research

Understanding and handling interference across multiple active cameras.

BY JONGHO LEE, MOHIT GUPTA, BHUVANA KRISHNASWAMY, AND SUMAN BANERJEE

When Two Cameras Are a Crowd

VISION AND ROBOTICS systems enabled by cameras that recover 3D scene geometry are revolutionizing several aspects of our lives via technologies such as autonomous transportation, robotic surgery, and 'hands-free' user interfaces. Modern 3D cameras are active devices, where a programmable light source emits coded illumination. The emitted light gets reflected from the scene and is received by a sensor to infer the 3D structure of the surroundings. In a multi-camera environment, such active 3D cameras may receive light from the sources of other cameras, resulting in large depth errors. This problem is becoming increasingly important due to the emergence of low-cost and compact active 3D cameras, which are becoming ubiquitous across a wide range of applications, from consumer devices to vehicular vision systems.

We observe that the multi-camera interference (MCI) problem shares several similarities and dissimilarities with common interference problems in the RF domain. Based on this observation, this article describes new and emerging challenges when multiple active 3D cameras operate in the same spatio-temporal region. The article also outlines some solutions, and more importantly, highlights the next steps.

The 3D Revolution

We are in the midst of a 3D revolution fueled by cameras that can recover 3D geometry of their surroundings (Figure 1). The key catalyst driving this revolution is the emergence of low-cost, time-of-flight (ToF) 3D cameras that emit coded light and infer distances (depths) based on reflections from surrounding surfaces. ToF cameras can be made into extremely compact devices,[a] and thus, can potentially measure accurate 3D shapes over a wide area.

Applications. Due to their low cost, compact form factors and low computational complexity, ToF-based active 3D

---

a In comparison, 3D cameras based on other principles, such as binocular stereo and parallax, require large form factors and are unsuitable for long-range 3D imaging applications such as automotive navigation.

» **key insights**

- **3D cameras are revolutionizing several aspects of our lives in many applications, such as autonomous vehicles, cellphones, tablets, AR and VR devices.**

- **3D cameras reconstruct 3D geometry of the surroundings by emitting and receiving light. As multiple 3D cameras operate in the same spatiotemporal region, the light interference between cameras can cause large depth errors.**

- **Multi-camera interference (MCI) is becoming an important issue as 3D cameras become ubiquitous. The MCI problem shares similarities and dissimilarities with the interference problem in wireless communications.**

- **We describe several challenges and solutions for MCI based on these similarities and differences.**

Cameras that recover the 3D structure of their surroundings are becoming ubiquitous in several application domains, including transportation, augmented and virtual reality, mobile robotics, and phones.

MS HoloLens

Apple iPad

Transportation   **Augmented reality**   **Robot navigation**   **Mobile phones**

Figure 2. Multidevice interference and the resulting depth errors in active 3D cameras.

(a) An active 3D camera consists of a programmable light source and a sensor.
(b) Multiple active 3D cameras simultaneously imaging the same scene point leads to MCI. (c) The estimated scene depths can be significantly different from the true depths due to interference. (d) An example scene.
(e) Ground-truth scene depths. (f) Depths measured in the presence of an interfering camera have large systematic errors. *Part of the figure adapted from Lee and Gupta[19] with permission.*

Sensor   Reflection   3D scene

Light source   Emission

(a) Active 3D camera

— True reflection   — Interfering reflection   — Total reflection

3D scene   True depth   Recovered depth

3D camera

Interfering 3D camera

Intensity   Time

(b) Multiple active 3D cameras   **(c) Depth errors by interference**

Large systematic errors   4m

0m

(d) Scene color image   **(e) Scene depths: no interference**   **(f) Scene depths: with interference**

cameras are now the method of choice in most commercial 3D imaging systems, including vehicular LiDARs, and more recently in commodity[1,2] and consumer-grade devices, such as cell phones, tablets (for example, Apple iPad), and gaming and AR/VR headsets (for example, Microsoft Kinect and Hololens). They are also used by inspection and delivery robots to navigate in warehouses[24] and perform 3D modeling,[9,13] directly impacting safety and functionality. Going forward, our dependence on them is only going to grow as these cameras will drive an even larger, potentially transformative set of applications such as human-machine interaction[6] and home robotics.

Active 3D cameras and interference. More generally, ToF cameras belong to the class of *active 3D cameras*, which consist of a *programmable light source* that emits spatially or temporally coded light.[b] For example, the light source could be a laser or an LED whose intensity is modulated over time. The emitted light signal travels to the scene of interest, gets reflected, and is captured by the sensor (typically co-located with the light source), as shown in Figure 2a. Scene depths (and hence, the 3D geometry) are extracted by comparing the emitted and the reflected light. Since the camera actively controls the illumination, it can reliably recover highly precise 3D geometry even in challenging real-world scenarios, including large lighting variations (bright sunlight to dark night sky) and optically uncooperative scenes (shiny materials, textureless and dark objects), which are otherwise difficult to handle for passive 3D cameras.

Although the coded light sources enable high-precision 3D recovery in isolation, **multiple active cameras**, each emitting their own light signals, can cause **mutual interference**. In such multi-use scenarios, an active camera's sensor may receive light emitted not just by its own source, but also by the sources of other cameras. This interfering signal prevents correct 3D depth estimation, resulting

---

b In contrast, passive 3D cameras (for example, stereoscopic 3D cameras) do not actively control the illumination, thus only passively observing the scene.

in potentially large, systematic depth errors. An example is shown in Figure 2b-f, where the scene depth observed by one active camera is significantly distorted in the presence of a second, interfering one. Such errors in low-level 3D measurements can make it near impossible to extract any actionable information (for example, object detection and recognition, image classification) for downstream machine-learning (ML) algorithms.

Why is now a good time to address MCI? Until recently, a large proportion of ToF cameras were based on a scanning laser beam that illuminates one (or a few) scene point at a time. Since they illuminate only a small portion of the scene at a time, the probability that two devices simultaneously illuminate the same point is relatively low (Figure 3a). This enabled multiple devices to co-exist without affecting each other. However, they require mechanical moving parts (for example, rotating mirrors), often leading to long acquisition times, high cost, bulkiness, and reliability issues.

