A SEEING MACHINES WHITEPAPER

3D SENSING FOR IN-CABIN MONITORING

SOLVING THE IN-CABIN SENSING PUZZLE

AUTHOR:
TIMOTHY EDWARDS,
CO-FOUNDER & VP STRATEGY

MARCH 2025

seeingmachines

3D Sensing for In-Cabin Monitoring

Contents

Introduction

The automotive industry has been progressing towards a future where vehicles are intelligent, private and where personalized spaces are increasingly able to adapt to the individual needs and characteristics of each occupant. Broadly, this phenomenon is driving demand for more sophisticated and dependable sensing technologies within vehicle interiors.

Cameras were first introduced into vehicle interiors in the mid-2010s, with vision-based driver monitoring entering production in 2017. Camera-based interior sensing is now expanding in scope to encompass real-time monitoring of the *entire cabin, including all occupants and objects that may be present*.

This whitepaper explores the key opportunities and challenges of in-cabin 3D sensing and presents a compelling solution born from the collaboration between Seeing Machines, a pioneer in driver monitoring and in-cabin AI sensing, and Airy3D, an innovator in 3D sensing technology.

Combining Airy3D's patented technology with Seeing Machines' in-cabin domain-expertise, the partnership has forged a solution that transcends key limitations of all other 3D sensing methods. This breakthrough paves the way for a new era of personalized safety, enhanced comfort and seamless human-machine interaction within the vehicle cabin.

The Challenge of Whole Cabin Monitoring

For well over a decade, Seeing Machines and its competitors have been pushing the limits of what can be achieved using real-time highly embedded vision-AI operating with traditional 2D monocular cameras. Advancements in 2D vision, particularly due to machine learning methods that now create very small and efficient AI models able to be executed on low cost SoCs (and in some cases, even on sensor chips themselves) have significantly improved the capabilities of these systems. However, expanding the scope of interior monitoring to encompass the entire cabin has presented car manufacturers with a labyrinth of options regarding the types of sensors to install within and around the cabin. Cabin sensor configuration presents a perpetual challenge, requiring engineers to estimate dozens of trade-offs between system cost, performance, while working within constraints of interior real estate. Additionally, beyond the sensor configuration issue lies the intricate task of integrating the data from these sensors to achieve optimal results.

Cost is King

A key factor affecting decisions concerning sensor configuration and vehicle processing architecture is overall solution cost. One of the main cost drivers is the number of sensor modules. The expense associated with incorporating supplementary cameras into the vehicle cabin is not solely attributable to the cost of advanced sensor chips or optical elements. It is more basic than that. Each sensor module possesses an inherent irreducible cost owing to its mechanical framework, which comprises metal, plastic, glass, components, connectors and wires. These modules require mass production, calibration, connection, and testing as integral steps within the vehicle assembly process, all of which adds up to dollars that are very difficult to reduce.

Consequently, augmenting the functionality of a single sensing module, as opposed to integrating additional modules, tends to be a more efficacious strategy for enhancing the value proposition of interior monitoring systems. Indeed, the industry has witnessed the emergence of 5MP RGB-IR image sensors, which exhibit superior capabilities compared to the initial

1.2MP IR image sensors employed in original driver-specific monitoring systems, as a testament to this logic. These advanced sensors now facilitate the acquisition of high dynamic range images characterized by a wide field-of-view with high frame-rates, and enable the monitoring of the entire front row of the vehicle from a single camera unit, including the precise gaze direction of the driver and passenger, their head orientation, their facial appearance and identity, their body-pose, hand gestures, whether they are eating, drinking, smoking, chewing, talking, resting, sleeping, and so on. The outcome of using a better image sensor, is a far more cost-effective approach than deploying two lesser cameras within the front row.

Be that as it may, achieving all necessary cabin monitoring functionality from the viewpoint of a single 2D camera is very technically demanding, with only some suppliers being able to technically support this configuration. Consequently, some automakers have incorporated multiple camera modules to achieve commercial supply objectives.

