

# SONDAR: Size and Shape Measurements Using Acoustic Imaging

Xiaoxuan Liang<sup>1,\*</sup>, Zhaolong Wei<sup>1,\*</sup>, Dong Li<sup>1,2</sup>, Jie Xiong<sup>1</sup>, Jeremy Gummeson<sup>1,†</sup>

<sup>1</sup>University of Massachusetts Amherst

<sup>2</sup>University of Maryland, Baltimore County

{xiaoxuanliang, zhaolongwei}@umass.edu, dli@umbc.edu, jxiong@cs.umass.edu, jgummeso@umass.edu

## Abstract

Acoustic signal has been used to sense important contextual information of targets such as location and movement. This paper explores its potential for measuring the geometric information of moving targets, i.e., shape and size. We propose SONDAR, a novel shape and size measurement system using Inverse Synthetic Aperture Radar (ISAR) imaging on commodity devices. We first design a Doppler-free echo alignment method to accurately align target reflections even if there exists a severe Doppler effect. Then we down-convert the received signals reflected from multiple scatter points on the target and construct a modulated downchirp signal to generate the image. We further develop a lightweight approach to extract the geometric information from a 2-D frequency image. We implement and evaluate a proof-of-concept system on both a Bela platform and a smartphone. Extensive experiments show that we can correctly estimate the target shape and achieve a millimeter-level size measurement accuracy. Our system can achieve high accuracies even when the target moves along a deviated trajectory, at a relatively high speed, and under obstruction.

## CCS Concepts

• **Human-centered computing** → Mobile devices; • **Computer systems organization** → Sensor networks.

## Keywords

acoustic sensing, acoustic imaging, size and shape measurement

## ACM Reference Format:

Xiaoxuan Liang<sup>1,\*</sup>, Zhaolong Wei<sup>1,\*</sup>, Dong Li<sup>1,2</sup>, Jie Xiong<sup>1</sup>, Jeremy Gummeson<sup>1,†</sup>. 2024. SONDAR: Size and Shape Measurements Using Acoustic Imaging. In *International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '24)*, October 14–17, 2024, Athens, Greece. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3641512.3686359>

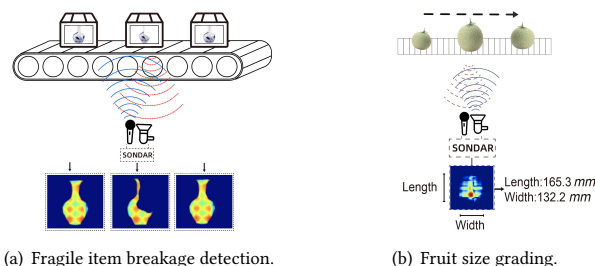
## 1 Introduction

Understanding the geometry of objects is crucial across various sectors, including supply chains, retail, industrial manufacturing, and home automation. For example, in supply chains, knowing the shape and size of items helps in detecting potential breakages during transit, ensuring that goods reach customers intact. In retail,

\*Both authors contributed equally to this work. †Corresponding authors.



This work is licensed under a Creative Commons Attribution International 4.0 License. *MobiHoc '24, October 14–17, 2024, Athens, Greece*  
© 2024 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-0521-2/24/10  
<https://doi.org/10.1145/3641512.3686359>



**Figure 1: Two applications for SONDAR. (a) Identifying whether the items inside the package are intact without opening the package. (b) Size grading of fruits on a conveyor belt.**

geometric data enhances computer vision systems, facilitating a smoother and more transparent shopping experience for customers. This knowledge allows businesses to streamline operations, reduce errors, and improve customer satisfaction by accurately handling products according to their dimensions.

Prior systems use LiDAR [17] and radar [9] to measure the target's size and shape, but their high cost and complexity limit integration into low-cost devices. Other methods like cameras paired with machine learning [10] raise privacy concerns and are less effective in poor lighting conditions. Acoustic imaging provides a promising alternative by using sound reflections to determine the target geometry. This technique does not require a light source, making it superior for low-light conditions. Unlike camera-based and LiDAR systems that need a clear line of sight, acoustic imaging can measure objects even when obstructed [23]. Moreover, acoustic sensing achieves millimeter-level accuracy using low-cost devices, making it suitable for widespread applications [4, 5, 15, 19, 24].

Acoustic sensing has primarily focused on extracting location and motion information, such as distance [20], angle [16], and speed of a target [18]. In this paper, we enhance the functionality of commodity acoustic devices to also determine the shape and size of objects, broadening their utility across various real-world applications. For instance, as illustrated in Fig. 1, shape estimation can identify breakages in fragile goods within supply chains without requiring a visual inspection. Furthermore, precise size measurements facilitate the sorting and grading of products like fruits on a conveyor belt [2], enhancing efficiency in agricultural processing.

The concept behind acoustic-based shape and size measurement involves generating a two-dimensional (2-D) image of a target from its acoustic reflections. Typically, this is achieved using a large array of speakers and microphones, similar to how ultrasound imaging [36] visualizes tissue structures in medical applications. However, standard acoustic devices often feature only a few speakers and microphones, limiting their capability for detailed imaging. One prior study, AIM, proposes to use a smartphone to perform acoustic imaging by capturing reflections as the user manually

moves the device along a set path [23], but this method is impractical for many applications where human involvement is not feasible.

We have developed a system named SONДАР that utilizes the Inverse Synthetic Aperture Radar (ISAR) technique for acoustic imaging without human intervention. In this system, when a target moves (e.g., on a conveyor belt), a speaker emits chirp signals that are reflected off the target from various positions. A microphone captures these signals, allowing us to construct a 2-D image of the target and extract detailed geometric information. Although promising, there are several challenges to overcome in order to implement an accurate size and shape measurement system:

- Acoustic sensing, compared to radar, is more vulnerable to detrimental impacts of the Doppler effect when the target is moving [16]. The Doppler effect introduces additional time delays in received signals, disrupting echo alignment and compromising image clarity. To overcome it, we have developed a solution that effectively neutralizes Doppler-induced time delays to achieve Doppler-free echo alignment. Our method starts by iteratively estimating Doppler velocities for each received echo, capitalizing on the moment when there is no Doppler shift (i.e., zero radial velocity) as the target passes directly by the device. With these velocities, we create a modified chirp template that mirrors the compressing or stretching in reflected chirp signals. This allows for a precise cross-correlation between the tailored chirp template and the received echoes, accurately eliminating the additional time delays caused by the Doppler effect.
- ISAR-based imaging differentiates itself from other sensing systems by treating a moving target not as a single scatter point, but as composed of multiple scatter points. In conventional radar imaging, the typical approach that processes signals reflected by these scatter points involves constructing a downchirp signal with a sweeping frequency range large enough to capture all reflections during the dechirp process [6]. However, this approach is impractical for acoustic imaging due to the limited bandwidth and sampling rates of acoustic devices. To address this limitation, we have developed a straightforward and effective method for dechirping received echo signals. This involves down-converting the received echoes from the ultrasonic band to the baseband and then using a modulated downchirp signal for the dechirping process. This adaptation allows for effective processing of acoustic signals within the constraints of current device capabilities.
- To accurately extract geometric information from targets, it is crucial to convert the generated image from frequency space to physical space. This conversion involves scaling the pixel coordinates of the frequency image by their respective frequency resolutions. If the frequency resolutions are not appropriately chosen, the resulting image in physical space may appear distorted. The selection of these resolutions is particularly complex as they depend on target motion parameters like angular velocity. Traditional radar imaging systems typically assume a constant angular velocity during the imaging process. This is a reasonable assumption given the short 50 ms duration of radar imaging [6]. However, this assumption does not hold in acoustic imaging, which requires a significantly longer duration of at least 2.5 s for effective imaging. To address this issue, we propose a lightweight method that maps the 2-D image from frequency space to physical space by mathematically defining the relationship between

**Table 1: Comparison with the SOTA studies.**

Technology		Performance					
		Implementation simplicity	Low cost	Low power	Privacy preservation	Low-visibility environment	NLOS Size & shape measurement
Light	Camera [8, 10, 21]	✓	×	×	×	×	×
	LiDAR [17, 25]	×	×	×	✓	×	✓
RF	mmWave [11, 13]	×	✓	✓	✓	✓	×
	WiFi [34]	×	✓	×	✓	✓	×
	RFID [33]	×	✓	×	✓	✓	×
	Ultrasound [1]	×	×	×	✓	✓	✓
Acoustic	Handheld SAR [23]	✓	✓	✓	✓	✓	×
	<b>Our method</b>	✓	✓	✓	✓	✓	✓

the frequency resolution and motion parameters of the target. This approach ensures that the physical space representation of the target is accurate and undistorted.

