources. Therefore, the total received intensity in that slot is higher as compared to no-clash slots, with high probability. Therefore, we compare the sum of all the correlation values $o = \sum_k C_k$ in each slot to a threshold. If o is larger, the corresponding slot is discarded. Finally, we compute a depth value d_m ($m \in \{1, ..., M_{noclsh}\}$ for each non-clash slot, and the final depth value d for each *frame* is estimated by averaging d_m. See the technical report for details.

4.3. Practical Considerations and Limitations

Being a DC-orthogonal approach, SEC achieves higher SNR than ACO (see Section 6 for details). On the other hand, SEC has stronger requirements: (a) it requires higher source peak power (for the same total energy) as compared to ACO, and (b) it needs to capture more data (multiple slots per frame). Fortunately, as we show below, there are relatively small upper bounds on these requirements.

Required source peak power amplification: Since the effective integration time of SEC is shorter than ACO, the SNR of SEC can be smaller than ACO if the source peak power amplification A is not sufficiently large. The required A for SEC to perform better than ACO in terms of SNR can be estimated from $\sigma_{SEC} \leq \sigma_{ACO}$:

$$\frac{1}{\sqrt{p_{noclsh}}} \frac{\sqrt{A + r_a}}{A} \leq \sqrt{1 + r_a + N r_i}, \tag{8}$$

where $r_a = e_a/e_s$ and $r_i = e_i/e_s$ are relative ambient light strength and relative interfering light source strength, respectively. Figure 4 shows the required peak power amplification A over different number of interfering cameras N at different ambient light strengths. Although the required A increases with N, it eventually converges, as stated in the following result (see technical report for a proof):

Result 1. *If the source peak power amplification of SEC is larger than* $\left(e + \sqrt{e\left(e + 2r_a r_i\right)}\right)/r_i$*, the depth standard deviation of SEC is always lower than ACO regardless of the number of interfering cameras. For example, the required* $A \approx 6.3$ *when* $r_a = r_i = 1$.

Practicality of achieving high peak power: Two factors should be considered regarding the practicality of increasing source peak power. First, in power-constrained devices

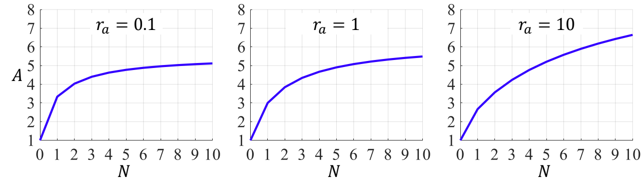


Figure 4. **Required source peak power amplification for SEC.** The required source peak power A increases with the number of interfering cameras N, but eventually converges, for various relative ambient light strengths r_a.

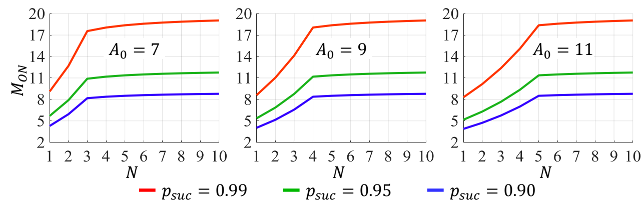


Figure 5. **Required number of ON slots for SEC.** More number of ON slots M_{ON} is required if the number interfering cameras N increases over various allowable peak power amplification A_0. However, the required M_{ON} eventually converges.

(e.g., cell-phones), in order to minimize total energy consumption, it may be desirable to operate the light source with low average power despite availability of higher peak power. Second, recent studies have shown the possibility of driving low-cost sources typically used in C-ToF cameras (e.g., laser diodes and LEDs) with high instantaneous peak power [22]. For example, a laser diode emitting at NIR (830 nm) with 1.5 W optical output power was successfully overdriven up to about 25 W [22].

Required number of slots: For correct depth estimation in SEC, we need at least one non-clashed ON slot. Let p_{suc} be the probability of getting at least one non-clashed ON slots during a frame. Then, the number of ON slots M_{ON} that a camera would need to capture per frame increases with N, but, is eventually bounded, as stated in the following result:

Result 2. *The required number of ON slots* M_{ON} *converges to* $e\left(z^2/2 + 1 - z\sqrt{z^2/4 + 1}\right)$ *regardless of the number of interfering cameras, where* z *is the z-score value, and is a function of* p_{suc}*. For example, when* $p_{suc} = 0.9$*, the required* M_{ON} *is upper bounded by 9.1.*

See supplementary report for a proof. Figure 5 shows M_{ON} over the number of interfering cameras N with various desired success probability p_{suc} and different allowable source peak power amplification A_0. M_{ON} increases with N, but *converges* as N increases. The total number of slots in a frame $M = M_{ON}/p_{SEC}$ can be large and affect the frame rate. However, the more pertinent factor that limits the frame rate is M_{ON} (the number of on slots), which is relatively small, thus making it possible to achieve sufficiently high frame rate for capturing dynamic scenes. See technical report for a detailed discussion and analysis.