Multi-Camera Systems

A central issue that troubles the single camera approach is the fact that extracting a 3D measurement from a 2D image of a scene requires algorithms that have a-priori knowledge about the scene's content, including the probability distribution of the dimensions of objects and features. For instance, to determine the location of the driver’s head in 3D, the probability distribution of adult head dimensions must be encoded into the algorithm. The 3D accuracy is then limited by the uncertainty in the scale of the object being observed. This is referred to as the 2D scale ambiguity problem and is covered in more detail later.

An advantage in employing multiple 2D cameras is the ability to triangulate measurements, yielding (opportunistically, for some frames) improvements in the quality of the 3d solution. Additionally, it can be argued that a second viewpoint provides better immunity to occlusion.

RADAR and UWB Technology

Another sensor technology becoming common in vehicle cabins is RADAR, which is primarily used to detect the presence of occupants and to alert vehicle owners if they accidentally leave a child or pet behind in a locked vehicle.

Additionally, Impulse Radio Ultrawideband sensing (IR-UWB) is a promising sensing and communication technology which is starting to appear in vehicles. The technology is already in Apple phones and AirTags and for vehicles to detect the proximity of a vehicle owner to intelligently unlock the vehicle when they are close-by. IR-UWB can also be repurposed as a form of RADAR, offering basic occupant sensing features at a low system cost.

The Cabin Sensing Maze

It is therefore unsurprising that we witness a complex ongoing debate in the industry as to the best configuration of sensors and what amounts to the necessary coverage for cabin sensors. The debate is reminiscent of what has previously occurred in the automotive industry regarding exterior sensing.

Today there are no clear standards for in-cabin sensor placement. Due to the variation and complexity of cabin interiors and the rapid progression of sensing and compute techniques, *such standards may never stabilize for very long*, even within the same car company.

A common approach to solving the sensor configuration puzzle is to undertake a “feature-based” analysis, where each existing cabin-monitoring feature is performance tested against a specific sensor configuration, where the best sensor configuration is assumed to be the one that yields the highest overall feature benchmark score for the least amount of cost. While this approach sounds sensible, it is also limited by the assumptions made in the development of the existing features used to drive the benchmark, which are usually algorithm paradigms that will have evolved with existing sensing technology (mainly 2D cameras). The "feature based" analysis approach does not offer comparisons with algorithms that may have been built to operate with 3D-sensor data from the ground up.

A more sensible and holistic approach is to go back to first principles, and ask what is the common underlying problem to be solved? If we forget about all the details of today's in-cabin features for a moment, the question of which sensor configuration is better, should arguably be expressed as *which system configuration can yield the highest quality three-dimensional real-time model of a vehicle's interior for any given cost*.

We refer to this real-time 3D model of the vehicle interior, as the Interior Perception Map (IPM). The IPM is a model that consists not only of 3D range information, but also colours and textures of shapes, the latter being formed from the more traditional 2D sensor information and which is often overlooked when evaluating 3D sensor performance.

Using this framework, Seeing Machines engineers have thrown away assumptions and have been studying the cabin sensing problem from this fundamental IPM perspective. In that context, we have been asking ourselves if future vehicle cabins will be best monitored by numerous 2D cameras, RADARs, UWB sensors, some configuration thereof, or, if there is a way to reduce the overall number of sensing modules, and reduce the software complexity, through incorporating 3D-vision sensing?

The obvious first conclusion was that compared to vision, both RADAR and UWB sensors carry less overall raw information about the cabin-scene compared to optical sensing, simply due to differences in wavelength. Optical systems are superior as they offer a far richer information stream and do so at a lower overall cost. The richness of the information stream also implies that not only can today's feature requirements be achieved, but that there is also a healthy roadmap for the future, unlocked by applying ever-improving AI.

Therefore, while RADAR and UWB will make limited improvements, the value derived from optical systems will improve at a faster rate, and be able to deliver the long-tail of concepts desired by OEMs, to differentiate their vehicle brands, and enrich the driving experience.

So, this leaves 2D vision sensors vs 3D vision sensors. Here we ask ourselves; can 3D-vision sensors really compete with the maturity, simplicity and low costs of 2D sensors?