To address its limitations, a new emerging class of solid-state ToF cameras *flood-illuminate* the entire scene (Figure 3b), making them considerably cheaper, smaller, lighter, and faster. Hence, solid-state cameras are fast replacing scanning-based cameras in autonomous driving and robotics applications (Figure 3c). But, there is a trade-off: Several active cameras flood-illuminating a scene simultaneously will interfere with each other (Figure 3b). Furthermore, a new generation of consumer devices (for example, cell phones) with 3D imaging capabilities are becoming ubiquitous. Due to stringent size and cost requirements, these devices also prefer solid-state 3D cameras (Figure 3d), which will create strong MCI whenever multiple active cameras are used in proximity.

Due to their compatibility with mainstream CMOS fabrication lines, the capabilities of solid-state cameras (spatial resolution, timing precision, signal-to-noise-ratio) continue to grow rapidly.[10,23,33] Therefore, these cameras are quickly becoming the method of choice in almost all applications that rely on 3D cameras, including vehicular navigation, robotics, and consumer mobile devices.

On the downside, this growth will add to the severity of MCI problems. Imagine several cars equipped with ToF camera-based LiDARs driving in close proximity, or an indoor setting with several consumer devices such as phones and headsets using their 3D cameras simultaneously for augmented reality (AR), localization, or 3D modeling.[9,13] It is, therefore, critical to address this problem now so that 3D cameras continue to function reliably across a broad spectrum of real-world scenarios, imaging devices, and application domains.
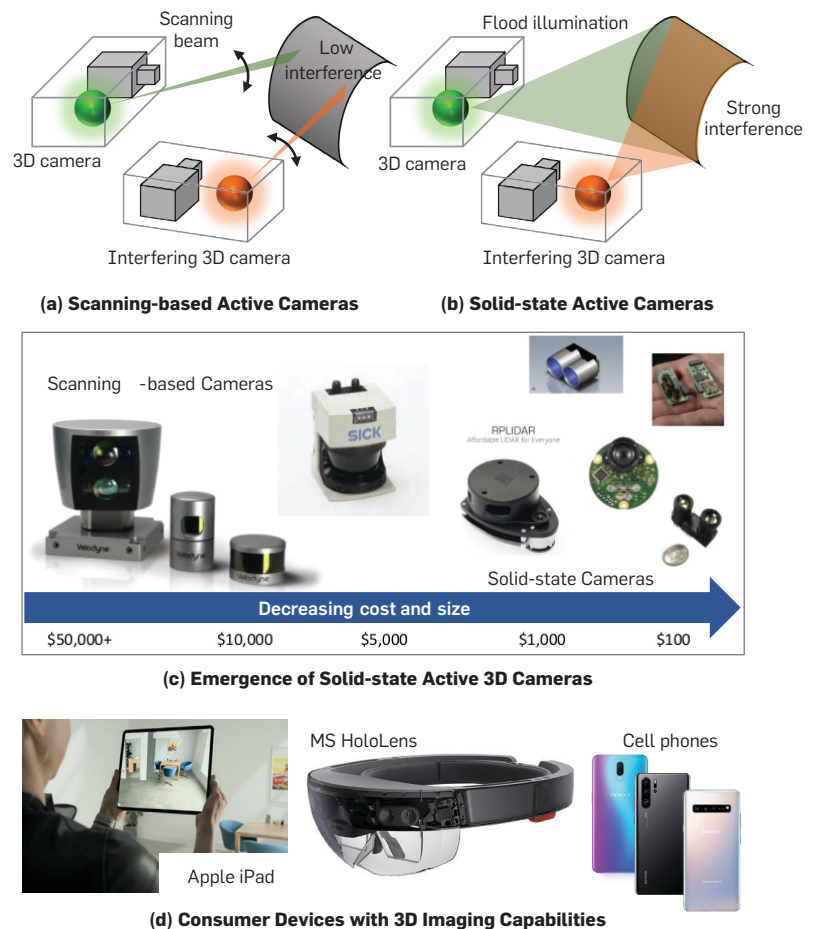
Therefore, managing interference across a multitude of these devices will be of profound importance as solid-state ToF cameras become ubiquitous in our personal mobile devices, vehicles, homes, and workplaces, as well as the basis for numerous critical applications. The goal of this article is to shine a light on this important problem (pun intended) that has received little attention so far: *interference of light signals emitted by active cameras and its impact in accurately recovering 3D scene information.*

Can research on wireless interference be used to address MCI? Typical wireless communication systems

---

Figure 3. Emergence of solid-state ToF and consumer-grade 3D cameras.

(a) With current LiDARs using a scanning laser beam, only a small portion of the scene is illuminated at a time. Thus, the likelihood of MCI is relatively low. (b) A new class of solid-state ToF devices is emerging where the light source flood-illuminates the entire scene without any moving parts, thus significantly increasing the likelihood of interference. (c) This new generation of solid-state 3D cameras is cheaper, smaller, and lighter, and they are fast replacing scanning-based devices. (d) Consumer devices with solid-state 3D cameras are becoming ubiquitous, increasing the possibility of strong interference in scenarios when multiple such devices are used simultaneously in a spatial neighborhood.

(a) Scanning-based Active Cameras

(b) Solid-state Active Cameras

(c) Emergence of Solid-state Active 3D Cameras

(d) Consumer Devices with 3D Imaging Capabilities

(for example, Wi-Fi, cellular systems) use radio waves propagating over a shared channel. These systems have dealt with multi-node interference for decades. This has led to the development of many techniques using whichever multiple wireless devices can co-exist and communicate over the same shared air medium. Hence, it is natural to evaluate whether strategies developed for mitigating wireless interference can be applied to address MCI. To answer the above question, we first discuss the similarities and differences of how interference manifests in these settings.

