Mitigating AC and DC Interference in Multi-ToF-Camera Environments

Jongho Lee and Mohit Gupta, *Member, IEEE*

Abstract—Multi-camera interference (MCI) is an important challenge faced by continuous-wave time-of-flight (C-ToF) cameras. In the presence of other cameras, a C-ToF camera may receive light from other cameras' sources, resulting in potentially large depth errors. We propose stochastic exposure coding (SEC), a novel approach to mitigate MCI. In SEC, the camera integration time is divided into multiple time slots. Each camera is turned on during a slot with an optimal probability to avoid interference while maintaining high signal-to-noise ratio (SNR). The proposed approach has the following benefits. First, SEC can filter out both the AC and DC components of interfering signals effectively, which simultaneously achieves high SNR and mitigates depth errors. Second, time-slotting in SEC enables 3D imaging without saturation in the high photon flux regime. Third, the energy savings due to camera turning on during only a fraction of integration time can be utilized to amplify the source peak power, which increases the robustness of SEC to ambient light. Lastly, SEC 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 thorough theoretical analysis, simulations and real experiments, across a wide range of imaging scenarios.

Index Terms—Multi-Camera interference, Time-of-flight cameras, 3D cameras, Time-division multiple access

1 Introduction

3D cameras are revolutionizing several aspects of our lives (Fig. 1 (a)). Autonomous vehicles and delivery robots use depth cameras to capture the geometry of the surroundings for safe navigation [2], [3]. 3D sensing is needed in interactive augmented reality (AR) [4] for providing a truly immersive virtual experience. In these, and several other 3D imaging applications, time-of-flight (ToF) cameras are fast becoming the method of choice. Their compact form-factors, low-cost, and low computational complexity have resulted in emergence of several commodity ToF cameras [5], [6], [7]. However, as these cameras become ubiquitous in our daily lives, an important problem arises: multi-camera interference (MCI). MCI is especially critical for continuous-wave ToF (C-ToF) imaging, where light is emitted and received continuously, and typically over a large field-of-view (FOV). When several C-ToF cameras capture the same scene concurrently, each sensor receives light from the light sources of other cameras, which results in large, systematic depth errors as shown in Fig. 1 (b) and (c).

Why is MCI an increasingly important problem now? Till recently, ToF cameras were largely based on scanning laser beams that illuminate only a few scene point at a time. Due to scanning, the probability that two lasers illuminate the same point simultaneously is relatively low, thus lowering the chances of MCI. However, scanning-based systems require long acquisition times, and are costly, bulky and have reliability issues due to mechanical moving parts.

To address these limitations, *solid-state ToF cameras* are increasingly being preferred especially in demanding real-world applications. These cameras flood-illuminate large scene regions without scanning, making them faster, smaller, and cheaper. But, there is a trade-off: Several ToF cameras flood-illuminating a scene considerably increase the likelihood of MCI. Considering that the capabilities of solid-state cameras are growing rapidly due to their compatibility with CMOS technology [8], [9], MCI is becoming a critical issue.

Conventional approaches to reduce MCI in C-ToF cameras: One way to mitigate MCI is to assign orthogonal coding functions to different C-ToF camreas. For example, periodic functions of different frequencies [10] or phases [11], [12], [13], or pseudo-random functions [14], [15], [16]. While these approaches can reduce interference in theory, they have practical limitations. The intensity of light from a 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 AC interference can be removed by the conventional methods, the remaining 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) degrades considerably, making reliable depth estimation challenging (Fig. 1 (d)). Besides, limited number of orthogonal frequencies and codes cannot accommodate the growing number of ToF cameras.

How to reduce both AC and DC interference? We propose a novel MCI reduction technique designed to mitigate *both* DC and AC interference. Our approach is based on time-division multiple access (TDMA), a well-known wireless communication scheme which facilitates multi-user access of a shared channel. In TDMA, a single shared channel is divided into multiple time slots, and only a single user can send a signal during each slot without any interference [17]. However, perfect temporal synchronization between users is essential for TDMA, which is done by a central authority

- *A preliminary version of this paper appeared in [1]*
- *J. Lee and M. Gupta are with the Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, 53706. E-mail: {jongho, mohitg}@cs.wisc.edu.*

Manuscript received August 8, 2023.

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2023.3307564

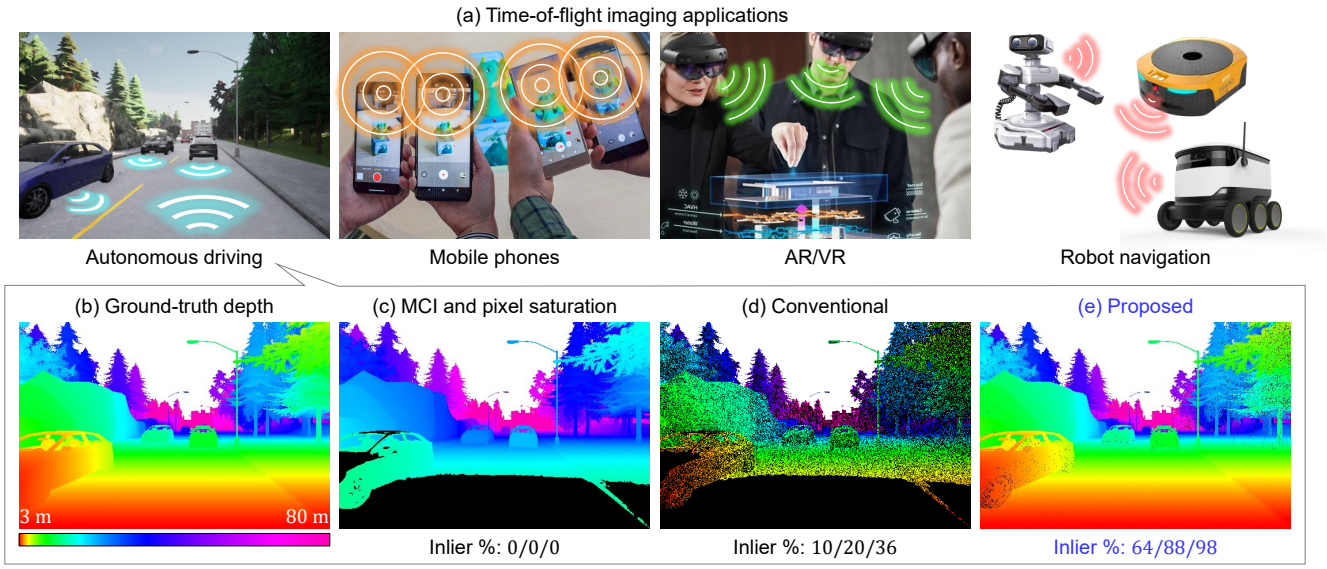

Fig. 1. **Time-of-flight imaging in multi-camera environments.** (a) Time-of-flight (ToF) cameras are becoming popular in several 3D imaging applications where multiple ToF cameras are used simultaneously in a shared physical space. (b, c) In a multi-ToF-camera environment, interference between ToF cameras results in large depth errors. Furthermore, high number of signal and ambient photons can cause pixel saturation (black pixels). (d) Conventional techniques to reduce interference cannot overcome low signal-to-noise ratio (SNR) and saturation. (e) The proposed approach achieves high SNR and reliable long-range depth imaging without saturation in high photon flux regime. The three numbers underneath each depth estimation are the percent of inlier pixels that lie within $0.5\,\%$, $1\,\%$, and $2\,\%$ of the true depth. Outlier pixels above $2\,\%$ depth error and saturated pixels are illustrated in black.

(e.g., a base station). This makes it challenging to apply TDMA directly in multi-ToF-camera environment, where there is no central authority for camera synchronization [15].

Stochastic exposure coding: Can we implement a TDMA-like approach to reduce MCI without synchronization? Our key idea is to leverage *stochasticity*. The proposed approach, called stochastic exposure coding (SEC), divides the total exposure time of each camera into multiple slots as TDMA. However, unlike TDMA, the camera and the source are turned on in each time slot randomly with a certain probability p_{ON}. 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. If a clash happens, the clash slots are discarded not to affect depth estimation.[1] We propose a compute-efficient and accurate clash-check algorithm to identify clash slots.

What is the optimal p_{ON} for SEC? The choice of p_{ON} is important since it determines the performance of SEC. If p_{ON} is too low, the clashes (multiple simultaneously active cameras) will be reduced, but the SNR will be lowered since the cameras are inactive during most of the integration time. If p_{ON} is too high, more clash slots will happen, resulting in lower SNR again. We derive the optimal p_{ON} as a function of system constraints and the number of interfering cameras based on a detailed theoretical analysis. The optimal p_{ON} enables each source to send light sufficiently sparsely to mitigate interference without synchronization. Given a fixed energy budget, this results in accurate depth estimates while maintaining a high SNR, as shown in Fig. 1 (e).

1. This approach is similar to random-access protocols in communication such as ALOHA [18] and CSMA [19] in that packets are sent randomly. However, while these protocols re-send packets whenever collision happens since each packet has unique information, SEC simply discards clash slots since all slots have the same depth information.

