When Two Cameras Are a Crowd
(Understanding and Handling Interference Across Multiple Active Cameras)

Fig 1: The 3D Revolution: 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.

ABSTRACT

Vision and robotics systems enabled by cameras which 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 the common interference problems in the RF domain. Based on this observation, we describe new and emerging challenges when multiple active 3D cameras operate in the same spatio-temporal region. We also outline some solution approaches and more importantly, highlight the next steps ahead.

1 THE 3D REVOLUTION

We are in the midst of a 3D revolution fueled by cameras which can recover 3D geometry of their surroundings (Fig. 1). The key catalyst behind 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\(^1\), and thus, can potentially measure accurate 3D shape over wide area.

\(^1\)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.

Applications: Due to their low-cost, compact form-factors, and low computational complexity, ToF-based active 3D 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 (e.g., Apple iPad), gaming and AR/VR headsets (e.g., Microsoft Kinect and Hololens). They are also used by inspection and delivery robots to navigate in warehouses [3] and perform 3D modeling [4, 5], 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.\(^2\) 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 Fig. 2 (a). Scene depths (and hence, the 3D geometry) is extracted by comparing the emitted and the reflected light. Since the camera actively controls the illumination, they 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 its 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

\(^2\)In contrast, passive 3D cameras (e.g., stereoscopic 3D cameras) do not actively control the illumination, thus only passively observing the scene.

Anonymous Author(s)
1.1 Why is Now a Good Time to Address Multi-Camera Interference (MCI)?

Till recently, a large proportion of ToF cameras were based on a scanning laser beam that illuminates only a small portion of the scene point at a time. Since they illuminate only a small portion of the scene at a time, the probability that two devices illuminate the same point simultaneously is relatively low (Fig. 3 (a)). This enabled multiple devices to co-exist without affecting each other. However, they require mechanical moving parts e.g., rotating mirrors, often leading to long acquisition times, high cost, bulky, and reliability issues.

To address its limitations, a new emerging class of solid-state ToF cameras *flood-illuminate* the entire scene (Fig. 3 (b)), making them considerably cheaper, smaller, lighter, and faster. Hence, solid-state cameras are fast replacing scanning-based cameras in autonomous driving and robotics applications (Fig. 3 (c)). But, there is a trade-off: several active cameras flood-illuminating a scene simultaneously will interfere with each other (Fig. 3 (b)). Furthermore, a new generation of consumer devices (e.g., cell phones) with 3D imaging capabilities are becoming ubiquitous, due to stringent size and cost requirements, these devices also prefer solid-state 3D cameras (Fig. 3 (d)), which will create strong multi-camera
interference (MCI) whenever multiple active cameras are used in close 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 [8–10]. Therefore, these cameras are rapidly becoming the method-of-choice in almost all applications which rely on 3D cameras, including vehicular navigation, robotics, and consumer mobile devices. On the downside, this growth will add to the severity of multi-camera interference problem. 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 AR, localization, or 3D modeling [4, 5]. 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, as solid-state ToF cameras become ubiquitous in our personal mobile devices, vehicles, homes, and workplaces and become the basis for numerous critical applications, managing interference across a multitude of these devices will be of profound importance. The goal of this paper is to shine light on this important problem (pun intended) that has received little attention so far: interference of light signals emitted by active cameras and its consequent impact in accurately recovering the 3D scene information.

1.2 Can Research on Wireless Interference be Used to Address MCI?

Typical wireless communication systems (e.g., WiFi, cellular systems) utilize radio waves propagating over a shared channel. These systems have dealt with the problem of multi-node interference for many decades. This has led to the development of many techniques using which multiple wireless devices are able to 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 multi-camera interference. To answer the above question, we first discuss the similarities and differences of how interference manifests in these settings.

At a first glance, an active 3D camera is very similar to a wireless transmitter as shown in Fig. 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 utilizes 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 time-of-flight 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 i.e., the light source and the sensor are both part of the same active 3D camera (Fig. 4(b)). 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. An active 3D camera on the other hand, 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) components; 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 (discussed in Section 2.2), 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 active 3D camera design, can together synthesize useful solutions. If successful, the resulting ideas will not just spur wide-spread 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, 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 paper, 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 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 Fig. 2 (a). The most widely used class of active 3D cameras are based on the time-of-flight (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, by computing the time-shift between the emitted and received waveforms (Fig. 5 (a–b)).