At first glance, an active 3D camera is very similar to a wireless transmitter (Figure 4). A wireless transmitter generates radio waves that propagate over the air (usually referred to as the channel/medium) to a wireless receiver located at a distance with a goal of communicating information. The information itself is encoded in the radio waves. An active 3D camera also uses a transmitter and a receiver. The active 3D camera transmitter (or light source) generates a signal (light), and a corresponding receiver (or sensor) captures a reflected instance of this signal. However, unlike the wireless communication systems, the goal of active 3D cameras is to estimate the depth, or distance, of objects in the environment based on the ToF of the received reflected signals. This could be considered as analogous to learning the channel in wireless systems.

The similarities further extend from a single transmitter-receiver (transceiver) pair to a network of transceivers. In a wireless network, when multiple nodes attempt to communicate simultaneously, the radio waves interfere with each other, leading to collisions. Similarly, when multiple light sources flood-illuminate a scene, they collide and corrupt the received signal, resulting in depth errors.

There are, however, two critical differences between these two systems. First, in active 3D cameras, the **transmitter and the receiver are co-located**, that is, the light source and the sensor are both part of the same active 3D camera (Figure 4b). Second, signals collide in different ways. In a wireless channel, collisions could result in constructive interference or destructive interference, where the signal gets strengthened or weakened. On the other hand, an active 3D camera modulates the *intensity* of light, as opposed to the underlying electric field. The key factor to notice is that the intensity of light is always positive with both a constant (DC) and a time-varying (AC) component; the depth is encoded in the time-shift of the AC component. Therefore, the interfering signals from multiple cameras always accumulate additively. Although orthogonal-coding approaches can remove the AC interference, the DC component still accumulates, resulting in higher photon noise.

Despite these differences, we believe that the experiences of the wireless communications and networking community in addressing interference problems can benefit the imaging and computer-vision community in addressing MCI. In particular, we believe that a good understanding of various approaches to mitigating wireless interference and the theoretical foundations and practical considerations of

Figure 4. Wireless and multi-camera environments.

(a) In an RF network, transmitter (Tx) nodes send signals via the channel to the receiver (Rx). (b) In a multi-camera environment, multiple active 3D cameras transmit light from their light sources (Tx) to the 3D scene (channel) and receive the reflected light by the sensor (Rx).

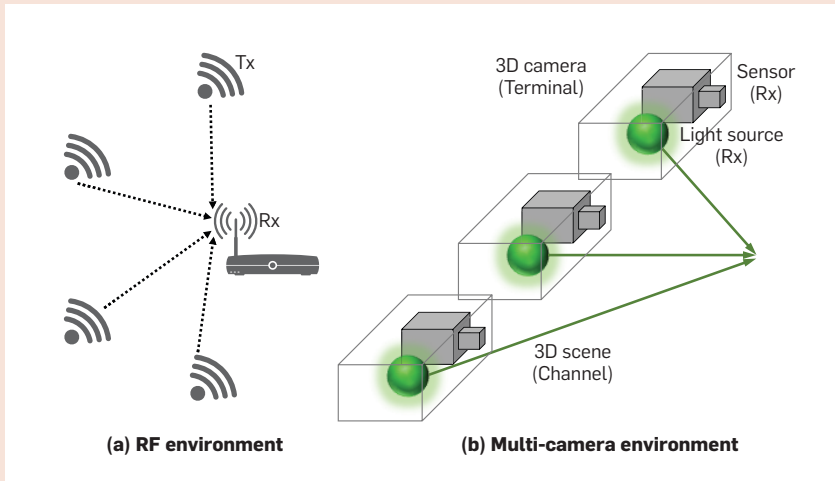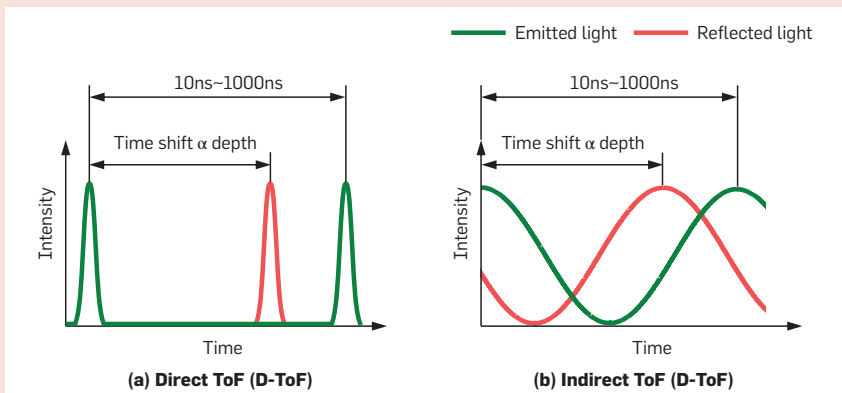(a) RF environment   (b) Multi-camera environment

Figure 5. Active 3D cameras.

(a) In direct ToF (D-ToF) approaches, the light source emits a periodic train of short light pulses. Scene depths are estimated by directly measuring the travel time of each pulse using high-speed timing circuits. (b) In indirect ToF (I-ToF) approaches, the light source emits continuously modulated periodic light waveforms. Depths are estimated by measuring the phase-shift between the emitted and reflected waveforms.

(a) Direct ToF (D-ToF)   (b) Indirect ToF (D-ToF)

active 3D camera design can together synthesize useful solutions. If successful, the resulting ideas will not just spur widespread adoption of existing technologies, but also enable emerging applications that were hitherto considered impossible. The theoretical tools and techniques developed as part of this work will find applications in a broad range of techniques which involve coded light sources and sensors, such as structured light, tomography, and microscopy, as well as optical communication systems using coded light sources.

In summary, we answer the question in this subsection affirmatively by indicating that research on wireless interference can be used to address MCI. In the rest of the article, we identify various MCI mitigation approaches by leveraging its similarity to wireless and propose new directions to address the differences between the two scenarios.