Enhanced dynamic range: For 3D imaging under bright sunlight, the camera source needs to emit a strong signal. Consequently, camera pixels imaging close scene points may receive a large number of signal and ambient photons, potentially saturating the sensor. SEC, as a by-product of time-slotting, prevents saturation thereby increasing the effective dynamic range resulting in correct depth estimation for a large depth range. In addition, SEC concentrates the source peak power into a fraction of the slots, thus enabling reliable depth estimation for distant scene points under high ambient light, which is challenging with conventional approaches as shown in Fig. 1 (d) and (e).

Layered view of C-ToF coding: A key benefit of SEC is that it does not require 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 [20]. This creates a layered view of C-ToF camera coding, as shown in Fig. 2. Conventional approaches to reduce MCI 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 real experiments that such combined multi-layer coding approaches (Fig. 2) 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.

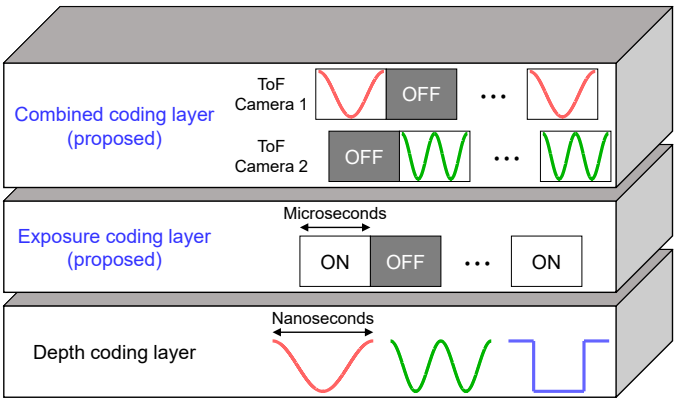

Fig. 2. **Layered C-ToF coding.** The proposed approach operates in the exposure coding layer, where the camera and the source are modulated at micro/millisecond time scales. In contrast, existing MCI reduction approaches operate in the lower depth coding layer, where modulation is performed at nanosecond time scales.

Limitations: Since SEC requires an integration-readout-reset cycle for each ON slot to read sensor measurements and check the slot clashes, it can result in increased data bandwidth (or a lower frame rate) and extra computation when the number of ON slots is high. However, there exist upper bounds for the required number of ON slots as we discuss in Section 4.3. In multi-layer coding where SEC is combined with the conventional MCI reduction approaches, the clash check is not necessary and there is no sacrifice of the frame rate as compared to the conventional MCI reduction approaches (Section 5).

2 Related Work

Multi-device interference in C-ToF imaging: Most existing approaches for MCI reduction rely on orthogonal functions, such as periodic functions of different modulation frequencies for different cameras [10], and pseudo-noise sequences [14], [15], [16]. Other approaches divide the total integration time into multiple time slots and randomly assign one of predetermined phases to each slot [11], [12], [13]. 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 [21]. Although this approach can also reduce DC interference, it requires mechanical scanning. In contrast, our approach can be implemented without moving parts.

Multi-path interference in C-ToF imaging: Multi-path interference (MPI) is also an important problem in C-ToF imaging, where a pixel receives light from several scene points, causing depth error. Many solutions have been proposed to address MPI [21], [22], [23], [24], [25], [26], [27]. Although the goal of both MCI and MPI reduction approaches is to mitigate the effect of unwanted signals, their solutions are typically different. In MPI, the corrupting signal is from the measuring camera itself, whereas in MCI, the interfering signal is from different cameras.

Learning-based approaches in C-ToF imaging: Recently, several learning-based methods [28], [29], [30], [31], [32] have been proposed for C-ToF imaging. However, the goal of these approaches is to remove depth errors by shot noise and multi-path interference (MPI), and there is no learning-based approach to mitigate multi-camera interference (MCI) to our best knowledge. This is because MPI is a local artifact which can be handled by using neighboring information while MCI is a global artifact which is hard to remove with learning at the moment.

3 Mathematical Preliminaries

C-ToF Image Formation Model: A C-ToF camera consists of a (typically co-located) camera and a light source [33]. 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 $M(t)$:

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

where s is a camera-dependent scale factor encapsulating sensor gain and sensitivity, T is the total capture 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 depth, several (≥ 3) different $C(\mathbf{p}; d)$ values are measured by using different pairs of modulation and demodulation functions [33].

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 $K \geq 3$ correlation 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. We drop the argument $\mathbf{p}$ from $C(\mathbf{p}; d)$ for brevity. 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 $\{C_k(d)\}$, $k \in \{1, \ldots, K\}$ 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. 3 (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 [34]. For example, for a 4-tap sinusoid coding scheme, d is given as:

$$d = \frac{c}{4\pi f_0} \tan^{-1}\left(\frac{C_4 - C_2}{C_1 - C_3}\right). \tag{3}$$

2. Several C-ToF camera architectures [14], [33] 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 this paper can be generalized to bipolar $D(t)$. See the supplementary report.

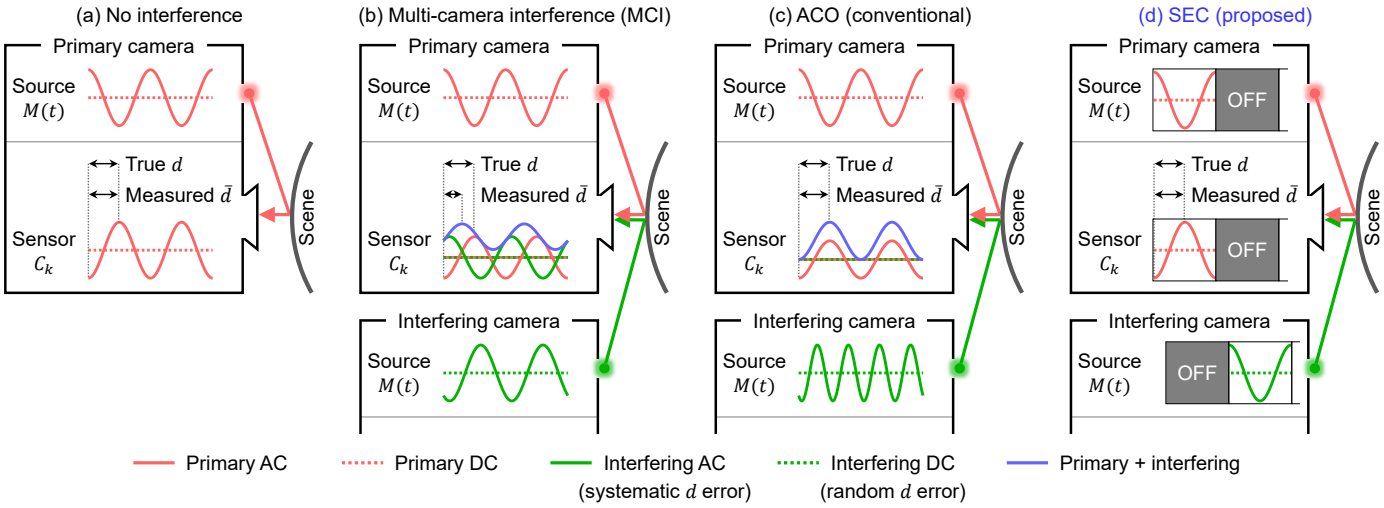

Fig. 3. **Multi-camera interference and interference reduction in C-ToF imaging.** (a) In C-ToF imaging, depth is recovered from the phases of the measured waveforms. (b) If there are multiple cameras, interfering sources corrupt the measured waveforms, resulting in systematic depth errors. (c) Conventional 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.

However, the estimated d by Eq. 3 differs from the true d due to shot noise. The standard deviation of d is given as:

$$\sigma \approx \frac{c}{2\sqrt{2}\pi f_0 \sqrt{T}} \frac{\sqrt{e_s + e_a}}{e_s}. \tag{4}$$

where f_0 is the modulation frequency, T is the total capture time for each measurement, and c is the light speed. $e_s = s\alpha P_s$ and $e_a = sP_a$ are the average number of electrons generated per unit time by the camera's own source and the ambient source, respectively. See the supplementary report for the derivation of Eq. 4.

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 correlation measured by one of the cameras (referred to as the primary camera) is given by:

$$C_{\text{tot}}(d) = C(d) + \underbrace{\sum_{n=1}^{N} C_n(d)}_{\text{multi-camera interference}}, \tag{5}$$

where N is the number of interfering cameras, $C(d)$ is the correlation 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 correlation due to the n^{th} interfering source. $R_n(t)$ is the radiance received by the primary camera due to light emitted by the n^{th} interfering source. The summation term in Eq. 5 corrupts the true correlation $C(d)$, thus resulting in erroneous depth estimates.

Assuming all the sources use sinusoids of the same frequency, the correlation values $\{C_{n,k}\}$, $k \in \{1, \ldots, K\}$ measured by the camera due to the n^{th} source *also form a sinusoid*. The total measurement $\{C_{\text{tot},k}\}$, $k \in \{1, \ldots, K\}$ (Eq. 5) is the sum of these individual sinusoids, and thus, also forms a sinusoid. This is shown in Fig. 3 (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 correlation values $\{C_{n,k}\}$, $k \in \{1, \ldots, K\}$ 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.[3] As a result, the total measurement waveform $\{C_{\text{tot},k}\}$, $k \in \{1, \ldots, K\}$ *has the same phase* as the measurement waveform 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, as shown in Fig. 3 (c).