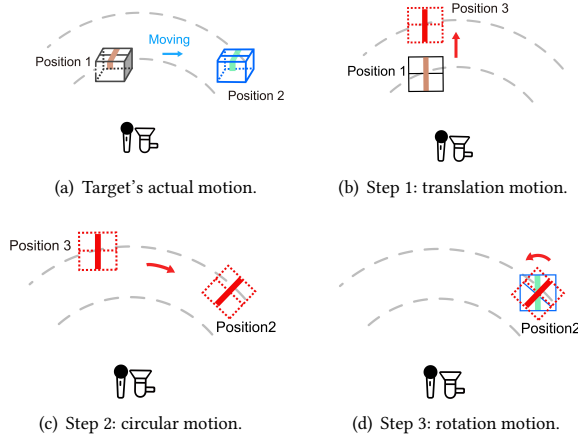
We have developed a proof-of-concept system using the Bela [29] platform and a Samsung Galaxy S9+ smartphone and systematically evaluated its performance under different conditions. Our contributions are summarized as follows:

- To our knowledge, SONДАР is the first system capable of measuring the shape and size of moving targets using just a single speaker and microphone pair on commodity devices. This enhanced functionality has potential applications in logistics, industrial manufacturing, security checks, and more.
- We develop a Doppler-free echo alignment algorithm for acoustic imaging, which can also be adapted to reduce the Doppler effect in other acoustic sensing applications. Additionally, we introduce a mapping technique that accurately transforms target images from frequency space to physical space.
- We implement and evaluate our proposed system using two commodity acoustic devices. Our extensive testing shows that the system accurately estimates target shapes and achieves millimeter-level measurement precision. It maintains high accuracy even when targets deviate from a set trajectory, move at high speeds, or are obstructed. We demonstrate the system's practicality through 3 real-life applications: grading fruits by size, detecting breakages in fragile items, and measuring box sizes.

## 2 Related Work

This section provides an overview of state-of-the-art (SOTA) studies, explaining our choice of acoustic signals for developing the SONДАР system. We categorize existing technologies based on the type of signals they use: light, RF, and acoustic. A comparative analysis of these methods with our approach is presented in Table 1.

While camera-based methods can measure target shape and size [8, 21], they struggle in low light conditions and raise privacy concerns. Additionally, they are ineffective for non-line-of-sight (NLOS) targets and require costly specialized equipment for optimal functionality. LiDAR [17, 25] can accurately capture shape and size through point clouds but shares similar limitations as camera-based methods. RF-based solutions excel in detecting NLOS targets but necessitate expensive hardware and extensive frequency bandwidths [11, 13, 33]. Ultrasound-based methods have been proposed for measuring target size and shape, but they require bulky and expensive equipment [1]. In contrast, a recent study utilizes readily available, low-cost acoustic devices [23]. However, its approach requires the device to move at a constant and low speed to minimize Doppler effects, which is impractical in many scenarios. Our



**Figure 2: (a) The target's actual motion can be decomposed into three steps, i.e., (b) translation motion, (c) circular motion, and (d) rotation motion.**

method effectively eliminates Doppler shifts and maintains high accuracy even when the target moves at a relatively high speed.

Furthermore, most existing solutions [8, 10, 13, 14, 21, 27] rely on deep learning techniques to achieve accurate results, necessitating significant computational resources and extensive data inputs. In contrast, our proposed system accurately measures the target's shape and size without relying on deep learning.

### 3 PRELIMINARIES

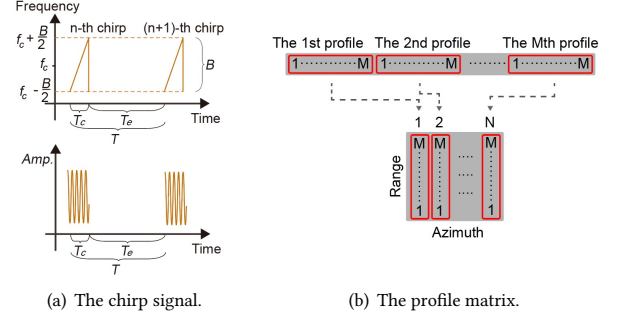
#### 3.1 The principle of ISAR

Inverse Synthetic Aperture Radar (ISAR) is widely used in military applications to determine the shape of moving targets like ships. ISAR differs from other technologies like motion tracking by treating the target as a collection of multiple scatter points instead of a single point. The principle behind ISAR imaging involves transforming the target's translational movement into rotational motion to capture reflected signals from various angles around the target.

As illustrated in Fig. 2(a), consider a target moving from Position 1 to Position 2. This movement can be broken down into three steps, depicted in Fig. 2(b)-2(d). Initially, the target moves radially from Position 1 to Position 3 (Fig. 2(b)), causing additional propagation delays for the signals reflected from each scatter point. These delays can complicate image accuracy and are corrected through our echo alignment technique. Next, the target moves along a circular path centered at the imaging device from Position 3 to Position 2 (Fig. 2(c)), which introduces no change in propagation delays and thus does not affect the imaging outcome. Finally, the target rotates around its geometric center by a specific angle (Fig. 2(d)), aligning with the azimuth angle traversed in its actual movement.

#### 3.2 The Basics of Chirp Signal

A speaker emits a series of chirp signals whose frequencies sweep linearly over time, as depicted in Fig. 3(a). Each chirp's frequency spans from  $f_c - \frac{B}{2}$  to  $f_c + \frac{B}{2}$ , where  $f_c$  is the carrier frequency and  $B$  is the bandwidth. The duration of each chirp,  $T_c$ , is followed by a pause,  $T_e$ , to ensure that all echoes from one chirp are received



**Figure 3: The signal illustration for imaging.**

before the next chirp begins. Each chirp can be represented as

$$s(t) = \text{rect}\left[\frac{t}{T_c}\right] \cos\left(2\pi\left(f_c t + \frac{k}{2}t^2\right)\right), \quad (1)$$

where  $k = \frac{B}{T_c}$  is the slope of the chirp. The term  $\text{rect}[\cdot]$  refers to the rectangle window function, which is defined as

$$\text{rect}[\cdot] = \begin{cases} 1 & |\cdot| \leq 0.5 \\ 0 & |\cdot| \geq 0.5. \end{cases} \quad (2)$$

Note that the total duration of the chirp and the gap interval is  $T = T_c + T_e$ . The transmitted signal is reflected by the target and then received by a microphone, which is often termed as an echo signal. The time delay ( $\Delta\tau$ ) of the echo signal is usually computed by performing correlation between the transmitted chirp signal and the received echo signal [22]. After getting the time delay, we can derive the distance  $D$  between the target and device using

$$D = \frac{\Delta\tau \times v_s}{2}, \quad (3)$$

where  $v_s$  is the sound propagation speed in the air.