5. Multi-Layer Coding for Mitigating MCI

The proposed SEC creates a layered view of C-ToF camera coding, as shown in Figure 1. Most existing approaches for MCI reduction operate in the bottom *depth coding layer* since they change the camera's coding functions at nanosecond time scales. In contrast, SEC operates at a higher *exposure coding layer* by modulating the camera and source at micro/millisecond time scales. Since SEC and conventional ACO techniques operate in different layers, these are orthogonal to each other, and, can be used in a complementary manner to combine the benefits of both. For example, it is possible to use sinusoid coding with different modulation frequencies for different cameras, while also using SEC. In such a multi-layer integrated approach (CMB), it is no longer necessary to discard the clashed slots since they do not introduce depth errors. This makes repeated clash check unnecessary, leading to simpler depth estimation and an efficient frame structure.

Depth standard deviation of CMB: Depth standard deviation of CMB σ_{CMB} can be easily derived from Eq. 4:

$$\sigma_{CMB} = \frac{c}{2\sqrt{2}\pi f_0 \sqrt{Tp}} \frac{\sqrt{Ae_s + e_a + NpAe_i}}{Ae_s}, \tag{9}$$

where $A = \min\left(\frac{1}{p}, A_0\right)$.

Optimal slot ON probability: The optimal slot ON probability for CMB p_{CMB} is defined as p minimizing Eq. 9:

$$p_{CMB} = \arg\min_p \sigma_{CMB} = \frac{1}{A_0}. \tag{10}$$

Note that p_{CMB} is independent of N. For derivation and depth estimation algorithm, see technical report.

6. Theoretical Performance Comparisons

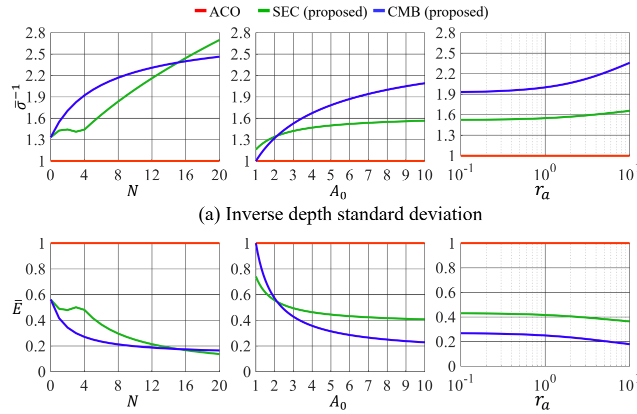
We present theoretical comparisons between ACO, SEC and CMB in terms of 1) depth standard deviation at the same energy consumption and 2) required energy to achieve the same depth standard deviation. All comparisons are *relative* to an ideal ACO. We define the normalized inverse depth standard deviations $\overline{\sigma}^{-1}$ (higher value is better):

$$\overline{\sigma}^{-1} = \frac{\sigma_{ACO}}{\sigma_{SEC}} = (1 - p_{SEC})^N \sqrt{\frac{A_0 (1 + r_a + Nr_i)}{A_0 + r_a}}, \tag{11}$$

and

$$\overline{\sigma}^{-1} = \frac{\sigma_{ACO}}{\sigma_{CMB}} = A_0 \sqrt{\frac{p_{CMB} (1 + r_a + Nr_i)}{A_0 + r_a + p_{CMB} N A_0 r_i}}, \tag{12}$$

for SEC and CMB, respectively. For ACO, $\overline{\sigma}^{-1} = 1$.



(a) Inverse depth standard deviation

(b) Energy consumption

Figure 6. **Theoretical comparison.** Different approaches are compared by (a) inverse depth standard deviation at the same energy consumption, and (b) required energy to achieve the same depth standard deviation. The relative performance of our approaches improves with the number of interfering cameras N, allowable peak power amplification A_0, and relative ambient light power r_a.