We call this *AC-Orthogonal (ACO)* approach since it removes the AC component of the interference. However, the offset (DC-component) of the total waveform still increases, as shown in Fig. 3 (c). The extra offset acts as additional ambient light, and thus lowers the SNR of the estimated depth due to increased shot noise [11].[4] For example, the depth standard deviation for a 4-tap sinusoid-based ACO method can be derived from Eq. 4 by adding the sum of DC components from all interfering sources to e_a:

$$\sigma_{\text{ACO}} \approx \frac{c}{2\sqrt{2}\pi f_0 \sqrt{T}} \frac{\sqrt{e_s + e_a + Ne_i}}{e_s}, \tag{6}$$

where e_i is the electron counts per unit time generated by the interfering source (see the supplementary report for the derivation of Eq. 6). Without loss of generality, we assume that e_i is the same for all interfering cameras. Note that ACO is not a specific approach. Instead, it represents *a family of conventional approaches* which do not mitigate DC interference, but reduce only AC interference (Sec. 2).

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

4. With bipolar demodulation functions, although the DC-offset is removed, the shot noise still increases. See the supplementary report.

Although an ACO approach prevents systematic depth errors due to MCI, random depth errors due to photon noise increase as the number of interfering cameras increases (Eq. 6). This is because each interfering source has a non-zero DC component, contributing additional photon noise to the correlation measurements. This raises the following question: Is it possible to design a *DC-Orthogonal* (DCO) approach, that removes *both* AC and DC interference?

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 as shown in Fig. 3 (d). 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_{ON}. 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 Fig. 3 (d). 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 raised: What is the optimal p? To address this, we derive the depth standard deviation of the SEC in terms of p.

Depth standard deviation of SEC: Consider a scene being imaged by $N+1$ C-ToF cameras (1 primary + N interfering cameras). For ease of analysis, we assume the cameras are identical. For each camera, the total capture time of a frame (the most basic unit to estimate the depth) is divided into multiple slots of the same duration as shown in Fig. 4. Every slot is turned on with a probability p. 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 (Fig. 4). The probability p_{noclsh} that a given slot does not produce a clash, i.e., only one camera is active during that slot, is:

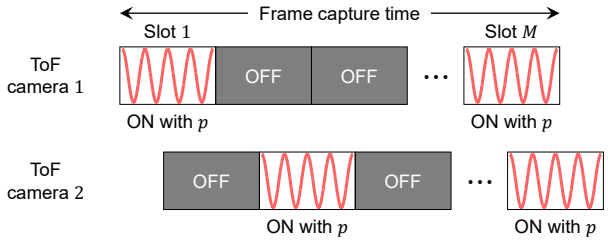

Fig. 4. **Stochastic exposure coding.** A frame, the most basic unit to estimate the depth, is divided into M number of slots. Each slot is activated with a probability p. A depth value for the frame is estimated from no-clash ON (activated) slots.

$$p_{\text{noclsh}} = p\,(1-p)^{2N}. \tag{7}$$

Assuming we can identify all the no-clash slots, the effective exposure time for each camera, on average, is $T\,p_{\text{noclsh}}$, where T is the total frame 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 $T\,p_{\text{noclsh}}$ and source power amplification A, the depth standard deviation of SEC can be derived from Eq. 4:

$$\sigma_{\text{SEC}} \approx \frac{c}{2\sqrt{2}\pi f_0 \sqrt{T p_{\text{noclsh}}}} \frac{\sqrt{A e_s + e_a}}{A e_s}. \tag{8}$$

See the supplementary report for the derivation of Eq. 8. Although randomness due to slot ON probability can influence the depth standard deviations, the effect of randomness is relatively small if a sufficient number of slots are used.

Optimal slot ON probability for SEC: The optimal slot ON probability p_{SEC} for SEC is obtained by minimizing σ_{SEC}:

$$\boxed{p_{\text{SEC}} = \arg\min_p \sigma_{\text{SEC}} = \min\left(\frac{1}{2N+1}, \frac{1}{A_0}\right).} \tag{9}$$

See the supplementary report for the derivation. As the number of interfering cameras N increases, the optimal slot ON probability p_{SEC} decreases so that the number of clashes remains low. When the allowable source peak power amplification A_0 increases, p_{SEC} also decreases (if peak power amplification is very high like a Dirac delta function, only one random slot might be optimal to avoid interference without reducing the SNR). If p is smaller or larger than p_{SEC}, the optimal SNR cannot be achieved since the effective exposure time is reduced.

Figure 5 shows the inverse depth standard deviation ${\sigma_{\text{SEC}}}^{-1}$ over p with different number of interfering cameras N and different source peak power amplification A_0. Without source peak power amplification ($A_0 = 1$), the optimal slot ON probability p_{SEC} (maximizing ${\sigma_{\text{SEC}}}^{-1}$) is determined by N. When N increases, p_{SEC} decreases, and ${\sigma_{\text{SEC}}}^{-1}$ at p_{SEC} decreases (σ_{SEC} at p_{SEC} increases). When source peak power

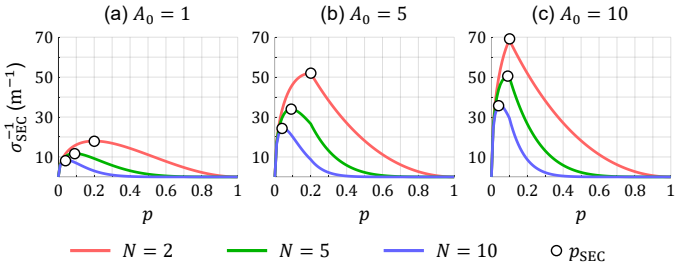

Fig. 5. **Optimal slot ON probability for SEC.** (a) The optimal slot ON probability p_{SEC} for SEC is determined only by the number of interfering cameras N when source peak power amplification is not allowed ($A_0 = 1$). (b, c) When peak power amplification is allowed ($A_0 > 1$), p_{SEC} is determined by N and A_0. p_{SEC} decreases when N and A_0 increase.

amplification is allowed, p_{SEC} is determined by N and A_0 (Eq. 9). $e_s = e_a = 10^6$ e$^-$/s, $T = 10$ ms, and $f_0 = 30$ MHz were used to create Fig. 5.

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.

Clash-check algorithm: Our clash-check algorithm is based on a simple intuition: In a clash slot, the camera receives light from multiple sources. Therefore, with high probability, the total amount of light received in that slot is higher as compared to no-clash slots. The total electron counts generated in a given slot by incident light is proportional to the sum of slot correlation values:

$$o_m = \sum_{k=1}^{K} c_{m,k},\ m \in \{1, \ldots, M_{\text{ON}}\}, \tag{10}$$

where $c_{m,k}$ is the k^{th} correlation value of the m^{th} ON slot (we use lower case c and upper case C to represent the slot and frame correlations, respectively) and M_{ON} is the number of ON slots within a frame. We compare o_m to a threshold, and if o_m is larger, the corresponding slot is discarded. The question is: How can we determine the clash threshold for correct identification of the clash slots?

If there is no clash and the number of generated electrons is sufficiently large, o_m can be well approximated as a Gaussian random variable with mean $\overline{o_m}$ and standard deviation $\sigma_{o_m} = \sqrt{\overline{o_m}}$ [35]. Thus, when there is no clash, stochastic upper and lower bounds of o_m can be defined as $\overline{o_m} \pm k\sigma_{o_m}$. We use the upper bound as the clash threshold o_{clsh} (i.e., we say that a clash happened if $o_m > o_{\text{clsh}}$):

$$o_{\text{clsh}} = \overline{o_m} + k\sigma_{o_m} = \overline{o_m} + k\sqrt{\overline{o_m}}. \tag{11}$$

$\overline{o_m}$ cannot be obtained by simply averaging measured o_ms over all ON slots since some of them have clashes. Instead, $\overline{o_m}$ is estimated from the minimum o_m, which is approximated as the lower bound of o_m. We define o_{min} as:

$$o_{\text{min}} = \min_m o_m,\ m \in \{1, \ldots, M_{\text{ON}}\}. \tag{12}$$

From $o_{\text{min}} = \overline{o_m} - k\sigma_{o_m} = \overline{o_m} - k\sqrt{\overline{o_m}}$,

$$\overline{o_m} = o_{\text{min}} + \frac{k^2}{2} + \sqrt{k^2 o_{\text{min}} + \frac{k^4}{4}}. \tag{13}$$

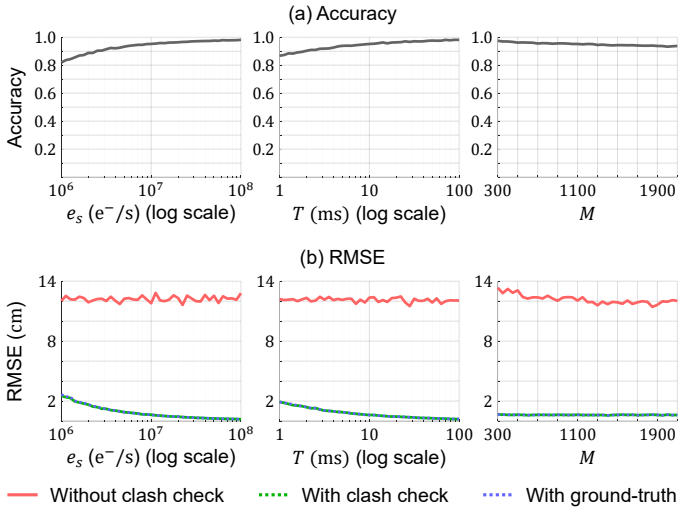

Fig. 6. **Performance of clash-check algorithm.** (a) The accuracy of our clash-check algorithm increases with the the electron counts per slot, which increase when e_s and T increase and M decreases. (b) Even when its accuracy is lower than 1, the clash-check algorithm is effective in filtering out the clash slots, thus improving depth estimation performance in terms of the root-mean-square error (RMSE). Depth estimation error with the no-clash slots identified by our clash-check algorithm is almost the same as that with the ground-truth no-clash slots.