#### 3.3 Signal Model for Imaging

Received echo signals are organized into a profile matrix for imaging, as shown in Fig. 3(b). This matrix arranges  $M$  samples per column with  $N$  consecutive collected profiles to construct the image. Then we apply the Hilbert transform to the echo signals, followed by dechirping each column to extract the intermediate frequency (IF) signal matrix. Subsequently, a 2-D Fast Fourier Transform (FFT) is performed on the IF signal matrix along both the range and azimuth dimensions. This step effectively maps each scatter point in physical space to a corresponding pixel in the image space,  $I$ , where a scatter point  $p$  is identified by a peak at position  $(p_r, p_a)$  in the 2-D image space, which can be represented as

$$I(i_a, i_r) = A \text{sinc}\left[\frac{2kT}{v_s}(i_r - p_r)\right] \times \text{sinc}\left[\frac{2\theta}{\lambda}(i_a - p_a)\right], \quad (4)$$

where  $A$  is a constant related to the amplitude and chirp duration,  $v_s$  is the sound speed,  $\lambda$  is the wavelength,  $\theta$  is the angle that the target rotates during the imaging process,  $i_r$  and  $i_a$  are the indexes along the range dimension and azimuth dimension, respectively.

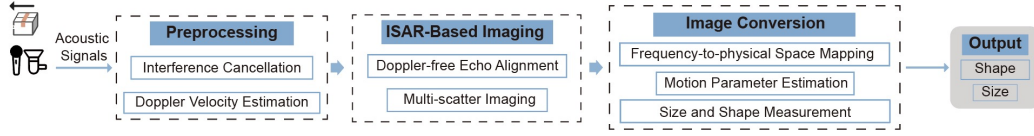


Figure 4: The signal processing pipeline of SONДАР.

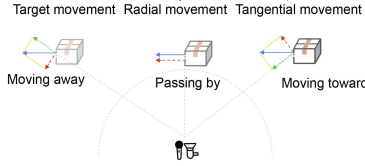


Figure 5: Target movement can be decomposed into radial movement and tangential movement, where radial movement causes the harmful Doppler effect.

## 4 SONДАР Design

Fig. 4 illustrates the core concept of SONДАР. It first generates a 2-D image of a moving target through ISAR-based imaging and then converts this image from frequency space to physical space.

### 4.1 Preprocessing

**4.1.1 Interference Cancellation.** In acoustic imaging, three primary types of interference complicate signal processing: ambient sound noises, direct path interference from the speaker to the microphone, and multipath reflections from static objects like walls. To mitigate ambient sound noises, which typically occur below 16 kHz, we use a finite impulse response (FIR) bandpass filter with a pass-band range of 18 – 22 kHz. This filter effectively isolates the received chirp signals by excluding lower frequency sounds. For the direct path and multipath reflections, which are relatively consistent over time, we apply a background subtraction technique [24]. This method effectively removes these static components from the signal, allowing for clearer analysis of the moving target’s echoes.

**4.1.2 Doppler Velocity Estimation.** Fig. 5 depicts the decomposition of target movement into radial and tangential components as the target approaches, passes, and moves away from the device. Unlike tangential movement, radial movement induces a Doppler velocity, which is also known as a Doppler frequency shift. Acoustic sensing is particularly vulnerable to this Doppler effect due to the relatively low speed of acoustic signals, making it highly sensitive to changes in target velocity [16]. For instance, if a target moves at a velocity of 30 cm/s at an angle of 34° relative to the device, the resulting radial velocity is about 25 cm/s. This movement can lead to an additional 3 cm error in distance estimation, potentially causing misalignment of echoes and blurring the resultant image. Accurately estimating and compensating for the Doppler velocity is therefore essential to obtain a clear and accurate image.

The key idea for estimating the Doppler velocity is to use the moment when the target is closest to the device, which represents the point of minimal radial movement and distance estimation error. At this point, the estimated distance serves as a reference for calculating the Doppler velocities (or radial velocities) for other profiles. We calculate a series of distances between the target and the device throughout the imaging process using the estimated time delays of echo signals as indicated in Equation (3). These



(a) Before echo alignment.

(b) After echo alignment.

Figure 6: Echo alignment removes extra delays caused by target movement and aligns echo signals across profiles.

distances are denoted as  $[D_1, \dots, D_{\min}, \dots, D_N]$ , where  $D_1$  and  $D_N$  are the distances at the first and last profiles, respectively.  $D_{\min}$  is the minimum distance.

After identifying the minimal distance  $D_{\min}$ , we calculate the radial velocities at adjacent profiles, specifically  $V_{r_{\min-1}}$  and  $V_{r_{\min+1}}$ . Take  $V_{r_{\min+1}}$  as an example. The difference between the distance at the  $(\min+1)$ -th profile  $D_{\min+1}$  and the distance at the  $(\min)$ -th profile  $D_{\min}$  involves two components, including the displacement  $V_{r_{\min+1}}(t_{\min+1} - t_{\min})$  and the velocity-induced deviation  $\frac{(f_c - \frac{B}{2})V_{r_{\min+1}}(t_{\min+1} - t_{\min})}{B}$ , where  $f_c$  and  $B$  are the carrier frequency and chirp bandwidth, respectively [16]. Since all the variables except the radial velocity  $V_{r_{\min+1}}$  are known, we can derive it as

$$V_{r_{\min+1}} = \frac{D_{\min+1} - D_{\min}}{(1 - \frac{(f_c - \frac{B}{2})}{B})(t_{\min+1} - t_{\min})}, \quad (5)$$

This procedure is iteratively applied to each profile until radial velocities are determined for all profiles. This approach allows for a precise computation of the target’s radial velocity and helps correct for any potential deviations in the imaging process caused by the Doppler effect.

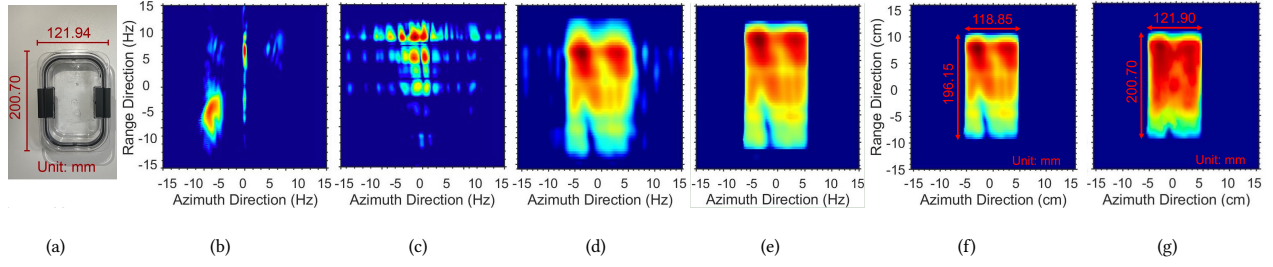
### 4.2 Doppler-free Echo alignment

The image of a moving target is constructed using the time delays from echo signals reflected by multiple scatter points. However, as shown in Fig. 6(a), target movement introduces additional time delays in echo signals that vary across different profiles, complicating the imaging process. These additional delays obscure meaningful image information, as visualized in Fig. 7(b), which shows the pre-alignment image of a glass food storage box in Fig. 7(a).

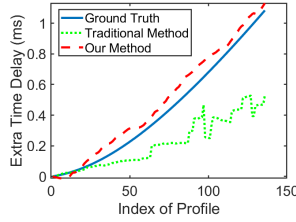
To address this issue, it is essential to align the echo signals across various profiles to remove the extra delays introduced by the target’s movement. Effective alignment, as illustrated in Fig 6(b), ensures that the resulting image accurately represents the target’s shape without distortion caused by its motion.