The History of Automotive Interior 3D Sensing

Looking back, it is a fact that the automotive industry has already explored introducing 3D sensing technology in vehicle cabins. Early examples include Time of Flight (ToF)-based gesture recognition systems first introduced in luxury European brands in the mid-2010s. Secondly, there have been stereo-vision systems, most notably the stereo-camera developed for the Mercedes-Benz S-Class (W233), which is integrated into the instrument cluster.

This latter stereo-camera measures precise 3D eye-position to support both a 3D instrument cluster display as well as an AR HUD system on the windshield. In particular, a very accurate 3D eye position is required to ensure accurate graphical overlays on the windshield, which must be precisely located as if to appear on the road surface itself, regardless of the driver's head position and orientation.

More recently, AR HUD systems have been introduced that use monocular 2D sensing, but while these systems work reasonably well, their capabilities are ultimately limited by the uncertainty in the range to the eye inherent from 2D-based eye-tracking, which in some circumstances will translate to incorrect graphical alignment being presented to the driver, representing a potential safety hazard.

While earlier forms of 3D sensors have so far failed to penetrate beyond luxury vehicles, it is important to remember that accurate and reliable 3D sensing remains the only way to deliver these (and many of other) value-adding vehicle features.

There should be no doubt that accurate 3D range data is powerful and useful information that can unlock new levels of sophistication in the vehicle interface. So why has 3D sensing not become commonplace in vehicle cabins? Because of the cost. But why have the costs not fallen by now? The reason is because the underlying technical methods are fundamentally resistant to cost optimization.

However, 3D technology is not standing still. Our research indicates that the cost issues associated with 3D sensing will soon to be overcome, and we therefore believe that there is significant future potential for the *right kind* of 3D sensing to unlock levels of safety, comfort, and convenience that, due to a mixture of physics and software complexity, 2D systems will be unable to compete with.

Importantly, falling costs will mean these advancements will not be limited to the luxury car segment and have the potential to fully permeate the global fleet.

Key Benefits of 3D Sensing

Key benefits include:

- **Improved System Dependability**: Today's vehicles regularly annoy occupants by issuing false warnings. In many scenarios these false warnings even create their own safety hazard by distracting or confusing the driver. Increasing the threshold or allowing a driver to disable the warning, is how car markers generally deal with this issue. However, this is not a great solution as doing so increases the chance that the system will fail to warn the driver when it really ought to. False detection phenomena are often caused by technical limitations of what can be achieved with software that operates on 2D images.
- **Personalized Safety**: the improved dependability that 3D sensing offers will allow deep integration of in-cabin vision sensing with other vehicle safety systems. For example, the response of a safety system will be able to be tuned to the characteristics of each individual in the cabin, adjusting the response of collision avoidance, airbag deployment and restraint systems, based on personal variables such as age, mass distribution, body posture and seat configuration.
- **Reduced Solution Cost and Complexity**: When compared at a system-level, 3D sensors will soon offer a simpler and cheaper solution versus combinations of 2D sensors and RADARs. Utilizing natively derived 3D range data, the cabin-monitoring software complexity is also reduced.
- **Comprehensive Object Detection and 3D Localization**: Beyond occupants, 3D sensing enables the detection and tracking of any object within the cabin, including objects that an AI model may never have witnessed in its training set. This capability will be crucial for identifying potential projectiles in emergency braking or collision scenarios, allowing for timely warnings or even proactive safety interventions that can account for loose objects in the cabin.
- **More Intuitive and Personalized Vehicle Interfaces**: The enhanced reliability and accuracy provided by 3D sensing will allow for safe and correct real-time adjustments of

seats, mirrors, directional interior lighting, augmented reality displays, directional and independent audio channels for each occupant, and more accurate interpretation of gestures combined with eye-gaze and voice commands. These are all features that 2D driver and occupant monitoring were supposed to enable, but so far, have failed to do so, largely due to the difficulty in ensuring a consistent experience for everyone in all driving conditions.