The clash threshold o_{clsh} can be obtained by replacing $\overline{o_m}$ in Eq. 11 with Eq. 13 (we use k=2.5). Our clash-check algorithm is compute-efficient due to the closed form solution of o_{clsh}.

Figure 6 shows the performance of our clash-check algorithm. The accuracy of the clash-check algorithm is defined as $\frac{|\mathcal{A} \cap \mathcal{B}|}{|\mathcal{A} \cup \mathcal{B}|}$, where $\mathcal{A}$ is a set of the no-clash slots identified by our clash-check algorithm and $\mathcal{B}$ is a set of the ground-truth no-clash slots. Figure 6 (a) shows the accuracy of our clash-check algorithm as a function of the average number of signal electrons per unit time e_s, total frame capture time T, and total number of slots M in each frame. When one of these parameters varies, the other parameters are fixed as $e_s = e_a = 10^7$ e$^-$/s, $T = 10$ ms, $f_0 = 30$ MHz, $A_0 = 9$, $M = 1000$, $d = 3$ m, and $N = 5$. The plotted accuracy is the average for 1000 trials. As shown in Fig. 6 (a), the accuracy of our clash-check algorithm increases as the electron counts per slot increase (the electron counts per slot increase when e_s and T increase and when M decreases). When its accuracy is lower than 1 (perfectly accurate), the clash-check algorithm is still effective in filtering out the clash slots and improves depth estimation performance in terms of the root-mean-square error (RMSE), as shown in Fig. 6 (b). The RMSE values obtained from $\mathcal{A}$ and $\mathcal{B}$ are almost the same since the contributions by the false no-clash slots are negligible.

Depth estimation in SEC: Given M_{noclsh} number of no-clash slots, the depth d_m, $m \in \{1, \ldots, M_{\text{noclsh}}\}$ for each *no-clash slot* is estimated first. If we use a 4-tap sinusoid coding scheme, d_m can be obtained by Eq. 3, but using slot correlation values $c_{m,k}$, $k \in \{1, \ldots, K\}$. The depth for each *frame* is obtained by averaging d_m values:

$$d = \frac{1}{M_{\text{noclsh}}} \sum_{m=1}^{M_{\text{noclsh}}} d_m. \tag{14}$$

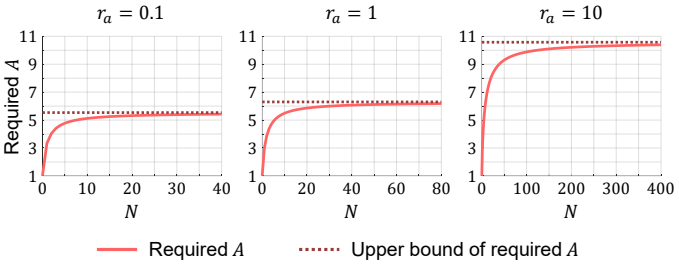

Fig. 7. **Required source peak power amplification for SEC.** The required source peak power amplification A for SEC to achieve better SNR than ACO increases with the number of interfering cameras N, but it eventually converges, for various relative ambient light strengths r_a

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 exposure 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 $\sigma_{\text{SEC}} \leq \sigma_{\text{ACO}}$:

$$\frac{1}{\sqrt{p_{\text{noclsh}}}} \frac{\sqrt{A + r_a}}{A} \leq \sqrt{1 + r_a + N r_i}, \tag{15}$$

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 7 shows the required peak power amplification A over different number of interfering cameras N at different ambient light strengths r_a ($r_i = 1$ was assumed). Although the required A increases with N, it eventually converges, as stated in the following result:

Result 1. If the source peak power amplification of SEC is larger than $\left(e + \sqrt{e\,(e + 2 r_a r_i)}\right)/r_i$, 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 $r_a = r_i = 1$.

The upper bounds of the required A for different r_as are shown in Fig. 7. See the supplementary report for the proof.

Practicality of achieving high peak power: Two factors should be considered regarding the practicality of increasing source peak power. First, in power-constrained devices (e.g., cell-phones), 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 used in C-ToF cameras (e.g., laser diodes) with high instantaneous peak power [36]. 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 [36].

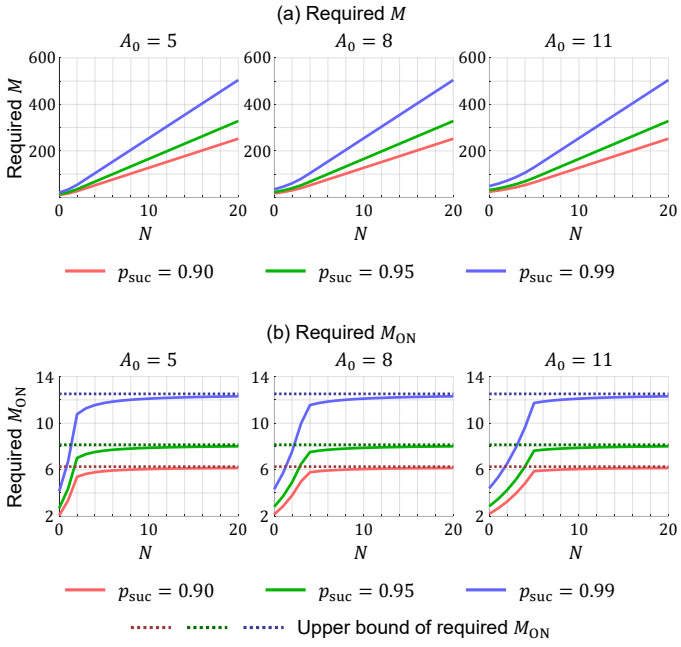

Fig. 8. **Required number of slots for SEC.** (a) More number of slots M is required if the number of interfering cameras N increases over various allowable peak power amplification A_0. (b) However, the required number of ON slots M_{ON} eventually converges, making it possible to achieve high frame rate for dynamic scenes.

Required number of slots: For correct depth estimation with SEC, at least one no-clash ON slot is needed. Then an important question is: How many slots are required during a frame to get at least one no-clash ON slot? Let p_{suc} be the probability of getting at least one no-clash ON slot during a frame. Then, p_{suc} is defined as $p_{\text{suc}} = 1 - (1 - p_{\text{noclsh}})^M$. From this, the required number of slots M given p_{suc} is

$$M = \frac{\ln(1 - p_{\text{suc}})}{\ln(1 - p_{\text{noclsh}})}, \tag{16}$$

where p_{noclsh} is defined by Eq. 7 with $p = p_{\text{SEC}}$. Figure 8 (a) shows the required number of slots M over the number of interfering cameras N at various allowable source peak power amplification A_0 and different desired success probability p_{suc}.

The number of ON slots that a camera would need to capture per frame is $M_{\text{ON}} = M p_{\text{SEC}}$. The required M_{ON} increases with N, but it is eventually bounded, as follows:

Result 2. The required number of ON slots M_{ON} for SEC to achieve the desired success probability p_{suc} converges to $-e \ln(1 - p_{\text{suc}})$ regardless of the number of interfering cameras. For example, when $p_{\text{suc}} = 0.9$, the required M_{ON} is upper bounded by 6.3.

See the supplementary report for the proof. Figure 8 (b) shows the required M_{ON} over the number of interfering cameras N with various desired success probability p_{suc} and allowable source peak power amplification A_0. The required M_{ON} increases with N, but eventually *converges*.

5 Multi-Layer Coding for Mitigating MCI

The proposed SEC creates a layered view of C-ToF camera coding, as shown in Fig. 2. 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 (MLC), it is no longer necessary to discard the clash slots since they do not introduce systematic depth errors. This makes repeated clash check unnecessary, leading to simpler depth estimation and an efficient frame structure.

Depth standard deviation of MLC: Depth standard deviation of MLC σ_{MLC} can be easily derived from Eq. 8:

$$\sigma_{\text{MLC}} \approx \frac{c}{2\sqrt{2}\pi f_0 \sqrt{Tp}} \frac{\sqrt{Ae_s + e_a + NpAe_i}}{Ae_s}, \tag{17}$$

where $A = \min\left(\frac{1}{p}, A_0\right)$. Tp is the effective exposure time of MLC, and $NpAe_i$ is the sum of DC components from all interfering sources. See the supplementary report for the derivation of Eq. 17.

Optimal slot ON probability: The optimal slot ON probability p_{MLC} for MLC is defined as p minimizing Eq. 17:

$$p_{\text{MLC}} = \arg\min_p \sigma_{\text{MLC}} = \frac{1}{A_0}. \tag{18}$$

Unlike p_{SEC}, p_{MLC} is independent of N. See the supplementary report for the derivation of Eq. 18.