ToF based depth imaging systems can be broadly classified into Direct and Indirect ToF systems. A direct ToF (D-ToF) system [11, 12] estimates scene depths by emitting a short light pulse into the scene, and directly measuring the travel time of the reflected pulse (Fig. 5 (a)). Most vehicular LiDARS are based on the D-ToF principle. An indirect ToF (I-ToF) system [13–15], 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 light source modulation and sensor demodulation functions can be any periodic and continuous functions such as sinusoids, square waves, or trapezoidal functions (Fig. 5 (b)) [16]. I-ToF cameras do not require expensive components, and hence, are routinely used in consumer devices such as Microsoft Kinect and Hololens.

### 2.1 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.

#### 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 [13, 14], or an impulse train function [17]. The period of $M(t)$ typically varies from 10 to 1000 ns, which corresponds to a measurable distance of 1 – 100 m. 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 $M(t)$:

$$R(p; t) = a M(t - \tau)$$

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 $p$, $c$ is the light speed. $a$ is a scene-dependent scale factor that encapsulates the scene’s reflectance properties. The camera computes $\tau$ (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 (Fig. 2 (b)), the brightness of the light incident at one of the cameras (referred to as the primary camera) is given by:

$$R_{\text{mul}}(t) = R(t) + MCI$$

where, $MCI = \sum_{n=1}^{N} R_n(t)$

where $N$ is the number of interfering cameras, $R(t)$, the radiance incident at the primary camera due to its own source (Eq. 1), and $R_n(t) = a_n M_n(t - \tau_n)$, the measured intensity due to the $n^{th}$ source. We drop $p$ for brevity. The summation term in Eq. 2 corrupts the true radiance $R(t)$, resulting in erroneous depth estimates. Fig. 2 (c) shows an example of ToF camera using sinusoid modulation. Assuming all the sources use sinusoids (or any other periodic shapes such as sinusoids, square waves, or trapezoidal functions).
squares) of the same frequency, the phase of $R_{\text{mult}}(t)$ may differ from the true phase of $R(t)$, resulting in systematic, potentially large depth errors as shown in Fig. 2 (d-f).

2.2 Current Approaches Addressing MCI
A trivial approach to prevent Multi-Camera Interference 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 $\approx 850$–$950$ nm, 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 (e.g., 5–10 nm). These constraints restrict the number of distinct wavelength bands preventing assigning a unique set of wavelengths to each active camera. Recent works addressing Multi-Camera Interference can be broadly classified into three categories, viz:

2.2.1 Orthogonal Coding: Majority of existing works rely on orthogonal coding, such as sinusoids of different modulation frequencies [18] or phases [19, 20], and pseudo-noise sequences [21, 22] for different cameras. However, they face challenges similar to Frequency Division Multiple Access (FDMA): limited set of orthogonal frequencies and codes 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.

2.2.2 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 [7]. These techniques do not scale with the number of interfering cameras.

2.2.3 Mechanical Approaches: Another method is to project a planar light sheet which is scanned over the scene. Since only a portion of the scene is illuminated at a time, the interference can be reduced [23–25]. Although these approaches successfully prevent interference, they require mechanical scanning, which increase the cost and size of the system.

Recently, optical phased arrays and MEMS (micro-electro-mechanical systems)-based technologies are attracting attention to resolve the limitations of mechanical approaches. These hardware-based approaches can be integrated into our software/firmware-based approaches (Sections 3 and 4) to create hybrid approaches to mitigate MCI. However, we focus on wireless-inspired software/firmware approaches to reduce MCI in this paper.

3 THE PROMISE AND PITFALLS OF WIRELESS-INSPIRED APPROACHES
Interference is a classical problem in wireless communication. A huge body of literature over multiple decades have examined this problem to enable efficient shared access of a common channel such as the wireless medium. These approaches are commonly referred to as the Medium Access Control (MAC) techniques in the wireless community. Since Multi-Camera Interference 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 Multi-Camera Interference. As discussed in §1.2, orthogonal coding across cameras can only reduce AC interference and not DC interference. A simple solution is to avoid collisions; we adopt a coordinated and distributed collision avoidance from wireless, and simulate them in a multi-camera setting and analyze their performance.