**4.2.1 Limitation of Traditional Method.** The traditional method for calculating extra delays in echo signals relies on maximizing the similarity between adjacent profiles using cross-correlation [12]. The green dotted line in Fig. 8 shows the delays computed by this method when a target moves away from the device at a velocity





**Figure 7: An illustrative example of imaging a glass food storage box moving at a speed of 10 cm/s and positioned 30 cm away from the device. (a) Photo. (b) Before echo alignment. (c) Traditional echo alignment. (d) Our echo alignment method. (e) After Doppler shift removal. (f) After image conversion. (g) The ground-truth image.**



**Figure 8: The comparison of echo alignment methods.**

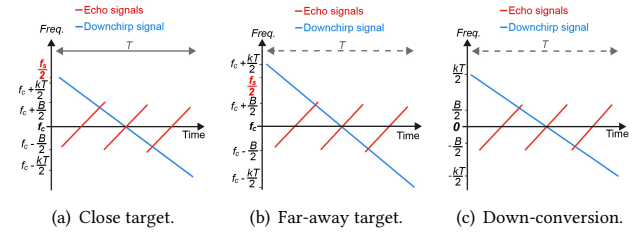
of 10 cm/s. The presence of Doppler shifts significantly impairs its effectiveness, causing deviations of the computed delays from their actual values. The deviated delays cause misalignment for profiles, resulting in a degraded image as shown in Fig. 7(c). The degradation is so severe that it becomes hard to discern any objects, highlighting the need for a more robust approach for echo alignment.

**4.2.2 Our Method.** We have developed a new chirp template designed to mitigate the effects of Doppler shifts on echo signals. The principle behind this approach is that movements of the target toward or away from the device cause the reflected echo signals to compress or stretch in both the time and frequency domains. Utilizing the Doppler velocities  $V_r$  estimated in Sec. 4.1.2, we crafted a chirp template that mirrors this compression or stretching.

The chirp template  $ST(t)$  compensates for these distortions by adjusting its frequency and time characteristics to align with the modified echo signals. It is mathematically defined as  $ST(t) = \cos\left(2\pi\left(f'_c t + \frac{B'}{2T'_c} t^2\right)\right)$ , where  $t \in [0, T'_c]$ , and

$$\begin{cases} f'_c = f_c \times \left(1 - \frac{V_r}{v_s}\right) \\ B' = B \times \left(1 - \frac{V_r}{v_s}\right) \\ T'_c = \frac{T_c}{1 - \frac{V_r}{v_s}} \end{cases} \quad (6)$$

$f_c$ ,  $B$ , and  $T_c$  represent the carrier frequency, bandwidth, and duration of the transmitted chirp, respectively, and  $v_s$  is the speed of sound. Note that, since the frequency of the chirp signal sweeps linearly over time, we need to compensate not just the central frequency of the chirp signal but the frequency at each timestamp, which is achieved by correcting  $B$  and  $T$ . By conducting a cross-correlation between the specifically designed chirp template and echo signals from each profile, we can accurately determine the extra delays, effectively removing the impact of Doppler shifts, as illustrated by the red dashed line in Fig. 8. We pinpoint these extra delays by identifying the strongest peaks in the correlation results.



**Figure 9: Compared to (a) a close target, (b) dechirping echo signals from a far-away target needs a longer duration, resulting in inadequate sampling. (c) Down-conversion is performed on received echo signals to avoid this issue.**

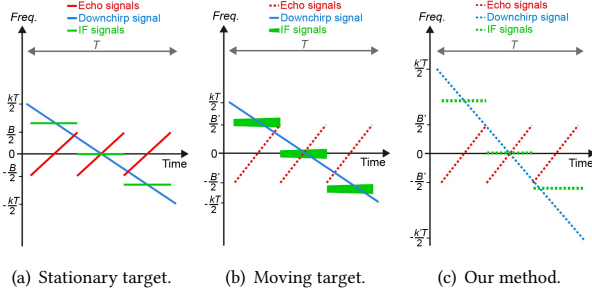
Then we implement echo alignment by synchronizing extracted delays across different profiles. Fig. 7(d) shows the results of applying our echo alignment technique. We can observe that the rough contour of the target is clearly visible, demonstrating the effectiveness of this method in enhancing image clarity.

### 4.3 Multi-scatter Imaging

Unlike previous acoustic sensing systems such as motion tracking, which view a moving target as a single scatter point, our proposed imaging system treats the target as a composition of multiple scatter points. This section outlines the methods for extracting image information from echo signals reflected by these scatter points.

**4.3.1 Addressing the Inadequate Sampling Issue.** The conventional method for generating a target image is constructing a downchirp signal and using it to perform the dechirp process. This process involves multiplying the downchirp signal with the received echo signals, which have been converted into complex signals via the Hilbert transform [28]. As illustrated in Fig. 9(a), the red upchirp signals represent the echoes from multiple scatter points received in a single profile. And the blue downchirp signal sweeps from  $f_c + \frac{kT}{2}$  to  $f_c - \frac{kT}{2}$  over a duration  $T$ . Here,  $f_c$  is the carrier frequency,  $k$  the chirp slope, and  $T$  the profile duration. Given that  $k = \frac{B}{T_c}$  and  $T = T_c + T_e$ , where  $B$ ,  $T_c$ , and  $T_e$  represent the chirp bandwidth, chirp duration, and empty duration, respectively, the maximum frequency of the downchirp signal is expressed as  $f_c + \frac{B(T_c + T_e)}{2T_c}$ .

The empty duration  $T_e$  is crucial as it ensures that all reflections from the target are captured. For targets close to the device, a smaller  $T_e$  suffices, and the conventional method functions effectively as depicted in Fig. 9(a). However, for distant targets, a larger  $T_e$  is necessary, leading to potential undersampling issues.



**Figure 10: (a) Dechirping echo signals for a stationary target outputs single-frequency IF signals. (b) However, the same dechirp process for a moving target results in wideband IF signals due to Doppler shifts. (c) Our method adopts a modulated downchirp signal to remove its impact.**

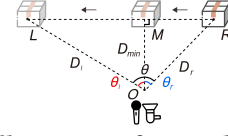
Specifically, the maximum frequency of the downchirp signal may exceed the device's maximum supportable frequency  $\frac{f_s}{2}$ , where  $f_s$  is the sampling rate, as demonstrated in Fig. 9(b).

To address this issue, we propose a straightforward yet effective approach that down-converts the received echo signals to the baseband. This is achieved by multiplying the received echo signals with  $\cos(2\pi f_c t)$  and subsequently applying a low-pass filter to eliminate high-frequency components. This results in a baseband signal that is centered at 0 Hz and sweeps from  $-\frac{B}{2}$  to  $\frac{B}{2}$ , as shown in Fig. 9(c). The corresponding downchirp signal is then configured to sweep inversely from  $\frac{kT}{2}$  to  $-\frac{kT}{2}$  over the same duration  $T$ .

**4.3.2 Removing the Impact of Doppler Shifts for Scatter Points.** Fig. 10(a) shows the dechirp process for echoes from multiple scatter points on a stationary target. It consolidates signal energies into distinct single-frequency intermediate-frequency (IF) signals, which carry spatial information about the scatter points. However, as depicted in Fig. 10(b), when the target moves away from the device, Doppler shifts extend the echo signals in time and frequency. Dechirping these stretched signals with the standard downchirp results in wideband IF signals, which have dispersed energy and increased noise susceptibility. This effect is evident in Fig. 7(d), where the target's blurred contour makes precise measurement difficult.

To solve this issue, we develop a modulated downchirp signal, tailored for each scatter point based on their estimated Doppler velocities, which we calculate in the preprocessing stage. The modulated downchirp, defined as  $MD(t) = \cos\left(2\pi\left(\frac{k'T}{2}t - \frac{k'}{2}t^2\right)\right)$ , where  $t \in [0, T]$ ,  $k' = k \times (1 - \frac{v_r}{v_s})^2$ ,  $k$  is the chirp slope,  $T$  the profile duration, and  $v_s$  the sound speed. This modulation aligns the downchirp signal with the stretched echoes, producing clean single-frequency IF signals as shown in Fig. 10(c). The image processed with this technique, as depicted in Fig. 7(e), shows a more distinct and clearer contour compared to the one without removing Doppler shifts.

**4.3.3 Phase Compensation.** Target movement introduces phase offsets in received echo signals, potentially blurring the resulting image [32]. To correct these phase offsets, we use the Minimum Entropy Autofocus (MEA) algorithm [7], commonly employed in radar imaging. Once the phases are compensated, we apply a 2-D Fast Fourier Transform (FFT) to the intermediate frequency (IF) signal matrix, generating the target image in frequency space.