Depth estimation in MLC: In MLC, both clash and no-clash slots participate in depth estimation. We compute the sum of correlation values from all ON slots within a frame:

$$C_k = \sum_{m=1}^{M_{\text{ON}}} c_{m,k},\ k \in \{1, \ldots, K\}, \tag{19}$$

where M_{ON} is the number of ON slots in the frame. The depth for each frame can be obtained by Eq. 3 when we use a 4-tap sinusoid coding.

6 Theoretical Performance Comparisons

We present theoretical comparisons between ACO, SEC and MLC in terms of (a) depth standard deviation at the same energy consumption and (b) 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 is better) as:

$$\overline{\sigma}_{\text{SEC}}^{-1} = \frac{\sigma_{\text{ACO}}}{\sigma'_{\text{SEC}}} = (1 - p_{\text{SEC}})^N \sqrt{\frac{A_0\,(1 + r_a + Nr_i)}{A_0 + r_a}}, \tag{20}$$

and

$$\overline{\sigma}_{\text{MLC}}^{-1} = \frac{\sigma_{\text{ACO}}}{\sigma_{\text{MLC}}} = \sqrt{\frac{A_0\,(1 + r_a + Nr_i)}{A_0 + r_a + p_{\text{MLC}} N A_0 r_i}}, \tag{21}$$

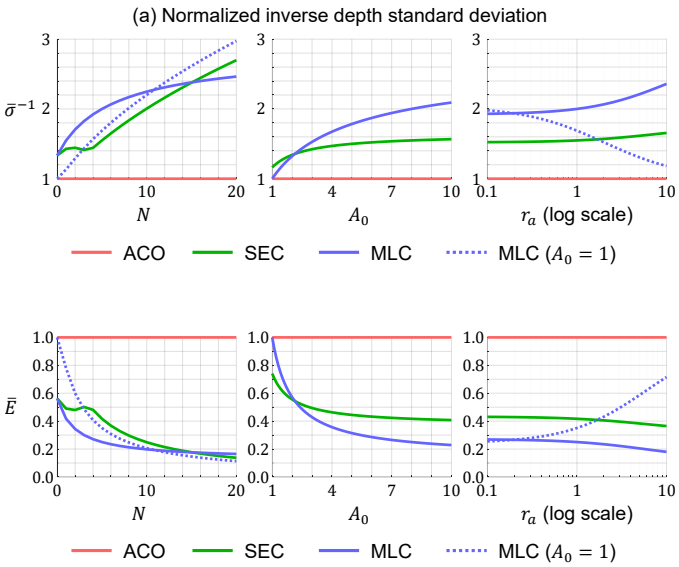

Fig. 9. **Theoretical comparison.** Different MCI reduction approaches are compared by (a) inverse depth standard deviation at the same energy consumption, and (b) required energy to achieve the same depth standard deviation. The relative performance of our approaches SEC and MLC improves with the number of interfering cameras N, allowable peak power amplification A_0, and relative ambient light power r_a.

for SEC and MLC, respectively. $\sigma'_{\text{SEC}} = \sigma_{\text{SEC}}\sqrt{p_{\text{SEC}}A}$ is the compensated σ_{SEC} to meet the same energy constraint. $\overline{\sigma}_{\text{ACO}}^{-1} = 1$ for ACO.

The required energy consumption to achieve the same depth standard deviation is also compared. Let E_{ACO}, E_{SEC} and E_{MLC} be the required energy consumption to achieve the same depth standard deviation for ACO, SEC and MLC, respectively. We define the normalized energy consumption $\overline{E}$ as:

$$\overline{E}_{\text{SEC}} = \frac{E_{\text{SEC}}}{E_{\text{ACO}}} = \frac{1}{(1 - p_{\text{SEC}})^{2N}} \frac{A_0 + r_a}{A_0\,(1 + r_a + Nr_i)}, \tag{22}$$

and

$$\overline{E}_{\text{MLC}} = \frac{E_{\text{MLC}}}{E_{\text{ACO}}} = \frac{A_0 + r_a + p_{\text{MLC}} N A_0 r_i}{A_0\,(1 + r_a + Nr_i)}, \tag{23}$$

for SEC and MLC, respectively. $\overline{E}_{\text{ACO}} = 1$ for ACO. Note that $\overline{E}_{\text{SEC}} = \overline{\sigma}_{\text{SEC}}^2$ and $\overline{E}_{\text{MLC}} = \overline{\sigma}_{\text{MLC}}^2$.

Figure 9 shows (a) $\overline{\sigma}^{-1}$ and (b) $\overline{E}$ of three MCI reduction 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 we discussed, $\overline{\sigma}$ and $\overline{E}$ are closely related to each other. In general, $\overline{\sigma}$ and $\overline{E}$ of SEC and MLC improve when N increases due to DC interference reduction which cannot be achieved by ACO. Although the relative performance of SEC and MLC 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 such as mobile devices and autonomous vehicles.

Performance without peak power amplification: If peak power amplification A_0 is 1, and the integration time is kept constant, the optimal ON probability of MLC is 1,

i.e., $p_{\text{MLC}} = 1/A_0 = 1$. In this case, MLC becomes the same as existing ACO approaches, with the same performance. The more interesting comparison is when the integration time is allowed to be increased. In this case, we can use lower ON probabilities to avoid clashes. Specifically, we set $p_{\text{MLC}} = 1/(2N+1)$. To keep the total energy constant, we increase the total integration time by $2N+1$. We show the performance of MLC without peak power amplification ($A_0 = 1$) in Fig. 9 with dotted lines. The performance of MLC with $A_0 = 1$ is higher than that with $A_0 = 8$ for large Ns due to reduced clash probabilities. However, no peak power amplification makes MLC vulnerable to ambient light.

SEC vs. MLC: Since MLC is a combined approach of SEC and ACO, the applicability of MLC is limited by ACO. For example, if the number of orthogonal frequencies for ACO is not sufficient to handle a huge number of C-ToF cameras, MLC is not directly applicable. One of the benefits of MLC over SEC is a higher frame rate because MLC does not need to read the accumulated charge of each ON slot for the clash check. The performance of SEC and MLC depends on the imaging conditions and system parameters. In certain conditions, MLC can achieve better performance than SEC, since MLC can increase the effective integration time by retaining the clash slots in addition to the no-clash slots. However, a large number of clash slots due to several interfering cameras can lower the performance of MLC due to increased DC interference, as shown in Fig. 9. In this case, higher source peak power amplification is useful if allowable.

7 Validation by Simulations

We have developed a physically accurate simulator for C-ToF imaging, which can emulate MCI and several MCI reduction approaches (ACO, SEC, and MLC). Given a scene with ground-truth intensity and depth values, our simulator can generate depth estimates with and without MCI reduction approaches over a wide range of operating parameters such as the number of interfering cameras, ambient light strength, frame capture time, modulation frequency, peak power amplification, and light source strength. Using this simulator, we compare the performance of different MCI reduction approaches in different imaging scenarios.

7.1 Verification of Depth Standard Deviation

We confirm the derived depth standard deviation equations of ACO (Eq. 6), SEC (Eq. 8), and MLC (Eq. 17) by simulations. Under the various MCI environments defined by different operating parameters, noise-free correlation values are computed by different MCI reduction approaches. Noisy correlation values are obtained from the Poisson distribution, and the depth is estimated from the corrupted correlation values. This procedure is repeated 2000 times to compute the depth standard deviations. Figure 10 shows the depth standard deviations σ of ACO, SEC, and MLC over the number of interfering cameras N, allowable peak power amplification A_0, modulation frequency f_0, frame integration time T, light source strength e_s, and ambient light strength e_a. When one of these parameters varies, the other parameters are fixed as $e_s = e_a = 10^7\ \text{e}^-/\text{s}$, $T = 10\,\text{ms}$,

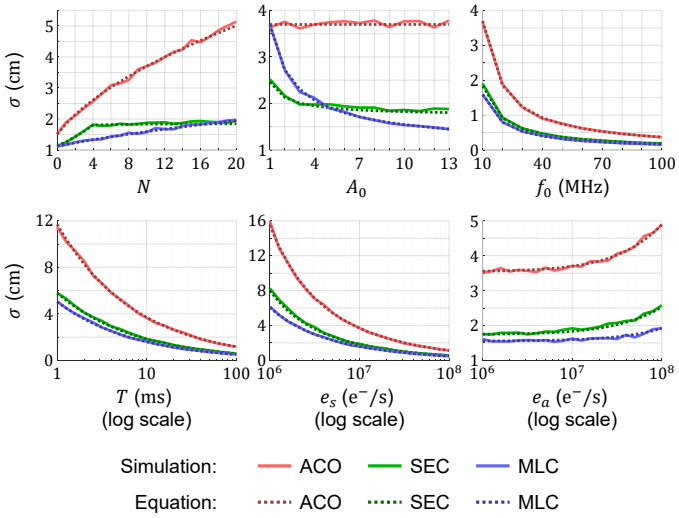

Fig. 10. **Depth standard deviations by simulations and equations.** Simulation results match well with the derived equations over various operating parameters. The proposed approaches outperform existing methods over a range of imaging scenarios.