The required energy consumption to achieve the same depth standard deviation is also compared. We define $\overline{E}$ as:

$$\overline{E}_{SEC} = \frac{E_{SEC}}{E_{ACO}} = \frac{1}{(1 - p_{SEC})^{2N}} \frac{A_0 + r_a}{A_0 (1 + r_a + Nr_i)}, \tag{13}$$

and

$$\overline{E}_{CMB} = \frac{E_{CMB}}{E_{ACO}} = \frac{A_0 + r_a + p_{CMB} N A_0 r_i}{A_0 (1 + r_a + Nr_i)}, \tag{14}$$

for SEC and CMB, respectively. $\overline{E} = 1$ for ACO.

Figure 6 shows (a) $\overline{\sigma}^{-1}$ and (b) $\overline{E}$ of three approaches as a function of the number of interfering cameras N, allowable peak power amplification A_0, and ambient light strength r_a. When one of these parameters varies, the other parameters are fixed as $N = 5$, $A_0 = 8$, $r_a = 1$, and $r_i = 1$. As can be seen from the figure, $\overline{\sigma}^{-1}$ and $\overline{E}$ are closely related to each other. In general, $\overline{\sigma}^{-1}$ and $\overline{E}$ of SEC and CMB improve when N increases due to DC interference reduction which cannot be achieved by ACO. Although the relative performance of SEC and CMB improves with A_0, it saturates for SEC. Lower energy consumption is one of the key benefits of our approaches, which is critical in power-constrained applications. For additional comparisons with the same total peak power, see technical report.

7. Validation by Simulations

7.1. Verification of Depth Standard Deviation

We confirm the derived depth standard deviation equations of ACO, SEC, and CMB by simulations. For each approach, correlation values are computed, Poisson noise

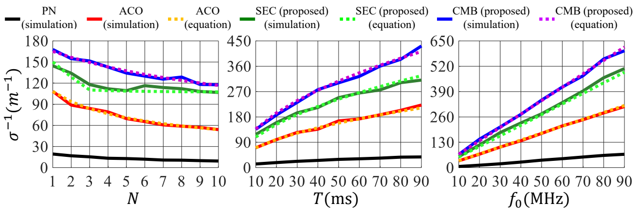


Figure 7. **Inverse depth standard deviations by simulations and equations.** Simulation results match well with the derived equations over various parameters. The proposed approaches outperform existing methods over a range of imaging scenarios.

is added, and the depth value is estimated from the noisy correlation values. This procedure is repeated 1000 times to compute the depth standard deviations. We also include the PN-sequence approach (PN) [6, 7] for simulations. We modified the original depth estimation algorithm [6] to accommodate unipolar demodulation functions and four correlation values for fair comparisons with other approaches.

Figure 7 shows the inverse depth standard deviations σ^{-1} of PN, ACO, SEC, and CMB over the number of interfering cameras N, total integration time T, and modulation frequency f_0 when the depth value is $1\,\mathrm{m}$. Solid and dotted lines indicate the results by simulations and equations, respectively. All simulation results match well with the derived depth standard deviation equations. The poor performance of PN id due to non-zero AC interference and relatively low modulation frequency to achieve the same measurable depth range as other approaches. See technical report for more details.

7.2. Simulations with a 3-D Model

Given a 3-D model, the depth values from a given camera position to all vertices of the model are computed. For each vertex, the correlation values are computed by 4 different approaches (PN, ACO, SEC, and CMB), photon noise is added, and the depth value is estimated from the corrupted correlation values. Once the model is reconstructed, root-mean-square error (RMSE) is computed for the objective quality comparison as well. Figure 8 compares the simulation results by different approaches over different number of interfering cameras N. RMSE values (in mm) are shown below the results. Although absolute performance of all approaches decreases with N, the relative performance of SEC and CMB increases compared to PN or ACO in both objective and subjective quality.