**Figure 11: The illustration of azimuth angle estimation.**

## 4.4 Image Conversion

The image generated so far is in frequency space and does not accurately capture the target's actual geometric details. This section explains how we transform the image from frequency space to physical space to precisely measure the target's size and shape.

**4.4.1 Frequency-to-physical Space Mapping.** The mapping operation is achieved by multiplying pixel coordinates of the frequency image along range and azimuth dimensions with their frequency resolutions, respectively. We show how frequency resolutions can be derived from signal parameters and target motion parameters. Specifically, the range resolution  $\rho_r$  can be computed using the FFT frequency resolution  $\Delta f = \frac{1}{T}$ , which can be derived as

$$\rho_r = \frac{\Delta f}{B} \cdot T_c \cdot \frac{v_s}{2} = \frac{v_s \cdot T_c}{2BT}, \quad (7)$$

where  $B$  and  $T_c$  are the chirp bandwidth and duration, respectively.  $T$  is the profile duration.  $v_s$  is the sound speed in the air. The azimuth resolution  $\rho_a$  is related not only to signal parameters but also to target motion parameters [35], which can be expressed as

$$\rho_a = \frac{\lambda}{2\omega I} = \frac{\lambda}{2\Theta}, \quad (8)$$

where  $\lambda$  is the carrier wavelength.  $\omega$ ,  $I$ , and  $\Theta$  are the target angular velocity, rotation interval, and rotation angle along its geometric center, respectively. Prior radar imaging systems [35] assume that the angular velocity is constant during the imaging process. This is a valid assumption since the imaging duration for radar signals is only 50 ms. However, the constant angular velocity cannot be guaranteed for acoustic signals that require at least 2.5 s for imaging. Next, we present an approximate estimation solution that allows us to accurately estimate the motion parameter, i.e., rotation angle.

**4.4.2 Motion Parameter Estimation.** According to the motion decomposition model in Sec. 3.1, the rotation angle at the target's geometric center equals to the azimuth angle the target traverses along its actual trajectory. We use accurately estimated distances from acoustic signals to compute this azimuth angle. As shown in Fig. 11, if a target moves from the right side  $R$  to the left side  $L$  of the device, the azimuth angle  $\theta$  is split into two components,  $\theta_l$  and  $\theta_r$ , by the line from the device's position  $O$  to the midpoint  $M$  where the target is closest to the device. Then we approximate  $\triangle OMR$  and  $\triangle OML$  as right triangles and estimate the azimuth angle as

$$\theta = \theta_l + \theta_r = \arccos \frac{D_{\min}}{D_l} + \arccos \frac{D_{\min}}{D_r}, \quad (9)$$

where  $D_l$ ,  $D_{\min}$ , and  $D_r$  are the estimated distances between the target and device, respectively. The above-mentioned approximation is based on the fact that the target moves approximately in a straight line in our application scenarios, e.g., on a conveyor belt or manually pushing the box. Our experimental results in Sec. 6.2.7 show that this approximation is valid even if the target trajectory deviates from the straight line by tens of centimeters.

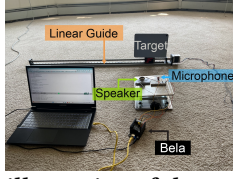


Figure 12: The illustration of the experiment setup.

**4.4.3 Size and Shape Measurement.** Fig. 7(f) shows the physical-space image of the glass food storage box generated by SONDAR, which aligns well with the ground-truth image as visualized in Fig. 7(g). To determine the size, we first calculate the differences in pixel values along the range dimension and identify boundary pixels based on an empirical threshold of 6 dB. The target's length is then determined by calculating the maximum distance between the upper and lower boundaries. The target's width is similarly computed along the azimuth dimension. For shape extraction, we utilize the OpenCV [3] library, a well-known open-source tool. Since the object size does not exceed  $700\text{ mm} \times 700\text{ mm}$ , we manually crop the generated image as this size and ensure the target is centered in the image. We input a physical-space image into OpenCV, which then classifies the shape.

## 5 Implementation

We implement our proof-of-concept system using two commodity devices: a Bela [29] platform and a Samsung Galaxy S9+ smartphone. Signal processing and analysis are carried out using MATLAB on a Hasee laptop equipped with an Intel i7 CPU and 16GB of memory. As depicted in Fig. 12, we perform controlled experiments by mounting the target on a FUYU FSL40 Linear Guide.

**Bela Platform:** The Bela [29] platform is favored in acoustic sensing research for its versatility in supporting various microphone and speaker configurations. We equipped the Bela with a PUI Audio AS03104MR-N50-R speaker for signal transmission and a SparkFun BOB-18011 MEMS microphone for signal reception.

**Smartphone:** Our prototype system was also implemented on a Samsung Galaxy S9+ smartphone using the existing LibAS framework [30]. The framework streams acoustic signals to the laptop without a need of considering smartphone-specific details.

**Acoustic Signals:** To avoid interference from environmental noise, which typically falls below 16 kHz [16], we configure our system to sweep frequencies from 18 kHz to 22 kHz. The duration of each chirp is set as 5 ms, and the empty duration is set as 20 ms. The Samsung Galaxy S9+ smartphone employs a sampling frequency of 48 kHz, while the Bela platform operates at 44.1 kHz.

**Ground Truth Measurements:** We utilize an electronic caliper to measure the ground truths of target sizes. Specifically, for each target, we measure its maximum length and width. In addition, we manually record the ground-truth shapes for targets.

## 6 Evaluation

We conducted a thorough evaluation of the SONDAR system using the Bela platform under various conditions and carried out two field studies to assess its applicability in real-life scenarios. For consistency across tests, unless specified otherwise, we used a cardboard box ( $140\text{ mm} \times 140\text{ mm}$ ) as the target. The minimum distance between the target and the device was maintained at 30 cm, with the target moving at a speed of 10 cm/s. Each experiment was repeated 20 times to ensure reliability.

Table 2: Performance comparison among different systems.

Methods	Signals	Scenes			
		LOS		NLOS	
		Shape accuracy	Size error(mm)	Shape accuracy	Size error(mm)
Zheng [37]	RGB	93.4%	18.37	-	-
Qin [26]	RGB	95.3%	2.50	-	-
Wang [33]	RF	89.7%	13.63	91.7%	13.63
Mao [23]	Acoustic	84.2%	27.76	82.1%	29.62
<b>Our system</b>	<b>Acoustic</b>	<b>92.9%</b>	<b>5.39</b>	<b>90.4%</b>	<b>6.47</b>

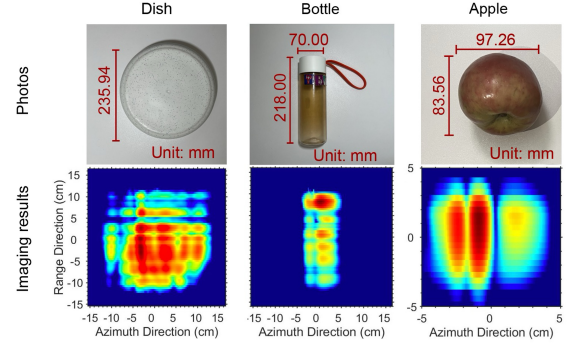


Figure 13: Imaging results for three common objects.

We report the average size measurement errors, which are defined as the difference between the estimated and actual sizes of the target. Additionally, we assess the accuracy of shape estimation, which is defined as the ratio of correctly identified shapes to the total number of shape measurements.

### 6.1 Overall Performance

**6.1.1 Imaging Results for Different Objects.** To intuitively demonstrate the effectiveness of SONDAR, we evaluate the imaging performance for objects with different shapes, including dish, water bottle, and apple, as illustrated in Fig. 13. We can observe that regular-shaped objects (i.e., dish and bottle) can achieve better imaging performance than irregular-shaped objects like apple. The reason is that irregular-shaped objects can cause diffusion for acoustic signals, leading to distortion in imaging results. Amazingly, even if the surfaces of objects are rough and uneven, we can still achieve reasonably accurate imaging results.