$f_0 = 10\,\text{MHz}$, $A_0 = 9$, total number of time slots $M = 2000$, and scene depth $d = 2\,\text{m}$. In Fig. 10, solid and dotted lines indicate the results by simulations and equations, respectively. All simulation results match well with the derived depth standard deviation equations as shown in Fig. 10.

7.2 Simulations with 3D Scenes

Comparisons over different number of interfering cameras: Figure 11 compares the simulation results by different MCI reduction approaches over different number of interfering cameras N. RMSE values are shown below the results. Although absolute performance of all approaches decreases with N, the relative performance of SEC and MLC increases compared to ACO in both objective and subjective quality. This is because the proposed approaches mitigate not only AC interference but also DC interference, which cannot be handled by ACO.

Dynamic range comparisons: There is a limitation on the maximum number of photon-electrons CMOS pixels can collect. This limit is called a full-well capacity (FWC). In C-ToF imaging, when the light source illuminates a scene with high signal power under high ambient light, the ToF camera pixels imaging close scene points may receive a large number of signal and ambient photons. If the amount of generated photo-electrons exceeds the FWC, the pixels are saturated. Figures 12 (a) and (b) show a long-range ($\sim 200\,\text{m}$) outdoor scene under bright sunlight and its ground-truth depth created using the CARLA simulator [37]. We assume 10 C-ToF cameras are imaging the same scene. In this challenging outdoor multi-camera environment, a C-ToF camera suffers from both interference and pixel saturation as shown in Fig. 12 (c). To reconstruct the whole scene without saturation, ACO requires the unrealistically high FWC (Fig. 12 (e)). In contrast, SEC considerably mitigates saturation and interference in these challenging scenarios even with two orders of magnitude lower FWC compared to ACO. This is because SEC can reduce the accumulated number of photo-electrons by dividing the integration time

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2023.3307564

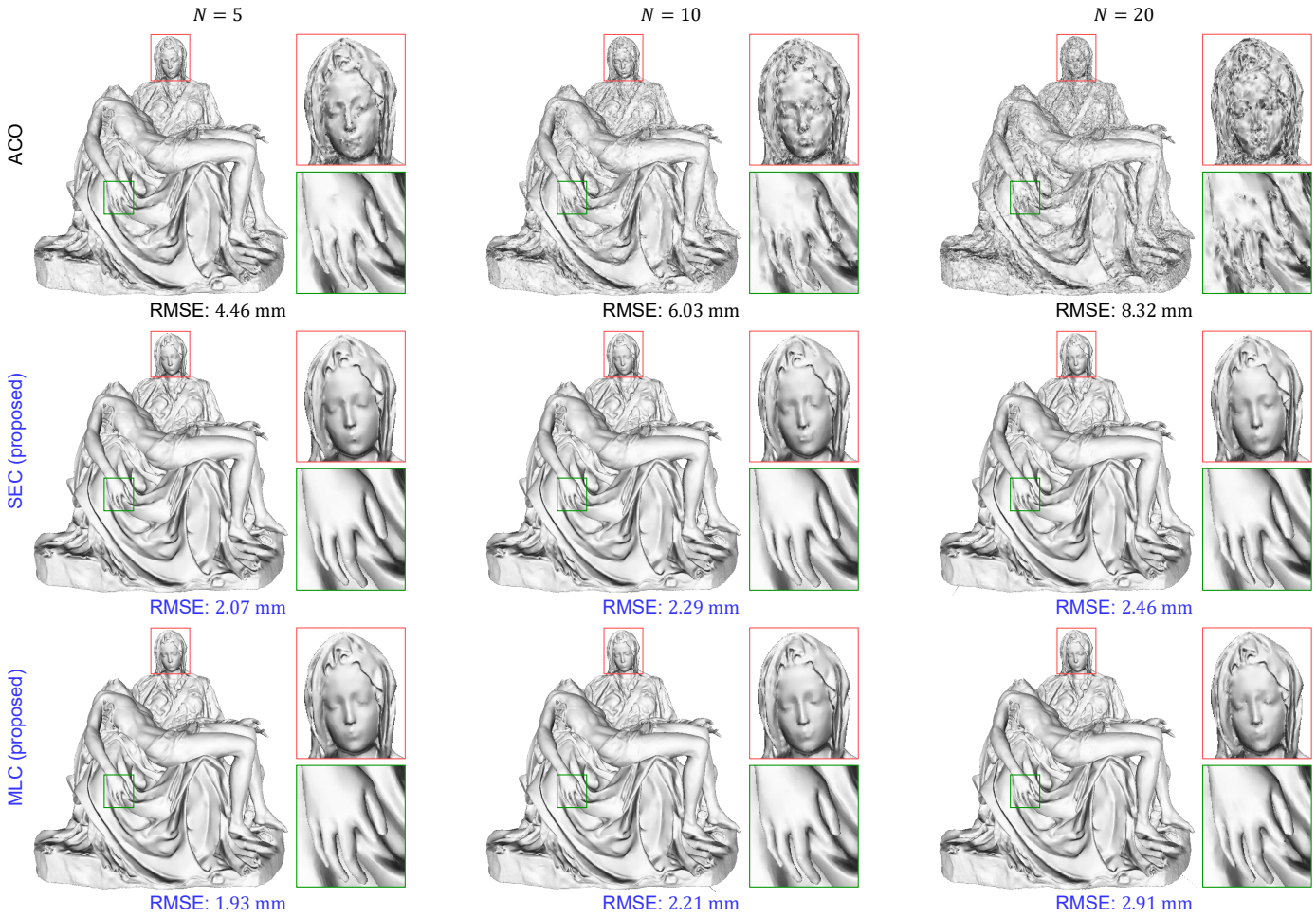

Fig. 11. **3D model reconstruction over different number of interfering cameras.** Our approaches achieve better performance than conventional ACO techniques in both subjective and objective quality over different number of interfering cameras N. This is because the proposed approaches reduce not only AC interference but also DC interference, which is challenging to mitigate with ACO.

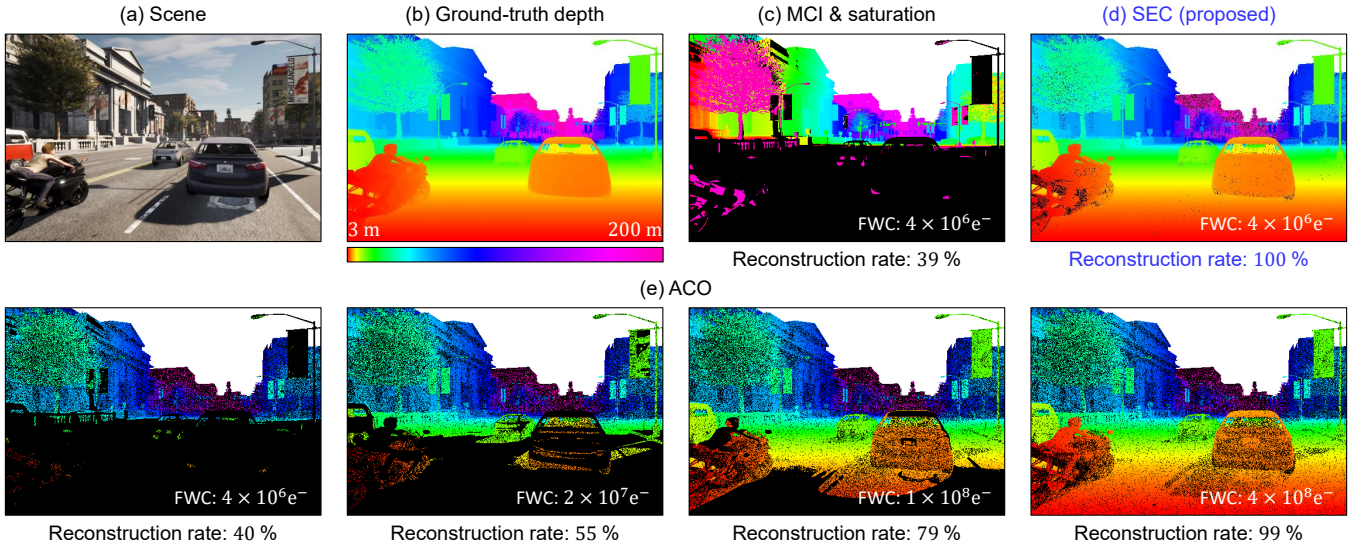

Fig. 12. **Long-range imaging under bright sunlight.** (a, b) A long-range outdoor scene under bright sunlight in the multi-camera environment and its ground-truth depth. (c) Reliable depth estimation is extremely challenging in these imaging conditions due to not only interference but also pixel saturation. (d) Our approach considerably mitigates interference and saturation, thus achieving high accuracy reconstructions. (e) ACO requires a high full-well capacity to overcome saturation in these scenarios. The three numbers below each depth estimation are the percent of inlier pixels that lie within $0.5\,\%$, $1\,\%$, and $2\,\%$ of the true depth. Outlier pixels above $2\,\%$ depth error and saturated pixels are shown in black.