Software Complexity

Today's 2D-based cabin monitoring systems employ software stacks composed of numerous, carefully trained but narrowly focused vision-AI models, typically connected in a logical hierarchy. In this hierarchy, each model operates on a specific portion of the 2D-image and must inherently and locally "solve for 3D" within the limited scope of the vehicle feature it is designed to support. This approach leads to several challenges:

- **Increased complexity**: As more features are added, the hierarchy grows increasingly complex, due to dependencies forming between models which then present a higher risk of unforeseen interactions and instabilities.
- **Difficult debugging**: Identifying and resolving issues within this hierarchical model structure can be challenging, as errors can propagate through the system and obscure the root cause.
- **Increased retargeting costs**: Adapting the software stack to new cameras, camera positions, or cabin geometries becomes increasingly expensive and time-consuming, as each model needs to be recalibrated and potentially retrained. Retraining any model potentially leads to retraining of all dependent models.

3D sensing opens the door to a fundamentally different approach. By directly capturing depth information, it provides a holistic and unambiguous measurement of the cabin environment in real time. This eliminates the need for multiple individual models to infer 3D information from limited 2D data, leading to several key benefits:

- **Simplified software architecture**: With 3D information readily available, the software stack can be streamlined, reducing the size and number of models and their interdependencies. This results in a more modular and maintainable system.
- **Improved robustness**: The reduced complexity and reliance on 2D inferences leads to a more robust system, less prone to errors and instabilities caused by cascading failures within the model hierarchy.
- **Easier retargeting**: Adapting the system to new configurations becomes simpler and more efficient, as the core 3D perception is less sensitive to changes in camera parameters or cabin geometry. Ultimately, a 3D perception layer offers a way to isolate vehicle features from the underlying sensor configuration.

The inherent software simplification offered by 3D sensing is a significant advantage hidden beneath the surface of these software systems and the magnitude of this complexity effect is only growing as in-cabin systems continue to evolve and incorporate more and more features.

By providing a solid foundation of reliable 3D information, 3D sensing thereby promotes a more elegant and scalable software architecture, paving the way for greater innovation and faster development cycles.

Challenges of 3D Automotive In-Cabin Sensing

While the appeal of 3D sensing is abundantly clear, achieving robust and reliable 3D sensing within the dynamic and complex environment of a vehicle cabin is no easy feat. Varying lighting conditions, occupant movements, reflective surfaces, and occlusions can all hinder the accuracy and consistency of the IPM.

Creating a robust 3D perception system within the confines of a vehicle cabin presents a unique set of challenges. Unlike other environments, the vehicle cabin presents numerous factors that work to impede accurate and reliable depth perception. To achieve the full potential of in-cabin 3D sensing, the following challenges must be addressed:

1. **Robustness**:
    - **Varying Lighting Conditions**: The vehicle cabin experiences dramatic shifts in ambient lighting, from bright sunlight streaming through the windshield to dim nighttime driving. A reliable 3D sensing system must perform consistently across this wide array of lighting conditions.
    - **Temperature**: Vehicle electronics typically need to operate from -40 to +85 degrees C ambient, which can mean that electronic components that dissipate thermal energy need to be able to function correctly above boiling point. Thermal dissipation and heat management is always a major design issue even for 2D cameras, often driving additional costs to mitigate heat.
    - **Occupant Dynamics**: Vehicles can impart significant g-forces, and passengers are constantly moving, adjusting their posture, and interacting with the vehicle's interior. The system needs to accurately capture every movement and maintain reliable 3D perception even with highly dynamic occupants.
    - **Scale Ambiguity**: Common objects such as phones, handbags or backpacks, do not have precise sizes, but occur in a distribution of sizes. Consider the example of a phone handset, where dimensions may vary by as much as 10% or more. A 2D system, relying solely on visual cues and prior knowledge, can only estimate the size of a detected phone within this range of uncertainty. This directly translates to uncertainty in the estimated distance to the phone. 3D sensing seeks to be immune to the object scale ambiguity phenomenon.
    - **Unknown Objects and Complex Occlusions**: Vehicle cabins may also contain unknown objects of unknown shapes, sizes, and materials. Occlusions caused by passengers, seats, and other objects can complicate depth perception. The ideal system must be able to handle these complexities and provide accurate depth information even when there are occlusions by random objects.
    - **Reflective and Absorptive Surfaces**: The presence of glass, mirrors, and other reflective surfaces within the cabin can cause interference and distort depth measurements. Similarly, surfaces that tend to absorb light, such as textured fabrics, can cause reliability issues for active illumination schemes.
    - **Near and Far Range Limits**: 3D sensing in cabins must handle occlusions from very close-range objects while also being able to operate to 2-3m. For example, if the camera is located near the rear-mirror, it will be common for occupants to place their hands very close to the sensor itself, but the sensor might see through the fingers of the hand and must still work to image the rear seat.
2. **Imaging Characteristics:**
    - **Wide Field of View**: To capture the entire cabin and all its occupants, a very wide field of view (FOV) is essential, typically 140 degrees or more. This is a particularly important requirement, as a reduced FOV just drives the need for additional cameras, which drives up system cost.
    - **High MTF[1]**: It is crucial that the optical system can resolve fine details sufficient particularly over the eye regions of the occupants. This is to support eyelid, pupil and gaze tracking