8. Hardware Prototype and Experiments

We developed a proof-of-concept hardware prototype to implement ACO, SEC, and CMB. Our setup consists of four C-ToF cameras (OPT8241-CDK-EVM, Texas Instruments [2]) and four microcontrollers (Arduino UNO) to generate random binary sequences (Figure 9). The square waves at 50% duty cycle are used as the modulation and

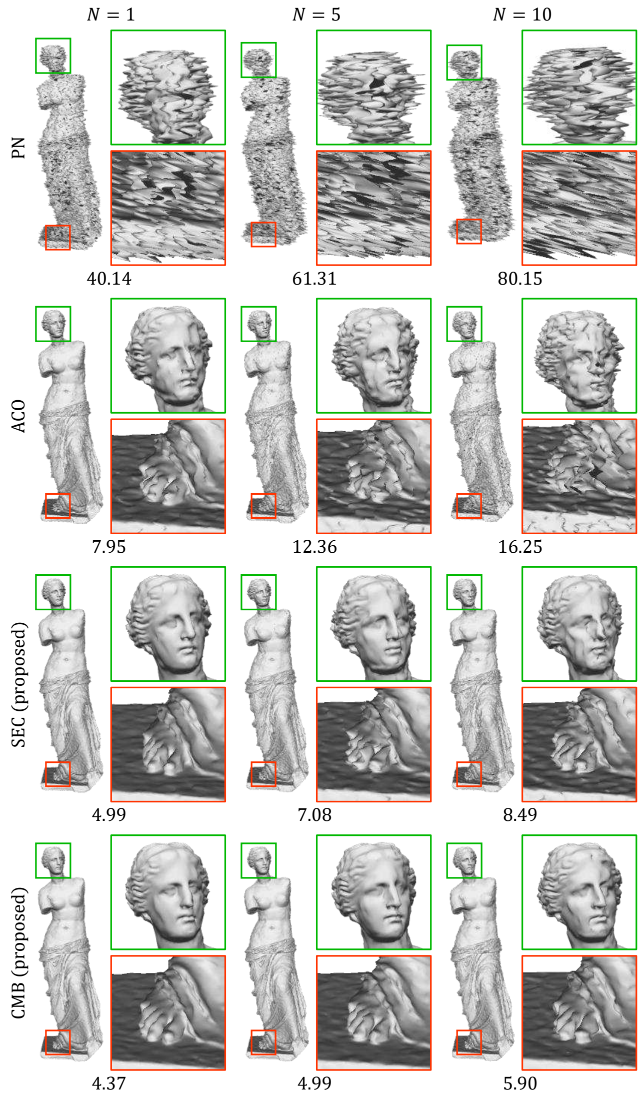


Figure 8. **3-D model reconstruction over different number of interfering cameras.** Our approaches achieve better performance in both subjective and objective quality over different number of interfering cameras N. The RMSE values (in mm) are shown.

demodulation functions. Since a frame is the most basic structure of the camera to access depth values, we used a frame as a slot. For ACO and CMB, four different modulation frequencies $\mathcal{B} = \{18, 20, 22, 24\}$ (MHz) are used for four different cameras. The depth values from all time slots of a primary camera are averaged to obtain a depth value for ACO. For SEC and CMB, the cameras operate in the slave mode to be activated by external pulses generated with an Arduino according to the given slot ON probability by which the slot activation is determined. The depth values from non-clashed ON slots and all ON slots are averaged to obtain depth values for SEC and CMB, respectively. Since it is challenging to amplify peak power of the light source for SEC and CMB, we lower it for ACO instead using the ND-filters (NE20A-B, Thorlabs) with an optical density fil-

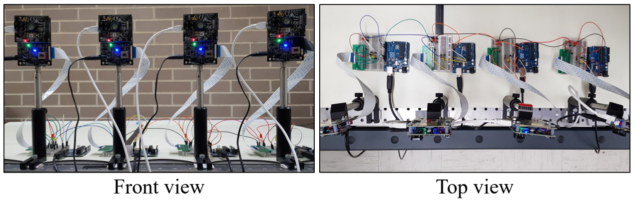


Figure 9. **Hardware prototype.** Front and top views of our setup to implement ACO, SEC, and CMB. The setup consists of four C-ToF cameras and four microcontrollers to generate random binary sequences to activate the cameras by given slot ON probabilities.

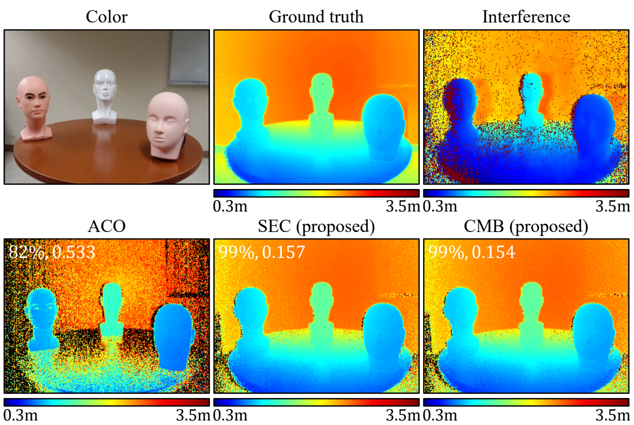


Figure 10. **Performance comparison via real experiments.** Multi-frequency coding is used in the three different approaches. The % of inliers (non-black pixels) and RMSE values (in m) at the inliers are represented for comparison between approaches.

ter, while keeping the total energy consumption the same.