**6.1.2 Comparison with the State-of-the-art System.** We conducted a qualitative performance comparison between our SONDAR system and other state-of-the-art systems when measuring size and shape for 30 daily objects. These tests were carried out under both a Line-of-Sight (LOS) scenario and a Non-Line-of-Sight (NLOS) scenario with objects inside a cardboard box. The tested objects varied widely in shape and size, including water bottles, and kitchen utensils, etc. We report the average accuracies across all objects.

As shown in Table 2, SONDAR achieves millimeter-level accuracy in size measurement and approximately 93 % accuracy in shape identification under the LOS scenario, which is comparable to vision-based methods. However, unlike vision-based methods, our system can accurately measure the object's size and shape under the NLOS scenario. Additionally, our system shows superior performance over the state-of-the-art system AIM [23], primarily due to the AIM's inability to handle the Doppler effect.



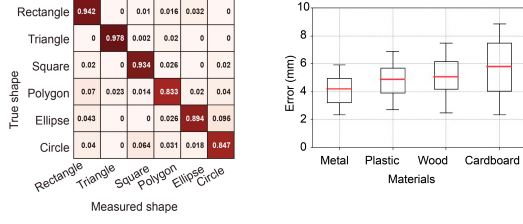


Figure 14: Impact of shape. Figure 15: Impact of material.

Table 3: Impact of cardboard boxes with different sizes.

Size (mm) \ Type	Small	Medium	Large
Actual length $L_a$	140.00	280.50	458.00
Actual width $W_a$	140.00	155.50	336.00
Average measured length $L_m$	136.25	285.70	485.40
Average measured width $W_m$	144.60	161.90	349.75
Average length error $ L_a - L_m $	3.75	5.20	27.40
The percentage of length error	2.68%	1.85%	5.98%
Average width error $ W_a - W_m $	4.60	6.40	13.75
The percentage of width error	3.29%	4.12%	4.09%

## 6.2 Impacting Factor Analysis

**6.2.1 Impact of Target Shape.** We classified the tested objects into six shape categories: circle, rectangle, square, triangle, ellipse, and polygon. Each contains 20 objects. The confusion matrix in Fig. 14 illustrates the shape measurement accuracy. The overall accuracy exceeds 90%, but the polygon's accuracy is lower. This discrepancy is caused by short edges in polygonal objects, which our system may not detect due to weak reflections. Note that our current system only measures the 2D shapes of objects. And it can be extended to obtain the 3D shapes of objects if two sets of devices are deployed.

**6.2.2 Impact of Target Size.** We evaluate the impact of target sizes by conducting experiments using cardboard boxes with different sizes. As listed in Table 3, we can achieve a millimeter-level size measurement accuracy for both small and medium boxes. We can also observe that, as the box size becomes large, the size measurement error also increases. The reason is that a larger box results in weaker signals reflected from edge scatter points due to the limited beam width caused by the speaker's directivity.

**6.2.3 Impact of Material.** Since different materials exhibit various capabilities for absorbing and reflecting acoustic signals, we study how the imaging results vary across different materials. We chose four materials with the same size of  $140\text{ mm} \times 140\text{ mm}$  as targets, including wood, plastic, cardboard, and metal. As we can observe from Fig. 15, the size measurement error for the metal is the smallest due to its highest reflection coefficient.

**6.2.4 Impact of Distance.** We evaluate the impact of the distance between the target and device by varying it from  $30\text{ cm}$  to  $150\text{ cm}$  with a step size of  $30\text{ cm}$ . As shown in Fig. 16, the performance degrades as the distance increases, which is expected due to the lower SNR. However, even when the distance is increased to  $90\text{ cm}$ , we can still achieve a millimeter-level size measurement error.

**6.2.5 Impact of Target Speed.** Target movement introduces the Doppler effect. We evaluate the performance with and without the elimination of the Doppler effect at different speeds. We set the target to move at five different speeds, ranging from  $10\text{ cm/s}$  to  $30\text{ cm/s}$ . As shown in Fig. 17, it is evident that, benefited from

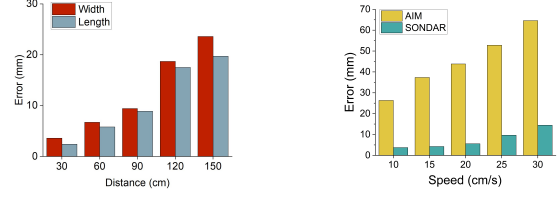


Figure 16: Impact of distance. Figure 17: Impact of speed.

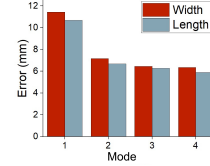
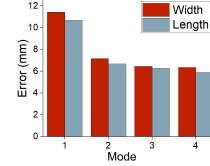


Figure 18: Impact of variable speed.

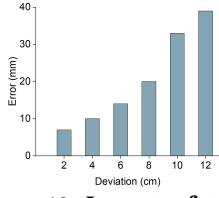


removing the Doppler effect, our system SONDAR outperforms the prior system AIM. Additionally, we find that even after eliminating the Doppler Effect, the error increases with speed. The reason is that, since the trajectory length is fixed, a higher speed indicates less data for imaging, making it more susceptible to noise.

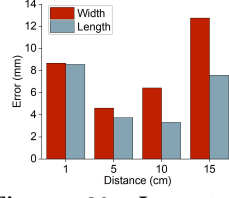
**6.2.6 Impact of Variable Speed.** Target movement may vary and not remain constant. We evaluate the impact of the variable target speed using four different speed modes. Mode 1 gradually increases the speed from  $15\text{ cm/s}$  to  $30\text{ cm/s}$ . Mode 2 gradually decreases the speed from  $15\text{ cm/s}$  to  $0\text{ cm/s}$ . Mode 3 increases the speed from  $15\text{ cm/s}$  to  $20\text{ cm/s}$  and then decreases the speed to  $10\text{ cm/s}$ . Mode 4 decreases the speed from  $15\text{ cm/s}$  to  $10\text{ cm/s}$ , then increases the speed to  $15\text{ cm/s}$ , and finally decreases the speed to  $10\text{ cm/s}$ . As shown in Fig. 18, we can achieve reasonably good performance when the target moves at a variable speed. Additionally, by comparing Fig. 18 with Fig. 17, we can observe that the error in variable speed is larger than that in constant speed. This is because the assumption where the target moves at a constant speed within each profile does not hold if the target speed is variable.

**6.2.7 Impact of Trajectory Deviation.** A target's trajectory might deviate from a straight line, e.g., an individual holding a box and passing through a security gate. We conduct experiments to evaluate how the trajectory deviation impacts our system performance. We asked a participant to manually move the target along predefined trajectories marked on the ground. The trajectory deviations are introduced in the radial direction, varying from  $2\text{ cm}$  to  $12\text{ cm}$  with a step size of  $2\text{ cm}$ . Fig. 19 shows that our system can achieve a reasonably high size measurement accuracy even if there exists a trajectory deviation of  $8\text{ cm}$ . Furthermore, we found that, although the size measurement error is high for a large trajectory deviation, we can still accurately recognize the target shape.

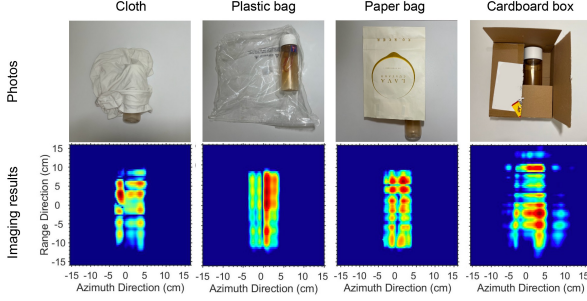
**6.2.8 Impact of Speaker-Microphone Distance.** The distance between the speaker and the microphone varies across devices. We change the speaker-microphone distance to evaluate its impact on imaging. Fig. 20 depicts the size measurement errors when the speaker-microphone distance is set to  $1\text{ cm}$ ,  $5\text{ cm}$ ,  $10\text{ cm}$ , and  $15\text{ cm}$ , respectively. We can observe that the size measurement error first decreases and then increases as the distance increases. The reason



**Figure 19: Impact of trajectory deviation.**



**Figure 20: Impact of speaker-mic distance.**



**Figure 21: Impact of obstruction.**

is that the device cannot capture the reflections from all scatters on the target if the speaker-microphone distance is too small or large.