An Overview of Active 3D Cameras

An active 3D camera consists of a light source that emits coded illumination toward the scene and a sensor that captures the reflected light, as shown in Figure 2a. The most widely used class of active 3D cameras is based on the ToF principle. ToF-based cameras have a light source which emits temporally coded illumination. For example, the light source could be a laser or a light-emitting diode (LED) that sends out short light pulses or a continuously modulated light. The emitted light travels to the scene of interest and is reflected back to the sensor. The cameras measure the scene depths by measuring the total time of travel, computing the time-shift between the emitted and received waveforms (Figure 5a-b).

ToF-based depth-imaging systems can be broadly classified into direct and indirect ToF systems. A direct ToF (D-ToF) system[16,34] estimates scene depths by emitting a short light pulse into the scene and directly measuring the travel time of the reflected pulse (Figure 5a). Most vehicular LiDARs are based on the D-ToF principle. An indirect ToF (I-ToF) system,[11,18,27] on the other hand, emits light continuously. The intensity of its light source and the exposure of the sensor are both modulated over time for measuring the scene depths.

The multi-camera interference problem shares several similarities and dissimilarities with common interference problems in the RF domain.

The light-source-modulation and sensor-demodulation functions can be any periodic and continuous functions such as sinusoids, square waves, or trapezoidal functions (Figure 5b).[12] I-ToF cameras do not require expensive components, and hence, are routinely used in consumer devices such as Microsoft Kinect and Hololens.

MCI in active 3D cameras. When multiple active 3D cameras illuminate a scene, the reflected signals interfere, corrupting the scene information. In this section, we provide the mathematical background for understanding MCI in active 3D cameras.[c]

ToF image formation model. The intensity of the light source in a ToF camera is temporally modulated as a continuous periodic function $M(t)$; it could be a sinusoid,[18,27] or an impulse train function.[17] The period of $M(t)$ typically varies from 10–1,000ns, which corresponds to a measurable distance of 1–100m. The light emitted by the source travels to the scene of interest and is reflected back toward the camera. The intensity of the reflected light incident on a sensor pixel **p** is a time-shifted and scaled version of $M(t)$:

$$R(\mathbf{p}; t) = \alpha M(t - \tau), \qquad (1)$$

where $\tau = \frac{2d}{c}$ is the time-shift of the waveform due to travel from the source to the sensor. d is the distance between the camera and the scene point imaged at **p** and c is the light speed. α is a scene-dependent scale factor that encapsulates the scene's reflectance properties. The camera computes τ (typically of the order of ns) by using high-speed, on-chip timing circuits, and the scene distance is estimated as $d = \frac{c\tau}{2}$.

Multi-camera interference. If multiple ToF cameras are simultaneously illuminating and imaging a scene point (Figure 2b), the brightness of the light incident at one of the cameras (referred to as the primary camera) is given by:

$$R_{mult}(t) = R(t) + MCI,$$
$$\text{where, MCI} = \sum_{n=1}^{N} R_n(t) \qquad (2)$$

where N is the number of interfering cameras, $R(t)$ is the radiance incident at the primary camera due to its own source (Eq. 1), and $R_n(t) = \alpha_n M_n(t - \tau_n)$ is the measured intensity due to the nth

---

c ToF image formation model is applicable to any active 3D camera.

source. We drop **p** for brevity. The summation term in Eq. 2 corrupts the true radiance $R(t)$, resulting in erroneous depth estimates. Figure 2c shows an example of a ToF camera using sinusoid modulation. Assuming all the sources use sinusoids (or any other periodic shapes such as squares) of the same frequency, the phase of $R_{mult}(t)$ may differ from the true phase of $R(t)$, resulting in *systematic*, potentially large depth errors as shown in Figure 2d-f.

Current approaches to addressing MCI. A trivial approach to prevent MCI is to assign different wavelengths to different cameras. This approach only eliminates AC interference, and it faces two practical constraints: (1) The set of available wavelengths is strongly limited by the sensitivity range of the sensor material (typically silicon), as well as the practical requirement for the emitted light to be invisible to humans. This limits the available wavelengths to be ≈ 850-950nm, which is the near-infrared region of the EM spectrum; (2) Due to laser and sensor hardware constraints, each sensor must be assigned a range of wavelengths (for example, 5-10nm). These constraints restrict the number of distinct wavelength bands preventing assigning a unique set of wavelengths to each active camera. Recent works addressing MCI can be broadly classified into three categories:

Orthogonal coding. The majority of existing works rely on orthogonal coding, such as sinusoids of different modulation frequencies[30] or phases,[20,36] and pseudo-noise sequences[7,8] for different cameras. However, they face challenges similar to frequency division multiple access (FDMA): a limited set of orthogonal frequencies and codes which proves inadequate for the rapidly growing cameras. These approaches often require a central authority that assigns a unique code to each camera, which is not practical.

Time division multiple access (TDMA). Other approaches divide the total capture time of the camera into multiple time slots and assign them to individual cameras randomly.[19] These techniques do not scale with the number of interfering cameras.

Mechanical approaches. Another method is to project a planar light

Several active cameras flood-illuminating a scene simultaneously will interfere with each other.

sheet which is scanned over the scene. Since only a portion of the scene is illuminated at a time, interference can be reduced.[4,26,35] Although these approaches successfully prevent interference, they require mechanical scanning, which increase system cost and size.

Recently, optical phased arrays and micro-electro-mechanical systems (MEMS)-based technologies have attracted attention to resolve the limitations of mechanical approaches. These hardware-based approaches can be integrated into our software/firmware-based approaches to create hybrid solutions to mitigate MCI. However, this article focuses on wireless-inspired software/firmware approaches to reduce MCI.