---

[1] Modulation Transfer Function (MTF) describes how well an optical system, like a lens and sensor stack, is able to transfer detail of the scene to an image of the scene.

which in turn are key for many advanced interface features. In particular the vision system needs to be able to resolve subtle facial features and eye movements to accurately assess forms of driver impairment.

- **Performance Parity**: *The 2D image quality from the 3D sensing system must be comparable to, or exceed, that of dedicated 2D cameras*. This is particularly important for demanding applications like driver monitoring, where high-resolution images with excellent contrast and dynamic range are essential. Performance parity also allows 2D software stacks to be compatible with the 3D sensing, allowing for a controlled software migration to 3D.

3. **Computational Efficiency and Latency:**
    - **Real-time Processing**: For time-critical safety applications like airbag deployment, pre-collision interventions, or 3D augmented reality, the latency of the real-time processing is crucial. Any processing delays will compromise the effectiveness of these safety-related systems.
    - **Balancing Complexity with Speed**: Achieving high accuracy and robustness in 3D sensing often requires complex algorithms and significant computational resources. The challenge is balancing this complexity with the need for low latency to ensure timely responses.
4. **Cost and Integration:**
    - **Cost-Effectiveness**: For widespread adoption, 3D sensing solutions must be highly cost-effective and offer a compelling value proposition compared to alternative technologies.
    - **Seamless Integration**: The system should be easily integrated into existing vehicle architectures, minimizing design complexity and manufacturing costs.

Overcoming these challenges is essential for realizing the full potential of 3D sensing for vehicle cabins.

Existing 3D Sensing Technologies

The quest for robust and reliable in-cabin 3D perception has led to the exploration of various sensing technologies, each with its own strengths and weaknesses. Understanding these trade-offs is crucial for selecting the optimal solution to meet the specific requirements of in-cabin applications, particularly given the constraints of the automotive environment.

Stereo Vision

Stereo vision, inspired by human binocular vision, employs two cameras to capture slightly different perspectives of the same scene. Depth information is extracted by analysing the disparity between pairs of images. While capable of generating detailed depth maps, stereo vision systems face several challenges in the automotive context:

- **Computational Complexity**: Processing two image streams and performing the necessary calculations for depth estimation demands significant computational resources. This can lead to latency issues, especially in real-time applications like occupant safety, where immediate responses are critical.
- **Low-light Performance**: Stereo camera systems struggle in low-light conditions that are common during nighttime driving and so demand additional illumination. This is because low light reduces image quality which hinders accurate disparity matching, impacting depth accuracy and reliability.
- **Sensitivity to Repetitive Patterns**: The abundance of repetitive patterns in car interiors, such as seat fabrics and dashboard textures, can confuse stereo algorithms and lead to inaccurate depth measurements due to localized visual ambiguity.

- **Hardware Complexity**: Stereo vision requires two image sensors, two lenses, and potentially specialized hardware for image synchronization and processing. This doubles the cost of image capture components and adds complexity to system design and calibration. Stereo-cameras must also be calibrated across temperature and be able to detect and self-report loss of stereo calibration.