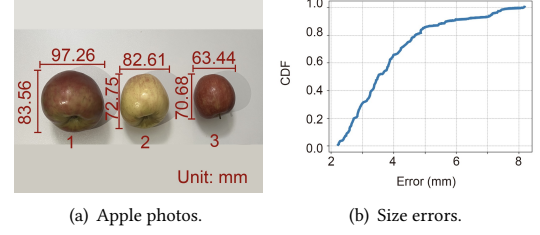
**6.2.9 Impact of Obstruction.** One appealing advantage of acoustic imaging is that it can image a target even under obstruction. As shown in Fig. 21, we conduct experiments to evaluate the imaging performance under different obstruction covers, including cloth, plastic bag, paper bag, and cardboard box. We can observe that we can clearly identify the shape of the water bottle under different covers. However, the contour of the water bottle becomes blurred when it is inside a cardboard box. This is because the cardboard box reflects most of the acoustic sensing signals.

**6.2.10 Impact of Environment Diversity.** We perform experiments in three environments with different room sizes and multipath, including a living room ( $6\text{ m} \times 3\text{ m}$ ), a laboratory ( $20\text{ m} \times 10\text{ m}$ ), and a conference room ( $5\text{ m} \times 7\text{ m}$ ). Benefitted from the background subtraction technique that removes the multipath interference, the size measurement errors are similar across the three environments.

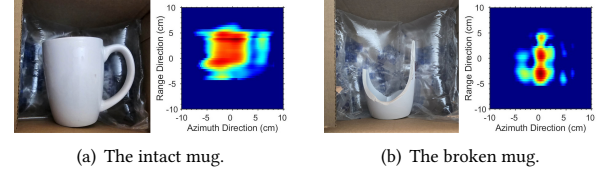
**6.2.11 Impact of the Ambient Noise.** We conduct experiments to evaluate the impact of ambient noises. We introduced four types of noises that are generated by human ( $58.1\text{ dB}$ ), dishwasher machine ( $50\text{ dB}$ ), air conditioner ( $70\text{ dB}$ ), and lawnmower ( $105\text{ dB}$ ). We observe that the size measurement errors for these noises are similar to those for quiet environments ( $36.8\text{ dB}$ ). The reason is that the frequencies of noises are below  $16\text{ kHz}$ , which is lower than the frequency band adopted for sensing. Our interference cancellation module can remove the impact of ambient noises.

## 6.3 Field Studies

**6.3.1 Apple Grading.** Due to variations in fruit sizes, farmers need to grade fruits according to specific size categories to optimize economic gains. However, the conventional method for fruit grading relies on labor-intensive manual processes or the utilization of expensive equipment [2]. We demonstrate that our proposed system makes it possible for fruit grading using cheap acoustic devices. To



**Figure 22: Apple grading.**



**Figure 23: Fragile item breakage detection.**

evaluate the performance of our system in fruit grading, we conducted experiments to image 20 different apples and measure their corresponding sizes. Fig. 22(a) illustrates apples in three grades, i.e.,  $6\text{ cm} - 7\text{ cm}$ ,  $7\text{ cm} - 8.5\text{ cm}$ , and  $\geq 8.5\text{ cm}$ , according to the American apple grading standards [31]. As shown in Fig. 22(b), the size measurement error for apples is less than  $1\text{ cm}$ , indicating that our system can be effortlessly applied to apple grading.

**6.3.2 Fragile Item Breakage Detection.** SONDAR can be used in the courier industry to check the completeness of items in packages on a conveyor belt during the sorting process or verify the integrity of items before customers sign for their packages. We showcase that our proposed system can detect the breakage of a mug without opening the cardboard box using a commodity smartphone. We put an intact mug inside a cardboard box and then manually pushed the box at  $30\text{ cm}$  in front of the smartphone. We did the same procedure for the broken mug. As shown in Fig. 23, if we adopt the image of the intact object as a reference, it is easy to determine if the object is complete without opening the package.

**6.3.3 Box Size Measurement using Smartphone.** We envision that our proposed system SONDAR can be potentially adopted in the express delivery industry, where a courier needs to measure the package size. Our system can accurately measure the packing box size when the courier manually pushes the box in front of a smartphone. As shown in Fig. 24(a), we asked a participant to sit at  $30\text{ cm}$  in front of the smartphone and manually push the box. Fig. 24(b) shows the size measurement errors for three different sizes of packing boxes, including Small ( $279.4\text{ mm} \times 381.0\text{ mm}$ ), Medium ( $279.4\text{ mm} \times 431.8\text{ mm}$ ), and Large ( $381.0\text{ mm} \times 533.4\text{ mm}$ ). Although the size measurement errors for the hand-pushing scenario are larger than those for the guide-trail scenario, the centimeter-level accuracy is sufficient for box size measurement.

## 7 Conclusion

This paper introduces SONDAR, the first acoustic imaging system for measuring the shape and size of moving targets on commodity devices. Inspired by ISAR-based radar imaging, we propose novel methods to address the unique challenges presented in acoustic



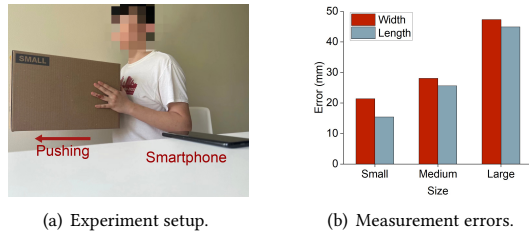


Figure 24: Box size measurement using smartphone.

signals, including Doppler-free echo alignment and frequency-to-physical space mapping. Extensive experiments demonstrate the feasibility and effectiveness of our proposed system. We believe our proposed system can be applied to many real-world size and shape measurement applications.