The Promise and Pitfalls of Wireless-Inspired Approaches

Interference is a classical problem in wireless communication. A huge body of literature over multiple decades has examined this problem to enable efficient shared access of a common channel such as the wireless medium. The wireless community commonly refers to these approaches as medium access control (MAC) techniques. Since MCI arises from sharing the common medium, it has similarities to MAC protocols in wireless communication. Here, we showcase the promise and pitfalls of applying well-known wireless MAC protocols to address MCI. As previously discussed, orthogonal coding across cameras can only reduce AC interference and not DC interference. A simple solution is to avoid collisions; we adapt a coordinated and distributed collision avoidance from wireless, simulate them in a multi-camera setting, and analyze their performance.

Comparisons using computer simulations of MCI. We developed a physically accurate computational simulator to emulate active 3D cameras, under a wide range of scene configurations and sensor-source parameters. The simulator models different steps of the image formation process as shown in Figure 6. Given a source modulation function $M(t)$ (≥ 0), the emitted light $E(t)$ is defined as $E(t) = P_s M(t)$, where P_s is average source power. $M(t)$ can be any periodic function. For example, for

sinusoid coding, $M(t) = 1 + \cos(2\pi ft)$, where f is modulation frequency. The light signal $L(t)$ received at the sensor is defined as the sum of source reflection at the scene and ambient light:

$$L(t) = \alpha E\left(t - \frac{2d}{c}\right) + P_a, \quad (3)$$

where α is a scale factor encapsulating scene reflectance and light fall-off, P_a is average ambient power, d is scene depth, and c is light speed. Noise-free sensor measurement C is the correlation between $L(t)$ and sensor demodulation $D(t)$:

$$C = \int_\tau L(t)D(t)\,dt, \quad (4)$$

where τ is integration time. If we assume sinusoid coding, $D(t) = 1 + \cos(2\pi ft)$. Final sensor measurement is obtained by adding photon noise, read noise, and ADC noise to C. To recover scene depth, we need multiple sensor measurements obtained by changing the phase of $D(t)$. For a 4-tap sinusoid coding scheme, those measurements C_k are obtained from Eq. 4 with $D_k(t) = 1 + \cos\left(2\pi ft - \frac{\pi}{2}(k-1)\right)$, $(k = 1, \ldots, 4)$. The scene distance estimate $\hat{d}$ is obtained by

$$\hat{d} = \frac{c}{4\pi f}\tan^{-1}\left(\frac{C_2 - C_4}{C_1 - C_3}\right) \quad (5)$$

Using this simulator, we compare the performance of the following two wireless-inspired MCI reduction approaches with the conventional orthogonal coding approach (OCA). We implement OCA by assigning orthogonal modulation frequencies to each camera. Visit https://bit.ly/3MhG4qu to access the code to simulate MCI reduction approaches.

CSMA-based MCI mitigation. A popular idea in many communication systems to share a common medium is the "listen before talk" approach. In this approach, interference is potentially mitigated by requiring each transmitter to listen to the channel before transmitting their data. Waiting until the channel is idle ensures that a transmitter does not interfere with an ongoing communication. In wireless parlance, this is called "carrier sensing," and the technique is often referred to as carrier sense multiple access (CSMA). To ensure no two nodes attempt to transmit simultaneously once the channel is sensed to be idle, CSMA with collision avoidance (CSMA/CA) is widely used in wireless

systems such as Wi-Fi. We draw inspiration from CSMA/CA, which requires nodes to wait a random amount of time after the channel is idle before transmitting. Combining carrier sensing and random wait times helps to avoid collisions. In particular, we design a "listen/view before illumination" approach at each camera to avoid interference from other cameras. Before illuminating the scene, if the channel is busy, the camera defers; else, it illuminates the scene and measures the depth information. In our implementation, we divide the total capture time (minimum time to estimate depth) of an active 3D camera into multiple time slots and sense the channel at each slot.

Co-operative random access-based MCI mitigation (CRA). We compare the depth accuracy of CSMA-based MCI with a time-division multiple access (TDMA) approach. If multiple cameras are synchronized to a global clock, their transmissions can be scheduled in fixed slots. Each camera chooses a random sequence of slots to be active and illuminate the scene. In our implementation, all the cameras are synchronized, such that the slot boundaries are the same.

In Figure 7, we compare OCA, CRA, and CSMA in terms of depth accuracy and power consumption as a function of the number of interfering cameras. For depth accuracy comparison at the same source power, we employ depth standard deviation since all compared approaches cause random depth errors due to noise, instead of structured errors. For both CSMA and CRA, we use a slot clash check algorithm[19] to check if slot interference occurred, and depth is estimated from the collection of non-clashed slots. Depth standard deviation of each MCI approach is computed from repeated

---

Figure 6. Simulation steps.

We developed a physics-based simulator that computationally emulates various steps of an active 3D camera's imaging process for evaluating the proposed interference mitigation techniques.
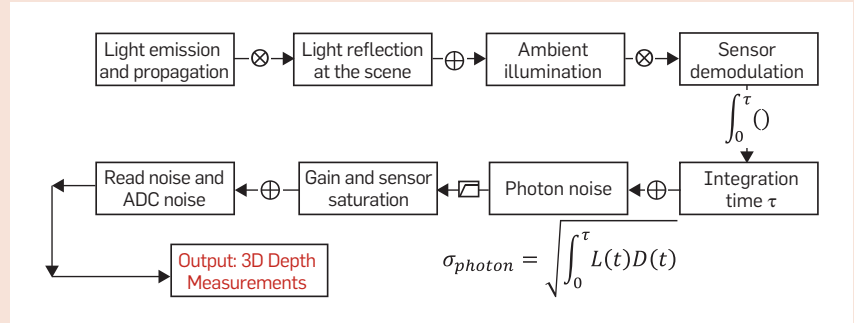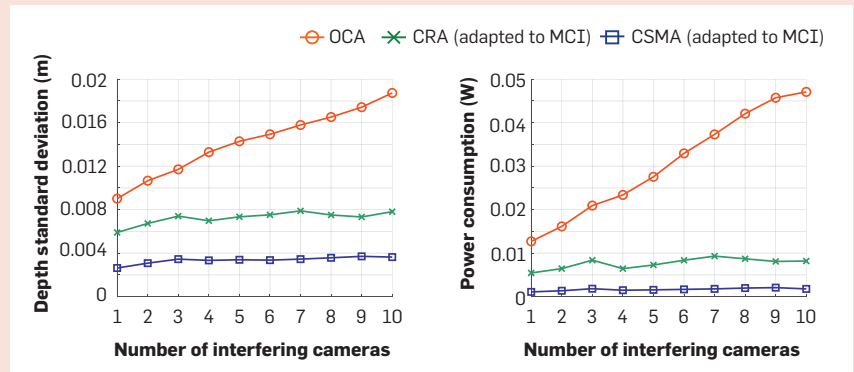

$$\sigma_{photon} = \sqrt{\int_0^\tau L(t)D(t)}$$

Figure 7. Depth standard deviations and power consumption of MCI approaches.