RADAR

Most interior RADAR systems operate using a modulated continuous wave method, known as Frequency Modulated Continuous Wave (FMCW). Here the transmitter emits “chirp” waveforms (where the frequency is swept across a band over time) and the receiver mixes (or multiplies) the transmitted waveform with the received energy from the scene, and in doing so effectively correlates the transmitted pulse to the received signal in the frequency domain, yielding a beat frequency that is directly proportional to the range of the target.

Frequency shift in the received chirp due to the Doppler effect can be used to detect very small motions. While RADAR range measurements are typically accurate to around 10cm, the Doppler-derived velocity measurement is very sensitive and precise, allowing for detection of tiny movements, such as the heartbeat or breathing of a living creature, or vibrational movements from the environment, such as those that can occur in concrete parking structures.

Another RADAR signal processing method applied to vehicle interiors is Pulsed Coherent RADAR (PCR). PCR differs from FMCW, by the energy being transmitted in pulses having relatively higher peak power. PCR measures the time delay of the pulse echo reflected from the target scene and observes the phase-difference between transmitted and received signals, yielding enhanced range accuracy. The increased peak power of PCR also allows greater penetration into materials than FMCW.

Most interior RADARs operate within 57-66 GHz, which are frequencies that can penetrate thin fabrics and some types of clothing (e.g. cotton but not leather). Higher frequency systems (e.g. 140GHz) are also becoming available, and these RADARs theoretically improve accuracy and reduce antenna and mechanical package size due to the shorter wavelengths.

It is often argued that RADARs are more able to deal with occlusion than regular visible or near-infrared cameras due to radio wavelengths being more penetrative, but this is a very marginal advantage over vision in vehicle cabins.

Of note, both 60GHz and 120GHz wavelengths are not able to penetrate the human body to “see the heart beating”. Rather, RADARs detect heart and respiration through detecting small movements on the surface of the body caused by internal heart and lung activity. The problem is these motion signals become buried by noise when a vehicle is on a rough road, jostling the occupants.

It may also be a surprise to learn that certain wavelengths of near-infrared light penetrate human tissue far more deeply than RADAR, due to an optical window in the absorption spectrum of water. Some wavelengths of near infrared penetrate human tissue by up to 4-5 cm which is why infrared light can be used effectively to monitor vital signs.

The radiative and scattering properties of interior scenes at 60 and 140GHz wavelengths also mean interior RADARs are prone to complex scene-dependent multipath reflections, making them susceptible to false measurement. For example, placing a bicycle into the cabin may trigger false occupancy warnings that will annoy the vehicle owner.

As a mental model, imagine the interior scene of a vehicle but with every surface coated with highly reflective metallic paint of very slightly varying reflectivity. A cave of mirrors.

Some vendors use multiple-input multiple output (MIMO) antenna setups. MIMO systems are more expensive, but offer two main advantages (1) they provide signal redundancy by receiving multiple independent reflections which help to mitigate complex multipath echoes, ultimately improving the system SNR, and (2) improving spatial resolution by effectively synthesizing a larger aperture than the physical antenna array (known as a “virtual array”).

RADARs ultimately suffer from very low spatial resolution due to their wavelengths, creating limitation in the practical beamwidths that can be achieved, having at best, angular resolutions of around 10 degrees, whereas a 5MP image sensor, with a 140 deg lens, offers better than *0.05 degrees* of angular resolution.

As a form of 3D sensor, an interior RADAR provides a very low-resolution image (maybe 15x5 pixels), but where each pixel has an associated motion signal (which is very sensitive), and a distance measurement with about 10% error.

Ultra-Wideband (UWB) Sensing

Most people will be familiar with UWB technology being deployed in beacon systems like Apple’s AirTag which is used for accurate short-range sensing; able to measure the distance of the beacon relative to the phone handset to within about 10cm. However, UWB can also be applied with fixed transmitter and receiver positions to operate as a form of 3D sensor.