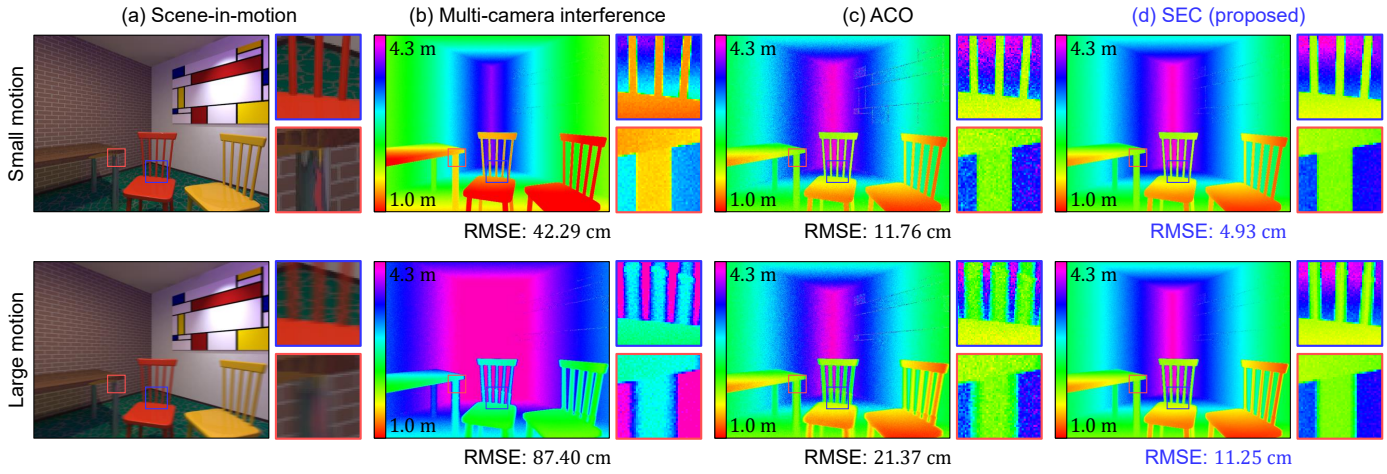

Fig. 13. **Dynamic scenes in multi-camera settings.** (a, b) If large motion exists between the scene and the capturing C-ToF camera in multi-camera environments, it causes large depth errors due to interference and motion artifacts. (c, d) Our approaches are more robust to both motion artifacts and multi-camera interference than ACO since the effective exposure time can be reduced while achieving higher SNR.

into multiple time slots and capturing the photons for only ON slots.

Comparisons with dynamic scenes: Motion between the scene and the C-ToF camera in multi-camera environments causes additional artifacts in depth estimation due to the mixed depth values for each pixel during the exposure time. These motion artifacts are common in all MCI reduction approaches when capturing dynamic scenes. One approach to mitigate the motion artifacts is to lower the exposure time. However, this comes at the cost of lower SNR of measurements. Our approaches are more robust to motion artifacts than the conventional ACO since the effective exposure time can be lowered while achieving higher SNR. In the worst case when the first and the last slots are ON, our approaches will cause the same amount of blur as ACO (but with lower noise). Figure 13 shows the results when there exist both MCI and motion between the scene and the capturing C-ToF camera. We assumed small and large camera translations along the x direction as shown in Fig. 13 (a). Large motions in multi-camera settings can cause large depth errors due to both interference and motion artifacts (Fig. 13 (b)). Our approach (SEC) reduces the error significantly since it lowers effective exposure time while increasing the SNR by removing both AC and DC interference (Figs. 13 (c) and (d)). See the supplementary report for the parameter values used for all simulation results.

8 Hardware Prototype and Experiments

We developed a proof-of-concept hardware prototype to implement ACO, SEC, and MLC. Our setup consists of four C-ToF cameras (OPT8241-CDK-EVM, Texas Instruments [6]) and four microcontrollers (Arduino UNO) to generate random binary sequences (Figure 14). The square waves at 50% duty cycle are used as the modulation and 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 MLC, four different modulation frequencies $\mathcal{F} = \{18, 20, 22, 24\}$ (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

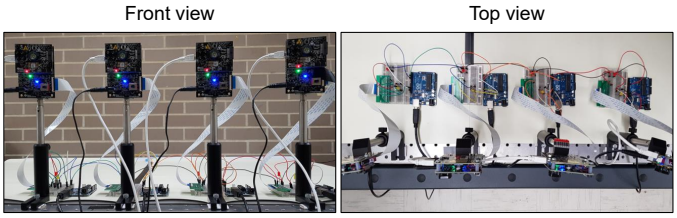

Fig. 14. **Hardware prototype.** Front and top views of our setup to implement ACO, SEC, and MLC. The setup consists of four C-ToF cameras and four microcontrollers to generate random binary sequences to activate the cameras by given slot ON probabilities.

MLC, 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 no-clash ON slots and all ON slots are averaged to obtain depth values for SEC and MLC, respectively. Since it is challenging to amplify peak power of the light source for SEC and MLC, we lower it for ACO instead using the ND-filters (NE10A-B, Thorlabs) with an optical density 1.0 (achievable A_0 is 8 in our cases), 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 [38]. We use the set of modulation frequencies $\mathcal{F} = \{18, 20, 22, 24\}$ (MHz) as the base frequencies, and $\mathcal{D} = \{27, 30, 33, 36\}$ (MHz) as the de-aliasing frequencies. Figure 15 shows the color image and ground truth depth map of a face mannequin along with interference result and estimated depth maps by three approaches. Depth at the regions with lowest 1% number of photons is 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 com-

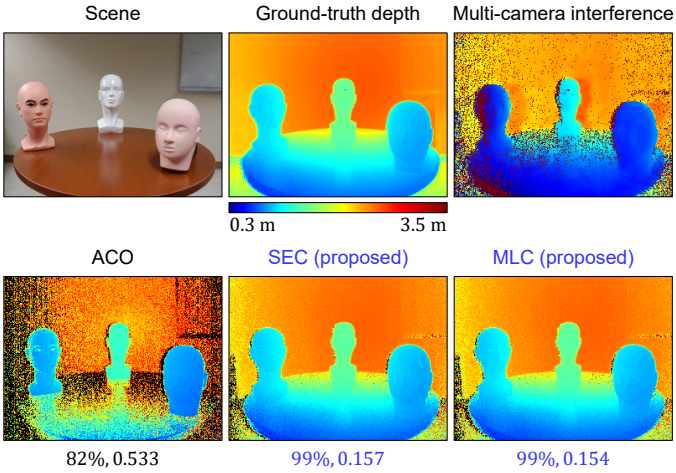

Fig. 15. **Performance comparison via real experiments.** Multi-frequency coding is used in the three different approaches. The % of inliers (non-black pixels) and RMSE values (in m) at the inliers are represented for comparison between approaches.

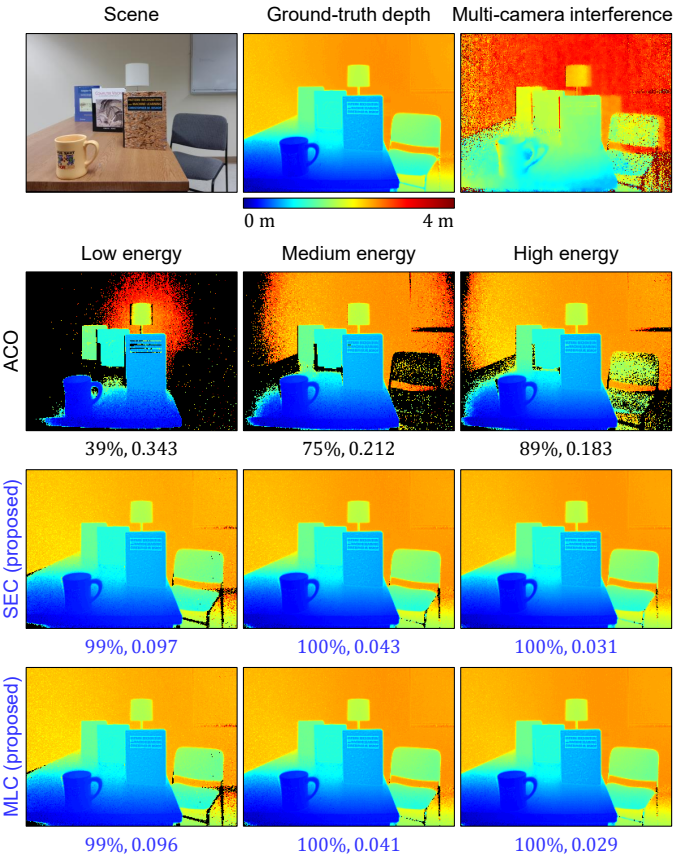

Fig. 16. **Depth estimation comparison over different energy consumption.** Our approaches show better performance at lower energy consumption than the conventional approach. The % of inliers (non-black pixels) and RMSE values (in m) at the inliers are represented for comparison between approaches.