Simulation results show that CSMA-based and cooperative random access-based MCI outperform conventional orthogonal coding.

depth estimation. The required power consumption to achieve the same depth standard deviation is also compared. CRA and CSMA, the wireless-inspired approaches, outperform OCA in terms of both depth accuracy and power consumption, as they avoid collisions and hence both AC and DC interference is reduced. Performance improves with the number of cameras, which is desired when active 3D cameras become more prominent in the near future.

Potential pitfalls. Although CRA and CSMA show promising results, the comparisons are based on purely theoretical and idealized simulations. In practice, additional overhead is required to account for various resources, such as power and time for clash check and carrier sensing. Furthermore, adapting wireless-based approaches for MCI has a tradeoff: The probability of collision increases as the network scales. To avoid collisions, more slots should be used, which leads to greater capture time and power consumption. The time to switch from carrier sensing to normal camera mode will further increase slot duration. These can be critical issues for real-time applications, such as vehicular networks. It is also challenging to synchronize multiple cameras to a global clock due to the absence of a central controller. Therefore, to accommodate the constraints of an active 3D camera system, careful algorithm design and post-processing are required. In the next section, we present open research areas that can address MCI using approaches inspired by wireless communication.

The Road Ahead: Challenges in Wireless-Inspired MCI Solutions

We identify the following research directions, each inspired by the rich work on MAC protocols in wireless networks, with open challenges in applying them to active 3D camera networks. We build upon existing approaches to be applied to active 3D cameras by leveraging the opportunities provided by cameras to address their unique constraints.

Distributed interference management. In large-scale active 3D camera networks, resource-efficient, distributed interference-management techniques are necessary. We propose spread spectrum strategies that can estimate ToF from the cumulative reflected signal, enabling **multiple cameras to co-exist**. Spread spectrum techniques are used to improve a sender's resilience to interference and enable co-existence with other transmitters by occupying a wider range of frequencies. Since throughput is not a metric of interest in an active 3D camera network, spread spectrum-based approaches are an appropriate choice for MCI.

One of the key opportunities to leverage in active 3D cameras is the co-existence of the transmitter and the receiver; this enables it to maximize the potential of spread spectrum without communication overheads. Most existing active 3D cameras can modulate multiple frequencies to achieve both high precision and large depth range (for example, Microsoft Kinect uses 120MHz, 80MHz, and 16MHz[25]). On the other hand, one of the b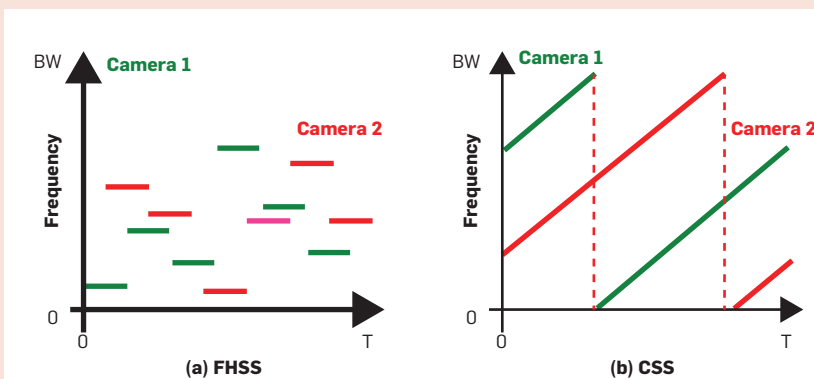iggest challenges is the need and ability to cancel background noise. As previously explained, *unlike wireless, orthogonal frequencies and codes still contribute to ambient noise* in an active 3D camera. A redesign of spread spectrum is needed to address this challenge. We explore two spread spectrum techniques (that are widely used in wireless) and discuss their challenges in applying to MCI.

Randomized frequency hopping spread spectrum. Frequency hopping spread spectrum (FHSS) is a spread spectrum technique where the carrier frequency of the transmitter follows a *hopping pattern* (Figure 8a). Due to its robustness to broadband interference, adaptive FHSS is used in Bluetooth, which enables Bluetooth devices to co-exist with Wi-Fi devices. Bluetooth devices follow a pseudo-random hop sequence to avoid interference, where the hopping pattern is known to the transmitter and the receiver. We adopt FHSS to reduce MCI. In this design, each light source randomly chooses a pseudo-random hopping pattern and hop duration. Since the light source and the sensor are co-located, the hopping pattern is known to both of them. The hopping pattern can also be varied without additional overheads to exchange the hopping pattern. Therefore, the sensor keeps track of ToF of the expected frequencies periodically. The cumulative ToF of the corresponding pattern is then used to estimate the depth of the entire scene.

While FHSS reduces the probability of two cameras colliding in a given slot, unlike RF signals, the intensities of light signals from interfering cameras add up, leading to DC interference even when the hopping patterns of two cameras do not match. With the help of successive interference cancellation, we propose to iteratively cancel DC interference across the frequency bands. For example, consider two cameras using FHSS colliding in three slots. The probability of the two cameras using the same frequencies in the same slots in round two is significantly lower. This probability can be further reduced by increasing the scan duration and introducing empty slots where a light source does not send anything.