Results with multi-frequency coding scheme: One of the key benefits of our approach is its ability to be used with any C-ToF coding scheme. To demonstrate this capability, we used a multi-frequency coding scheme with two frequencies [9]. We use the set of modulation frequencies $\mathcal{B} = \{18, 20, 22, 24\}$ (MHz) as the base frequencies, and $\{27, 30, 33, 36\}$ (MHz) as the de-aliasing frequencies. $0.83\,\mathrm{ms}$ is used for slot integration time. Figure 10 shows the color image and ground truth depth map of a face mannequin along with interference result and estimated depth maps by three approaches. Depths at the regions with lowest 1% number of photons are not recovered, and shown in black as outliers. For each approach, $\%$ of inliers and RMSE values (in m) for inliers are represented on the results. Although systematic depth errors are removed by all approaches, our approaches show significantly reduced noise compared to ACO.

Energy consumption comparison: We obtain depth estimation results with different energy consumption and compare them between different approaches. Different energy consumption is achieved by changing slot integration time: low energy ($0.83\,\mathrm{ms}$), medium energy ($1.83\,\mathrm{ms}$), and high energy ($2.83\,\mathrm{ms}$). Multi-frequency mode is deactivated and the set of modulation frequencies $\mathcal{B}$ are used as the base frequencies. Figure 11 shows the depth estimation results by different approaches over different energy consumption along with color image, ground truth depth map and interference result. Our approaches can obtain better results than ACO with only 30% of the energy consumed for ACO.

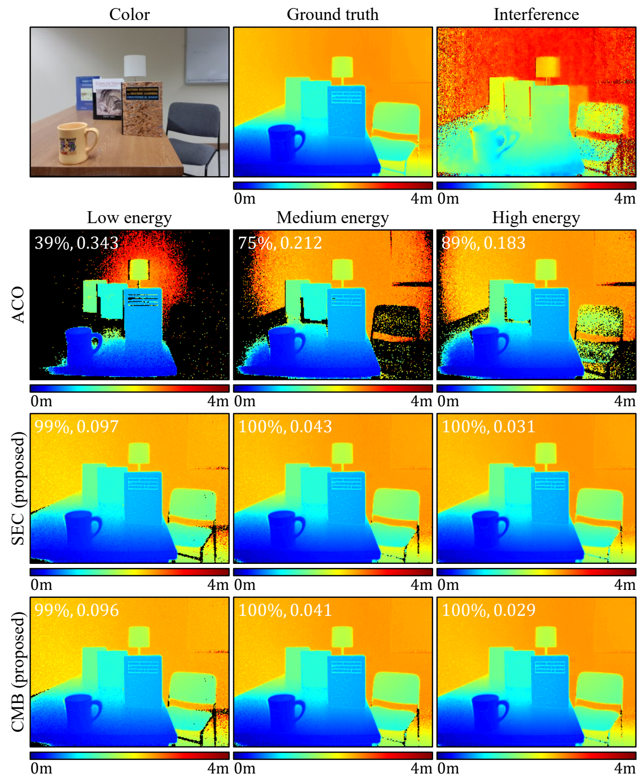


Figure 11. **Depth estimation comparison over different energy consumption.** Our approaches show better performance at lower energy consumption than the conventional approach. The % of inliers (non-black pixels) and RMSE values (in m) at the inliers are represented for comparison between approaches.

9. Discussion and Future Outlook

We propose stochastic exposure coding, a novel approach for mitigating both both AC and DC components of multi-camera interference in C-ToF imaging. This capability enables high precision depth estimation with low energy consumption. We demonstrat the performance benefits of the proposed approaches with theoretical analysis, simulations and real experiments. The proposed approach operates in an independent layer in C-ToF coding such that it can be incorporated with wide range of C-ToF coding functions, and various hardware platforms.

Acknowledgement. This research was supported in parts by the ONR grant number N00014-16-1-2995, and the DARPA REVEAL program.

References

[1] 3D ToF Development Kits, PMD. https://pmdtec.com/picofamily/.