UWB operates between 3.1 and 10.6 GHz and the “wideband” in UWB refers to the fact that the bandwidth of the emission spectra, relative to the mid frequency is greater than 20%.

UWB follows the same fundamental signal techniques as pulsed time-domain RADAR, but at lower frequencies (longer wavelengths). Intuitively, longer wavelengths ought to reduce resolution, however UWB technology emits radio pulses that have extremely short pulse widths (measured in nanoseconds, and thus a wide bandwidth). The spread-spectrum nature of the pulse design ensures that individual pulses are largely immune to sources of noise within the UWB band from other wireless transmissions. Importantly, UWB employs pulse groups, with inter-group pulse amplitude and timing variations that enact encoding schemes. These pulse groups are also known as symbols. Symbol codes offer improved timing resolution versus a single pulse (trading range with accuracy), but also allow reduced power, improved sensitivity and further immunity to noise.

Sequences of symbols are also used to carry additional information payloads, such as encrypted data, allowing communication between a UWB system in the vehicle, and UWB in a phone handset, as well as other devices within or around the vehicle. So UWB not only offers RADAR-like positioning and sensing, it provides a secure communications channel, making it a very attractive proposition for car-makers.

Like a RADAR, a UWB receiver operates as a matched filter, convolving the transmitted multi-pulse encoding with the received energy reflected from the scene, but importantly due to the lower frequencies (relative to RADAR), much of the signal processing can be performed digitally, with fewer analog mixing stages needed before received energy is converted to digital information via high-speed ADCs. Consequently, compared to RADARs, UWB systems reduce the number of high-frequency analogue components in the receiver compared to the FMCW and PCR RADAR approaches, reducing cost.

From a signal processing standpoint, UWB can be optimized to achieve a near-perfect channel impulse-response in the receiver. This means very faint signal echoes can be detected from surfaces despite low received-power levels, while also achieving very accurate range measurements relative to the wavelength. At present UWB typically achieves 10cm accuracy over 1-2 m range, relative to a mid-wavelength of about 4.4cm.

Similar to RADAR, velocity information can be measured by observing doppler shift in the received pulses, and so tiny movements can be detected. UWB is therefore able to detect surface movements of the chest like the other RADAR techniques. In addition, the longer UWB wavelengths penetrate the human body, meaning that at higher power levels, the technology has the (long-term) potential to detect vital signs more directly than just observation of surface motion. However, the issue of physical motion masking heart and lung movements will remain difficult to overcome.

UWB systems employ angle of arrival (AoA) techniques (either time-difference or phase-difference approaches) using multiple antenna elements. This allows the reflected pulse echoes to be resolved spatially, but the angular resolution is anticipated to be at best, 3 degrees depending on the number of antenna elements, which remain large and bulky, representing one of the main drawbacks for UWB modules.

UWB will also be compatible with MIMO techniques to improve spatial resolution. Additionally with pulse encoding, Multiple UWB systems can potentially transmit simultaneous orthogonal signals (that do not interfere with each other) despite being transmitted at the same time across the same frequencies. The output of those systems could be combined to further improve spatial resolution.

In short, the true capabilities of UWB for interior sensing are yet to be fully explored. However, based on the fundamentals, seen as a form of 3D camera, a very advanced UWB system might “see” a cabin scene as a series of semi-transparent surfaces, perhaps revealing *very blurry*, ghostly forms for cabin occupants and surfaces, while also offering an encrypted short-range inter-device communication network.

UWB is an exciting technology and is expected to become available in most new electronic devices and vehicle cabins. However, as a 3D sensor it remains very limited. The world we experience and have built is designed for human eyes and ears. UWB as a form of sensing is limited by its use of radio frequencies, which are unable to achieve the resolution required to effectively interpret the human environment.

Both RADAR and UWB methods fail to provide powerfully useful information such as colors, texture, fine-structures, and key human features such as eyelids, iris, pupils, facial expressions, mouth movements, which collectively represent the natural means of human social engagement, and which therefore remain essential signals for modern intelligent vehicles to understand about their occupants.