Randomized chirp spread spectrum (CSS). Chirp signals have been used

Figure 8. Distributed Interference Management approaches.

(a) FHSS

(b) CSS

for precision ranging in RADAR,[15] and more recently, CSS modulation has been used in LoRa[21,22] to enable long-range communication. CSS spreads energy by linearly increasing the frequency of operation over time. Since this linear increase makes it robust to interference, it can be leveraged to enable multiple cameras to coexist: Each light source is assigned a unique starting frequency such that interference from other sources does not affect the AC component of the reflected signal. Figure 8b illustrates a chirp signal transmitted by a source, where the frequency sweeps the entire assigned bandwidth in a given duration.

The probability of collision is inversely proportional to the number of chirps. It is therefore desirable to have a large pool of chirps to reduce collisions. While decreasing step size will increase the number of chirps, the minimum step size is limited by hardware switching speeds. To address this challenge, non-linear CSS can be explored. A non-linear chirp with initial frequency f_i is similar to that of linear chirp in Figure 8b, whose frequency progression with time is given by a function $g(t)$. By choosing orthogonal non-linear functions, the probability of collisions can be reduced.[15,22] Similar to FHSS, CSS also must consider DC interference due to cumulative energy from other frequencies. We propose to successively cancel DC interference from one band to another and, leveraging the lack of time synchronization, this offers more possibilities to explore in MCI.

Centralized, networked coordination. While distributed approaches to MCI are necessary for many applications, there are multiple environments that lend themselves well to networked coordination with centralized control. For example, a factory floor with mobile autonomous robots operating in a single administrative domain can easily be coordinated through a central server. In such indoor applications, RF-based wireless connectivity to network devices via a central server that helps with synchronization could be integrated with the cameras. Therefore, combining RF and camera-based networks could offer real-time control of the cameras to

As solid-state ToF cameras become ubiquitous in our personal mobile devices, vehicles, homes, and workplaces, managing interference across a multitude of these devices will be of profound importance.

a central server. Vehicles with LiDARs on roadways could also accomplish similar goals with roadside infrastructure providing the control function. Finally, home environments are likely the most challenging as there might be multiple environment sensing devices from diverse manufacturers, but if standardization of centralized coordination were to be developed, then all such devices could coordinate through a common hub within the home to manage active camera activities. We propose the following approaches in such centralized coordination scenarios.

Creating a conflict graph in real-time. To schedule simultaneous transmissions in wireless systems, one needs to infer the "conflict graph" that maps the potential of interference between transmitters.[3,28,29] Creating such a conflict graph in real time in the active camera domain is more challenging since the goal is to learn the channel. One possible approach is to time-synchronize different light sources and arrange them to send sustained pulses, sometimes in tandem and sometimes in isolation. If there are differences in received outcomes, we may conclude that such transmit-receive pairs interfere. Of course, a challenge lies in achieving various synchronization accuracies. We can combat them by adaptively choosing pulse durations and pulse structures that outlast potential clock synchronization errors. Other sources of inaccuracies can stem with other transmitters in the vicinity that are operating under the control of the centralized coordinator. Time-series analysis to identify and eliminate such interferers is an interesting research problem toward a practical deployment of a centralized coordination. A bigger challenge is the ability to meet the same goals as in a passive setting, that is, by simply observing activities on the channel and using time information to determine the same information. This, however, may be effective if all 3D cameras are sufficiently active. Overall, a hybrid passive-active method would optimize the best of both alternatives.

Networked schedule of activity under centralized control. Prior work on centralized WLANs has shown that

it is possible to create efficient partial-traffic scheduling for best performance.[31,32] Traffic scheduling is particularly effective when a vast majority of the traffic can be unscheduled, as it can be addressed through simple distributed mechanisms. Centralized control can benefit the fraction that cannot be addressed effectively by those mechanisms. Such traffic scheduling can also be used to minimize collisions in the active camera environments. The real-time conflict graphs discussed previously will inform a centralized coordinator which transmitters might benefit from scheduling while allowing the remaining to be unscheduled. Note that scheduling requirements might depend on specific environments and the frequency with which transmitters need to rediscover their 3D environment. It also depends on how frequently the environment changes and external, out-of-band information can be used for this purpose. Based on this information, each transmitter-receiver pair that cannot be managed using a distributed approach may be scheduled to time slots to accomplish their scene-sensing goals. Note that a sensing attempt may be scheduled (or not), depending on the other devices attempting to sense concurrently.

Conclusion

We strongly believe that due to their impending growth for consumer applications and their similarities to wireless communications, active 3D camera networks will benefit from new research from wireless experts. In particular, as interference between these cameras grows, sophisticated techniques are needed to handle MCI. Due to some fundamental differences between active cameras and wireless networks identified in this work, careful considerations in designing interference mitigation, cancellation, and coordination approaches are needed. It is interesting to note that some of the early work on interference-mitigation strategies in wireless environments (MACA,[14] MACAW[5]) started by exploring the potential use of wired interference-mitigation strategies in the wireless environment (Ethernet and related CSMA/CD approaches) and their consequent limitations as well as through the process discovering improvements that led to eventually useful solutions. We believe that a similar approach can be taken by using the state-of-the-art wireless protocols as a starting point to uncover how similar strategies can be designed for active 3D camera systems.

We believe this problem domain is ripe for multiple communities to address collaboratively. Successful and effective solutions to this problem can have a big impact in many applications in our homes and society.