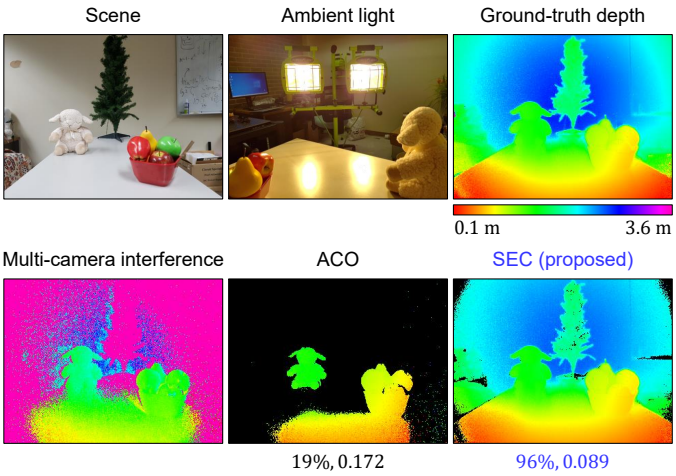

Fig. 17. **Robustness to ambient light.** In the proposed approaches (SEC and MLC), the saved energy by activating the C-ToF camera during only a fraction of integration time can be used to amplify the source peak power. This enables reliable depth estimation under strong ambient light. The % of inliers (non-black pixels) and RMSE values (in m) at the inliers are represented for comparison between approaches.

pare 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 $\mathcal{F}$ are used as the base frequencies. Figure 16 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.

Robustness to ambient light: Since a C-ToF camera is active during only a fraction of integration time in SEC and MLC, the saved energy can be used to amplify the source peak power, given a fixed energy constraint (when source peak power amplification is allowed in the system). The source peak power amplification achieves higher robustness to ambient light. In order to increase the effect of ambient light, we reduce the source energy such that depth of only foreground objects can be estimated robustly in ACO. The same source energy is used for ACO and our approach (SEC). As shown in Fig. 17, our approach provides more reliable depth estimation result than ACO under strong ambient light by a work lamp.

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 demonstrate the performance benefits of the proposed approach 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 a 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, the DARPA

REVEAL program, NSF CAREER Award 1943149, NSF award CNS-2107060 and Intel-NSF award CNS-2003129.

REFERENCES

[1] J. Lee and M. Gupta, "Stochastic exposure coding for handling multi-tof-camera interference," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7880–7888.

[2] Y. Li and J. Ibanez-Guzman, "Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 50–61, 2020.

[3] S. May, B. Werner, H. Surmann, and K. Pervolz, "3d time-of-flight cameras for mobile robotics," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. Ieee, 2006, pp. 790–795.

[4] J. Fischer, B. Huhle, and A. Schilling, "Using time-of-flight range data for occlusion handling in augmented reality." *IPT/EGVE*, vol. 109116, 2007.

[5] "Azure kinect dk, microsoft," https://azure.microsoft.com/en-us/products/kinect-dk/, accessed: 2023-08-04.

[6] "Time-of-flight sensors, texas instruments," https://www.ti.com/sensors/specialty-sensors/time-of-flight/products.html, accessed: 2023-08-04.

[7] "3d depth sensing development kits, pmd," https://3d.pmdtec.com/en/3d-cameras/flexx2/, accessed: 2023-08-04.

[8] C. S. Bamji, S. Mehta, B. Thompson, T. Elkhatib, S. Wurster, O. Akkaya, A. Payne, J. Godbaz, M. Fenton, V. Rajasekaran *et al.*, "Impixel 65nm bsi 320mhz demodulated tof image sensor with 3μm global shutter pixels and analog binning," in *2018 IEEE International Solid-State Circuits Conference-(ISSCC)*. IEEE, 2018, pp. 94–96.

[9] C. Bamji, J. Godbaz, M. Oh, S. Mehta, A. Payne, S. Ortiz, S. Nagaraja, T. Perry, and B. Thompson, "A review of indirect time-of-flight technologies," *IEEE Transactions on Electron Devices*, 2022.

[10] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein, "Computational imaging with multi-camera time-of-flight systems," *ACM Transactions on Graphics (ToG)*, vol. 35, no. 4, p. 33, 2016.

[11] R. Z. Whyte, A. D. Payne, A. A. Dorrington, and M. J. Cree, "Multiple range imaging camera operation with minimal performance impact," in *Image Processing: Machine Vision Applications III*, vol. 7538. International Society for Optics and Photonics, 2010, p. 75380I.

[12] D.-K. Min, I. Ovsiannikov, Y. Noh, W. Kim, S. Jung, J. Lee, D. Shin, H. Jung, L. Kim, G. Waligorski *et al.*, "Pseudo-random modulation for multiple 3d time-of-flight camera operation," in *Three-Dimensional Image Processing (3DIP) and Applications 2013*, vol. 8650. International Society for Optics and Photonics, 2013, p. 865008.

[13] L. Li, S. Xiang, Y. Yang, and L. Yu, "Multi-camera interference cancellation of time-of-flight (tof) cameras," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 556–560.

[14] B. Büttgen, F. Lustenberger, P. Seitz *et al.*, "Pseudonoise optical modulation for real-time 3-d imaging with minimum interference," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 10, pp. 2109–2119, 2007.

[15] B. Büttgen and P. Seitz, "Robust optical time-of-flight range imaging based on smart pixel structures." *IEEE Trans. on Circuits and Systems*, vol. 55, no. 6, pp. 1512–1525, 2008.

[16] T. Fersch, R. Weigel, and A. Koelpin, "A cdma modulation technique for automotive time-of-flight lidar systems," *IEEE Sensors Journal*, vol. 17, no. 11, pp. 3507–3516, 2017.

[17] B. Sklar, *Digital communications*. Prentice Hall Upper Saddle River, 2001, vol. 2.

[18] N. Abramson, "The aloha system: another alternative for computer communications," in *Proceedings of the November 17-19, 1970, fall joint computer conference*. ACM, 1970, pp. 281–285.

[19] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part i-carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE transactions on Communications*, vol. 23, no. 12, pp. 1400–1416, 1975.

[20] R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: motion deblurring using fluttered shutter," in *ACM transactions on graphics (TOG)*, vol. 25, no. 3. ACM, 2006, pp. 795–804.

[21] S. Achar, J. R. Bartels, W. L. Whittaker, K. N. Kutulakos, and S. G. Narasimhan, "Epipolar time-of-flight imaging," *ACM Transactions on Graphics (ToG)*, vol. 36, no. 4, p. 37, 2017.

[22] S. Fuchs, "Multipath interference compensation in time-of-flight camera images," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 3583–3586.

[23] D. Jiménez, D. Pizarro, M. Mazo, and S. Palazuelos, "Modeling and correction of multipath interference in time of flight cameras," *Image and Vision Computing*, vol. 32, no. 1, pp. 1–13, 2014.

[24] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin, "Phasor imaging: A generalization of correlation-based time-of-flight imaging," *ACM Transactions on Graphics (ToG)*, vol. 34, no. 5, p. 156, 2015.

[25] 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 (ToG)*, vol. 32, no. 6, 2013.

[26] S. Su, F. Heide, G. Wetzstein, and W. Heidrich, "Deep end-to-end time-of-flight imaging," in *IEEE CVPR*, 2018.

[27] N. Naik, A. Kadambi, C. Rhemann, S. Izadi, R. Raskar, and S. B. Kang, "A light transport model for mitigating multipath interference in time-of-flight sensors," in *IEEE CVPR*, 2015.

[28] J. Marco, Q. Hernandez, A. Munoz, Y. Dong, A. Jarabo, M. H. Kim, X. Tong, and D. Gutierrez, "Deeptof: off-the-shelf real-time correction of multipath interference in time-of-flight imaging," *ACM Transactions on Graphics (ToG)*, vol. 36, no. 6, pp. 1–12, 2017.

[29] S. Su, F. Heide, G. Wetzstein, and W. Heidrich, "Deep end-to-end time-of-flight imaging," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6383–6392.

[30] G. Agresti and P. Zanuttigh, "Deep learning for multi-path error removal in tof sensors," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0–0.

[31] G. Agresti, H. Schaefer, P. Sartor, and P. Zanuttigh, "Unsupervised domain adaptation for tof data denoising with adversarial learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5584–5593.

[32] G. Dong, Y. Zhang, and Z. Xiong, "Spatial hierarchy aware residual pyramid network for time-of-flight depth denoising," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*. Springer, 2020, pp. 35–50.

[33] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE Journal of quantum electronics*, vol. 37, no. 3, pp. 390–397, 2001.

[34] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (tof) cameras: A survey," *IEEE Sensors Journal*, vol. 11, no. 9, pp. 1917–1926, 2011.

[35] S. W. Hasinoff, "Photon, poisson noise." 2014.

[36] A. Stylogiannis *et al.*, "Continuous wave laser diodes enable fast optoacoustic imaging," *Photoacoustics*, vol. 9, pp. 31–38, 2018.

[37] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.

[38] D. Droeschel, D. Holz, and S. Behnke, "Multi-frequency phase unwrapping for time-of-flight cameras," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 1463–1469.

Jongho Lee received his B.S. degree in Electronic Engineering from Ajou University and M.S. degree in Electrical Engineering and Computer Science from Gwangju Institute of Science and Technology, South Korea, in 2006 and 2010, respectively. He is currently a doctoral student in Computer Sciences, University of Wisconsin-Madison. His research interests include computer vision and computational imaging.

Mohit Gupta is an Associate Professor of Computer Sciences at the University of Wisconsin-Madison. He received Ph.D. from the Robotics Institute, Carnegie Mellon University, and was a postdoctoral research scientist at Columbia University. He directs the WISION Lab with research interests broadly in computer vision and computational imaging. He has received the Marr Prize honorable mention at IEEE ICCV, a best paper honorable mention at IEEE ICCP, a Sony Faculty Innovation Award and an NSF CAREER award.