## References

- [1] Alan Agurto, Yong Li, Gui Yun Tian, Nick Bowring, and Stephen Lockwood. 2007. A Review of Concealed Weapon Detection and Research in Perspective. In *2007 IEEE International Conference on Networking, Sensing and Control*. 443–448.
- [2] Vahid Farzand Ahmadi, Peyman Ziyaee, Pourya Bazayar, and Eugenio Cavallo. 2020. Development and Testing of a Low-Cost Belt- and-Roller Machine for Spheroid Fruit Sorting. *AgriEngineering* 2, 40 (12 2020), 596–606.
- [3] Gary R. Bradski and Adrian Kaehler. 2008. *Learning OpenCV - computer vision with the OpenCV library: software that sees*. O'Reilly.
- [4] Chao Cai, Rong Zheng, Jun Li, Linwei Zhu, Henglin Pu, and Menglan Hu. 2020. Asynchronous Acoustic Sensing Localization and Tracking for Mobile Targets. *IEEE Internet of Things Journal* 7, 2 (2020), 830–845.
- [5] Shirui Cao, Dong Li, Sunghoon Ivan Lee, and Jie Xiong. 2023. PowerPhone: Unleashing the Acoustic Sensing Capability of Smartphones. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–16.
- [6] Victor C. Chen and Marco Martorella. 1753. Inverse Synthetic Aperture Radar Imaging; Principles, Algorithms and Applications. <http://gbv.ebib.com/patron/FullRecord.aspx?p=1782251> Description based upon print version of record.
- [7] Yi-Chang Chen, Gang Li, Qun Zhang, Qing-Jin Zhang, and Xiang-Gen Xia. 2017. Motion Compensation for Airborne SAR via Parametric Sparse Representation. *IEEE Transactions on Geoscience and Remote Sensing* 55, 1 (2017), 551–562.
- [8] Yan Cui, Sebastian Schuon, Derek Chan, Sebastian Thrun, and Christian Theobalt. 2010. 3D shape scanning with a time-of-flight camera. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1173–1180.
- [9] Fabian Friederich, Wolff von Spiegel, Maris Bauer, Fanzhen Meng, Mark D. Thomson, Sebastian Boppel, Alvydas Lisauskas, Bernd Hils, Viktor Krozer, Andreas Keil, Torsten Löffler, Ralf Henneberger, Anna Katharina Huhn, Gunnar Spickermann, Peter Haring Bolivar, and Hartmut G. Roskos. 2011. THz Active Imaging Systems With Real-Time Capabilities. *IEEE Transactions on Terahertz Science and Technology* 1, 1 (2011), 183–200.
- [10] Longsheng Fu, Shipeng Sun, Rui Li, and Shaojin Wang. 2016. Classification of Kiwifruit Grades Based on Fruit Shape Using a Single Camera. *Sensors* 16, 1012 (July 2016). <https://doi.org/10.3390/s16071012>
- [11] Akshay Gadre, Deepak Vasishth, Nikunj Raghuvanshi, Bodhi Priyantha, Manikanta Kotaru, Swarun Kumar, and Ranveer Chandra. 2022. MiLTON: Sensing Product Integrity without Opening the Box using Non-Invasive Acoustic Vibrometry. *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 390–402.
- [12] Michael Y. Jin and Chialin Wu. 1984. A SAR correlation algorithm which accommodates large-range migration. *IEEE Transactions on Geoscience and Remote Sensing* GE-22, 6 (1984), 592–597.
- [13] Atsutake Kosuge, Satoshi Suehiro, Mototsugu Hamada, and Tadahi Kuroda. 2022. mmWave-YOLO: A mmWave Imaging Radar-Based Real-Time Multiclass Object Recognition System for ADAS Applications. *IEEE Transactions on Instrumentation and Measurement* 71 (2022), 1–10.
- [14] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. 2019. PointPillars: Fast Encoders for Object Detection From Point Clouds. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12689–12697.
- [15] Dong Li, Shirui Cao, Sunghoon Ivan Lee, and Jie Xiong. 2022. Experience: Practical Problems for Acoustic Sensing. In *Proceedings of the 28th Annual International Conference on Mobile Computing and Networking (Sydney, NSW, Australia) (MobiCom '22)*. Association for Computing Machinery, New York, NY, USA, 381–390.
- [16] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2020. FM-track: pushing the limits of contactless multi-target tracking using acoustic signals. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems (Virtual Event, Japan) (SenSys '20)*. Association for Computing Machinery, New York, NY, USA, 150–163.
- [17] Zhichao Li, Feng Wang, and Naiyan Wang. 2021. LiDAR R-CNN: An Efficient and Universal 3D Object Detector. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7542–7551.
- [18] Jie Lian, Xu Yuan, Ming Li, and Nian-Feng Tzeng. 2021. Fall Detection via Inaudible Acoustic Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 114 (sep 2021), 21 pages.
- [19] Jialin Liu, Dong Li, Lei Wang, and Jie Xiong. 2021. BlinkListener: "Listen" to Your Eye Blink Using Your Smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–27.
- [20] Jialin Liu, Dong Li, Lei Wang, Fusang Zhang, and Jie Xiong. 2022. Enabling contact-free acoustic sensing under device motion. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–27.
- [21] Ryan Lloyd and Scott McCloskey. 2014. Recognition of 3D package shapes for single camera metrology. In *IEEE Winter Conference on Applications of Computer Vision*. 99–106. <https://doi.org/10.1109/WACV.2014.6836113>
- [22] Bassem R. Mahafza. 2013. *Radar systems analysis and design using MATLAB* (3rd ed. ed.). Boca Raton, FL: CRC Press.
- [23] Wenguang Mao, Mei Wang, and Lili Qiu. 2018. AIM: Acoustic Imaging on a Mobile. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (Munich, Germany) (MobiSys '18)*. Association for Computing Machinery, New York, NY, USA, 468–481. <https://doi.org/10.1145/3210240.3210325>
- [24] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. 2019. RNN-Based Room Scale Hand Motion Tracking. In *The 25th Annual International Conference on Mobile Computing and Networking (Los Cabos, Mexico) (MobiCom '19)*. Association for Computing Machinery, New York, NY, USA, Article 38, 16 pages. <https://doi.org/10.1145/3300061.3345439>
- [25] K L Navaneeth, Ansu Mathew, Shashank Kashyap, Wei-Chih Hung, Varun Jampani, and R. Venkatesh Babu. 2020. From Image Collections to Point Clouds With Self-Supervised Shape and Pose Networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 1129–1137. <https://doi.ieeecomputersociety.org/10.1109/CVPR42600.2020.00121>
- [26] Fangbo Qin, De Xu, Blake Hannaford, and Tiantian Hao. 2023. Object-Agnostic Vision Measurement Framework Based on One-Shot Learning and Behavior Tree. *IEEE Transactions on Cybernetics* 53 (2023), 5202–5215. Issue 8.
- [27] Jacqueline M Schellberg, Hem Regmi, and Sanjib Sur. 2023. mmSight: Towards Robust Millimeter-Wave Imaging on Handheld Devices. *2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 117–126.
- [28] Sibao Sun, Yingchun Chen, Longhao Qiu, Guangpu Zhang, and Chunhui Zhao. 2020. Inverse Synthetic Aperture Sonar Imaging of Underwater Vehicles Utilizing 3-D Rotations. *IEEE Journal of Oceanic Engineering* 45, 2 (2020), 563–576.
- [29] Bela Team. 2023. *Bela Platform*. <https://bela.io>
- [30] Yu-Chih Tung, Duc Bui, and Kang G. Shin. 2018. Cross-Platform Support for Rapid Development of Mobile Acoustic Sensing Applications. In *In Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 455–467.
- [31] USDA. 2019. U.S. Grade Standards for Apple. <https://www.ams.usda.gov/grades-standards/apple-grades-standards>
- [32] Junfeng Wang and D. Kasilingam. 2003. Global range alignment for ISAR. *IEEE Trans. Aerospace Electron. Systems* 39, 1 (2003), 351–357.
- [33] Ju Wang, Jie Xiong, Xiaojiang Chen, Hongbo Jiang, Rajesh Krishna Balan, and Dingyi Fang. 2017. TagScan: Simultaneous Target Imaging and Material Identification with Commodity RFID Devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (Snowbird, Utah, USA) (MobiCom '17)*. Association for Computing Machinery, New York, NY, USA, 288–300. <https://doi.org/10.1145/3117811.3117830>
- [34] Xuanchi Wang, Kai Niu, Anlan Yu, Jie Xiong, Zhiyun Yao, Junzhe Wang, Wenwei Li, and Daqing Zhang. 2023. WiMeasure: Millimeter-level Object Size Measurement with Commodity WiFi Devices. (12 6 2023), 1–26. <https://doi.org/10.1145/3596250>
- [35] Yong Wang, Rui Cao, and Xin Huang. 2020. Cross-Range Scaling of ISAR Image Generated by the Range-Chirp Rate Algorithm for the Maneuvering Target. *IEEE Sensors Journal* 20, 16 (2020), 9124–9131.
- [36] Edward Zhang, Jan Laufer, and Paul Beard. 2008. Backward-mode multiwave-length photoacoustic scanner using a planar Fabry-Perot polymer film ultrasound sensor for high-resolution three-dimensional imaging of biological tissues. *Applied Optics* 47, 4 (2008), 561–577. <https://opg.optica.org/ao/abstract.cfm?URI=ao-47-4-561>
- [37] Bowen Zheng, Guiling Sun, Zhaonan Meng, and Ruili Nan. 2022. Vegetable Size Measurement Based on Stereo Camera and Keypoints Detection. *Sensors* 22, 4 (2022), 1617.