Comparisons using computer simulations of Multi-Camera Interference: 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 Fig. 6. Given a source modulation function $M(t) \geq 0$, 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:

$$L(t) = aE(t - \frac{2d}{c}) + P_a,$$

where $a$ 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 correlation between $L(t)$ and sensor demodulation $D(t)$:

$$C = \int_{\tau} L(t)D(t) \, dt,$$

where $\tau$ 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 $D_k(t) = 1 + \cos \left(2\pi f t - \frac{\pi}{4} (k - 1)\right), (k = 1, ..., 4)$. The scene distance estimate $\hat{d}$ is obtained by

$$\hat{d} = \frac{c}{4\pi f} \tan^{-1} \left( \frac{C_3 - C_4}{C_1 - C_3} \right).$$
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. The code to simulate MCI reduction approaches is available at https://github.com/purmod/MCI.

**CSMA-based Multi-Camera Interference 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. This approach of waiting till 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). In order 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 WiFi. We draw inspiration from CSMA/CA, which requires nodes to wait for a random amount of time after the channel is idle, before transmitting. Combining carrier sensing and random wait times enables avoiding 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 Multi-Camera Interference mitigation (CRA)**: We compare the depth accuracy of CSMA-based Multi-Camera Interference 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.

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. The code to simulate MCI reduction approaches is available at https://github.com/purmod/MCI.

**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 trade-off: the probability of collision increases as the network scales; to avoid collisions, more number of slots should be used, leading to increased capture time and power consumption. The time to switch from carrier sensing to normal camera mode will further increase the 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, in order to accommodate these 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.
4 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.

4.1 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 (e.g., Microsoft Kinect uses 120 MHz, 80 MHz and 16 MHz [26]). On the other hand, one of the biggest challenges is the need and ability to cancel background noise. As explained in §1.2 and §2.2, 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.

4.1.1 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 (Fig. 8 (a)). Adaptive FHSS is used in Bluetooth, due to its robustness to broadband interference, which in turn enables Bluetooth devices to co-exist with WiFi 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 multi-camera interference. 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. Additionally, the hopping pattern can 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 3 slots. The probability of the two cameras using the same frequencies in the same slots in round 2 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.

4.1.2 Randomized Chirp Spread Spectrum (CSS). Chirp signals have been used for precision ranging in RADAR [27] and more recently, CSS modulation has been used in LoRa [28, 29] 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 the interference from other sources do not affect the AC component of the reflected signal. Fig. 8 (b) 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 the step size will increase the number of chirps, the minimum step-size is limited by the 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 Fig. 8 (b), 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 [27, 29]. Similar to FHSS, CSS also needs to 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.
### 4.2 Centralized, Networked Coordination

While distributed approaches to MCI is necessary for many applications, there are multiple environments that lend themselves well to networked coordination with centralized control. For e.g., factory floor with mobile autonomous robots that operate in a single administrative domain can easily implement such mechanisms. In such indoor applications, RF-based wireless connectivity to network devices to a central server, that can in turn synchronize devices could be integrated with the cameras. Therefore, combining RF and Camera-based networks could offer a real-time control of the cameras to 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.

#### 4.2.1 Creating a conflict graph in real-time

In wireless systems, to schedule simultaneous transmissions, one needs to infer the “conflict graph” that maps the potential of interference between transmitters [30–32]. 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 towards a practical deployment of a centralized coordination. A bigger challenge is to meet the same goals as that in a passive setting, i.e., by simply observing activities on the channel, and using time information to determine the same information. This, however, maybe effective if all 3D cameras are sufficiently active. Overall, a hybrid passive-active method would optimize the best of both alternatives.

#### 4.2.2 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 [33, 34]. Traffic scheduling is particularly effective when a vast majority of the traffic can be unscheduled, as they can be addressed through simple distributed mechanisms. For the fraction that cannot be addressed effectively well by those mechanisms can benefit from centralized control. Such traffic scheduling can be used to minimize collisions in the active camera environments too. The real-time conflict graphs discussed above will inform a centralized coordinator which transmitters might benefit from scheduling, while allowing the remaining to be unscheduled. Note that the scheduling requirements might depend on specific environments, and the frequency with which transmitters need to re-discover 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, maybe scheduled to timeslots to accomplish their scene sensing goals. Note that a sensing attempt maybe scheduled (or not), depending on the other devices attempting to sense concurrently.

### 5 CONCLUSIONS

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 [35], MACAW [36]) started by exploring the potential of wired interference mitigation strategies in the wireless environment (Ethernet and related CSMA/CD approaches) and their consequent limitations and 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.

**Acknowledgements.** This research was supported in part by NSF CAREER Awards 1943149 and 2142978, NSF awards CNS-2107060, ECCS 2034415, Intel-NSF award CNS-2003129.
REFERENCES