[2] OPT8241 3D Time-of-Flight (ToF) Sensor Evaluation Module, Texas Instruments. http://www.ti.com/tool/OPT8241-CDK-EVM/.

[3] Norman Abramson. The aloha system: another alternative for computer communications. In *Proceedings of the November 17-19, 1970, fall joint computer conference*, pages 281–285. ACM, 1970.

[4] Supreeth Achar, Joseph R Bartels, William L Whittaker, Kiriakos N Kutulakos, and Srinivasa G Narasimhan. Epipolar time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 36(4):37, 2017.

[5] Pia Breuer, Christian Eckes, and Stefan Müller. Hand gesture recognition with a novel ir time-of-flight range camera–a pilot study. In *International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*, pages 247–260. Springer, 2007.

[6] Bernhard Büttgen, Felix Lustenberger, Peter Seitz, et al. Pseudonoise optical modulation for real-time 3-d imaging with minimum interference. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 54(10):2109–2119, 2007.

[7] Bernhard Büttgen and Peter Seitz. Robust optical time-of-flight range imaging based on smart pixel structures. *IEEE Trans. on Circuits and Systems*, 55(6):1512–1525, 2008.

[8] Yan Cui, Sebastian Schuon, Derek Chan, Sebastian Thrun, and Christian Theobalt. 3d shape scanning with a time-of-flight camera. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1173–1180. IEEE, 2010.

[9] David Droeschel, Dirk Holz, and Sven Behnke. Multi-frequency phase unwrapping for time-of-flight cameras. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1463–1469. IEEE, 2010.

[10] Thomas Fersch, Robert Weigel, and Alexander Koelpin. A cdma modulation technique for automotive time-of-flight lidar systems. *IEEE Sensors Journal*, 17(11):3507–3516, 2017.

[11] Jan Fischer, Benjamin Huhle, and Andreas Schilling. Using time-of-flight range data for occlusion handling in augmented reality. *IPT/EGVE*, 109116, 2007.

[12] Sergi Foix, Guillem Alenya, and Carme Torras. Lock-in time-of-flight (tof) cameras: A survey. *IEEE Sensors Journal*, 11(9):1917–1926, 2011.

[13] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments. In *In the 12th International Symposium on Experimental Robotics (ISER*. Citeseer, 2010.

[14] Leonard Kleinrock and Fouad Tobagi. Packet switching in radio channels: Part i-carrier sense multiple-access modes and their throughput-delay characteristics. *IEEE transactions on Communications*, 23(12):1400–1416, 1975.

[15] Robert Lange and Peter Seitz. Solid-state time-of-flight range camera. *IEEE Journal of quantum electronics*, 37(3):390–397, 2001.

[16] Lianhua Li, Sen Xiang, You Yang, and Li Yu. Multi-camera interference cancellation of time-of-flight (tof) cameras. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 556–560. IEEE, 2015.

[17] Stefan May, Bjorn Werner, Hartmut Surmann, and Kai Pervolz. 3d time-of-flight cameras for mobile robotics. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 790–795. Ieee, 2006.

[18] Dong-Ki Min, Ilia Ovsiannikov, Yohwan Noh, Wanghyun Kim, Sunhwa Jung, Joonho Lee, Deokha Shin, Hyekyung Jung, Lawrence Kim, Grzegorz Waligorski, et al. Pseudo-random modulation for multiple 3d time-of-flight camera operation. In *Three-Dimensional Image Processing (3DIP) and Applications 2013*, volume 8650, page 865008. International Society for Optics and Photonics, 2013.

[19] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *ACM transactions on graphics (TOG)*, volume 25, pages 795–804. ACM, 2006.

[20] Shikhar Shrestha, Felix Heide, Wolfgang Heidrich, and Gordon Wetzstein. Computational imaging with multi-camera time-of-flight systems. *ACM Transactions on Graphics (ToG)*, 35(4):33, 2016.

[21] Bernard Sklar. *Digital communications*, volume 2. Prentice Hall Upper Saddle River, 2001.

[22] Antonios Stylogiannis et al. Continuous wave laser diodes enable fast optoacoustic imaging. *Photoacoustics*, 9:31–38, 2018.

[23] Refael Z Whyte, Andrew D Payne, Adrian A Dorrington, and Michael J Cree. Multiple range imaging camera operation with minimal performance impact. In *Image Processing: Machine Vision Applications III*, volume 7538, page 75380I. International Society for Optics and Photonics, 2010.