References
1. 3D Time-of-Flight (ToF) Sensors, Texas Instruments; https://bit.ly/3Fy0yaD.
2. 3D ToF Development Kits, PMD; https://pmdtec.com/picofamily/.
3. Ahmed, N., Ismail, U., Keshav, S., and Papagiannaki, K. Online estimation of RF interference. In *Proceedings of the 2008 ACM CoNEXT Conf.* (December 2008), 1–12.
4. Achar, S. et al. Epipolar time-of-flight imaging. *ACM Transactions on Graphics (ToG) 36*, 4 (2017), 37.
5. Bharghavan, V. et al. MACAW: A media access protocol for wireless LANs. In *Proceedings of ACM Sigcomm 24*, 4 (Oct. 1994).
6. Breuer, P. et al. Hand gesture recognition with a novel IR time-of-flight range camera–a pilot study. In *Proceedings of the Intern. Conf. on Computer Vision/Computer Graphics Collaboration Techniques and Applications*. Springer (2007), 247–260.
7. Büttgen, B. et al. Pseudonoise optical modulation for real-time 3D imaging with minimum interference. *IEEE Transactions on Circuits and Systems I: Regular Papers 54*, 10 (2007), 2109–2119.
8. Büttgen, B. and Seitz, P. Robust optical time-of-flight range imaging based on smart pixel structures. *IEEE Trans. on Circuits and Systems 55*, 6 (2008), 1512–1525.
9. Cui, Y. et al. 3D shape scanning with a time-of-flight camera. In *Proceedings of the 2010 IEEE Conf. on Computer Vision and Pattern Recognition* (2010), 1173–1180.
10. Dutton, N.A.W. et al. A SPAD-based QVGA image sensor for single-photon counting and quanta imaging. *IEEE Transactions on Electron Devices 63*, 1 (Jan. 2016), 189–196.
11. Gupta, M. et al. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics 34*, 5 (2015), 156.
12. Gupta, M. et al. What are optimal coding functions for time-of-flight imaging? *ACM Transactions on Graphics (TOG) 37*, 2 (2018), 13.
13. Henry, P. et al. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments. In *Proceedings of the 12th Intern. Symp. on Experimental Robotics (ISER)*. Citeseer, (2010).
14. Karn, P. MACA: A new channel access method for packet radio. In *Proceedings of the 9th ARRL Computer Networking Conf.* (1990).
15. Klauder, J.R. et al. The theory and design of chirp radars. *Bell System Technical J. 39*, 4 (1960), 745–808.
16. Koechner, W. Optical ranging system employing a high power injection laser diode. *IEEE Trans. AES 4*, 1 (1968).
17. Kolb, A. et al. Time-of-flight cameras in computer graphics. *Eurographics*. (2010).
18. Lange, R. 3D ToF distance measurement with custom solid-state image sensors in CMOS-CCD-technology. *Ph.D. Thesis*. (2000).
19. Lee, J. and Gupta, M. Stochastic exposure coding for handling multi-ToF-camera interference. In *Proceedings of the IEEE Intern. Conf. on Computer Vision* (2019), 7880–7888.
20. Li, L. et al. Multi-camera interference cancellation of time-of-flight (ToF) cameras. In *Proceedings of the 2015 IEEE Intern. Conf. on Image Processing*. 556–560.
21. LoRa; https://www.semtech.com/lora.
22. LoRa Modulation Basics; https://bit.ly/3Qy7MBU.
23. Ma, J. et al. Photon-number-resolving megapixel image sensor at room temperature without avalanche gain. *Optica 4*, 12 (2017), 1474–1481.
24. May, S. et al. 3D time-of-flight cameras for mobile robotics. In *Proceedings of the 2006 IEEE/RSJ Intern. Conf. on Intelligent Robots and Systems*, 790–795.
25. Pagliari, D. and Pinto, L. Calibration of Kinect for Xbox One and comparison between the two generations of microsoft sensors. *Sensors* (2015).
26. Pan, Y. et al. Accurate depth extraction method for multiple light-coding-based depth cameras. *IEEE Transactions on Multimedia 19*, 4 (2017), 685–701.
27. Payne, J.M. An optical distance measuring instrument. *Rev. of Scientific Instruments 44*, 3 (1973).
28. Rayanchu, S. et al. Airshark: Detecting non-Wi-Fi RF devices using commodity Wi-Fi hardware. In *Proceedings of the Intern. Measurement Conf.* (Nov. 2011).
29. Rayanchu, S. et al. Catching whales and minnows using Wi-FiNet: Deconstructing non-Wi-Fi interference using Wi-Fi hardware. In *ACM/USENIX Networking Systems Design and Implementation* (April 2012).
30. Shrestha, S. et al. Computational imaging with multi-camera time-of-flight systems. *ACM Transactions on Graphics (ToG) 35*, 4 (2016), 33.
31. Shrivastava, V. et al. Centaur: Realizing the full potential of centralized WLANs using a hybrid data path. In *Proceedings of the 15th Annual Intern. Conf. on Mobile Computing and Networking* (Sept. 2009), 297–308.
32. Suresh, L. et al. Towards programmable enterprise WLANs with Odin. In *Proceedings of the 1st Workshop on Hot Topics in Software Defined Networks* (2012), 115–120.
33. Ulku, A.C. et al. A 512x512 SPAD image sensor with integrated gating for widefield flim. *IEEE J. of Selected Topics in Quantum Electronics 25*, 1 (Jan. 2019), 1–12.
34. Velodyne. Pulsed LIDAR sensor; http://www.velodynelidar.com/lidar/lidar.aspx.
35. Wang, J. et al. 3D scene reconstruction by multiple structured-light based commodity depth cameras. In *Proceedings of the 2012 IEEE Intern. Conf. on Acoustics, Speech and Signal Processing*, 5429-5432.
36. Whyte, R.Z. et al. Multiple range imaging camera operation with minimal performance impact. In *Image Processing: Machine Vision Applications III 7538*, Intern. Society for Optics and Photonics, (2010).

Jongho Lee (jlee567@wisc.edu) is a research associate of Computer Sciences at the University of Wisconsin-Madison, WI, USA.

Mohit Gupta is an associate professor of Computer Sciences at the University of Wisconsin-Madison, WI, USA.

Bhuvana Krishnaswamy is an assistant professor of Electrical and Computer Engineering at the University of Wisconsin-Madison, WI, USA.

Suman Banerjee is a professor of Computer Sciences at the University of Wisconsin-Madison, WI, USA.