Time-of-Flight (ToF) Cameras

From a signal processing standpoint, ToF cameras follow similar techniques to UWB and RADAR systems, just at optical wavelengths. ToF cameras have the obvious advantage of having far denser spatial resolution, with new sensors having over 1 million pixels.

Direct (dToF) and indirect (iToF) sub-variants exist, with iToF being the most promising approach for vehicle cabins. The fundamental difference between the two methods is like the differences in PCR and FMCW RADAR described earlier. dToF directly measures the time it takes for a light pulse to travel to an object and back, offering greater accuracy over longer distances and better ambient light immunity, resulting in cleaner, less noisy depth maps, but potentially at a lower 2D image resolution due to sensor design constraints and higher cost. Conversely, iToF determines distance by measuring the phase shift of modulated light, enabling higher resolution and frame rates, leading to potentially sharper 2D images and more detailed depth information in short to medium-range applications like gesture recognition and in-room mapping, but it can be more susceptible to interference, have limitations with long-range accuracy, and produce noisier depth maps. In essence, dToF excels in range, accuracy, and depth map clarity, while iToF prioritizes resolution, speed, and potentially better 2D image quality.

State of the art iToF cameras available on the market today are required to work with infrared light sources that modulate the light at megahertz frequencies. They also employ specialized pixel structures and custom ASICs to receive the light and derive the depth information.

As a form of 3D sensor, there are important limitations for ToF sensors when applied to cabin monitoring:

- **Compromised 2D Image Quality**: The specialized pixel design of ToF sensors requires a design trade-off to accommodate the range measurement functionality. The result is a far lower image quality compared to dedicated 2D sensor pixel designs, particularly in terms of resolution, dynamic range, SNR and sensitivity. Unfortunately, this has so far ruled out ToF for applications that require measurement of small image features over wide FOVs.
- **Limitations in Resolution and FOV**: Achieving both high resolution and a wide field of view can be challenging with current ToF technology. This can limit its applicability in scenarios where both detailed depth information and comprehensive scene coverage are required.
- **Cost and Complexity of High-Frequency Modulation**: To achieve the necessary precision in time-of-flight measurements, ToF sensors require high-frequency modulation of the emitted infrared light. This necessitates specialized circuitry capable of switching several amperes of drive current at megahertz frequencies, leading to increased component cost, design complexity, and the significant risk of EMC issues.

Structured Light

Structured light is a 3D vision approach that operates by projecting a known pattern of light into the scene and analysing information in the reflected pattern to determine depth. Our studies have shown that the approach can achieve about 2% range accuracy between 0.2m and 1.2m of range in a vehicle cabin environment, but also faces several challenges:

- **Sensitivity to Ambient Light**: Ambient light, particularly strong sunlight, can interfere with the projected pattern and degrade the accuracy of depth measurements. This makes it challenging to maintain consistent performance throughout varying lighting conditions.
- **Challenges with Materials**: Shiny or reflective surfaces, or dark and absorptive surfaces, both common in car interiors, can scatter or absorb the projected light, making it impossible to accurately reconstruct the pattern and extract depth information in some situations, leading to “scene-dependent” performance.
- **Optics Hardware Overhead and Thermal Considerations**: Structured light necessitates a dedicated light source, typically a VCSEL (Vertical-Cavity Surface-Emitting Laser) or a projector, along with associated driver circuitry and optics. This adds a non-trivial component cost, increases power consumption, and introduces another heat source that must be managed within the vehicle's thermal design. Due to the coherent phase of the

light, the power of these light sources must be carefully restricted to avoid injuring the human eye.

- **Low Spatial Resolution:** Structured light systems are only able to sparsely sample the cabin scene, providing depth information at a limited number of points spaced across the field of view.
- **Near and far field range limits**: Structured light systems must trade spatial density with near vs far range limits. The far range limit may also be limited by the power of the VCSEL illumination against the reflection of dark materials, which in turn is bounded by eye-safety concerns.

Monocular Depth Estimation (MDE)

Monocular Depth Estimation is a field of machine learning concerned with recovering depth information from only 2D image information. It is best thought of as a form of *generative AI*.

The approach offers the major advantage of being a purely software solution, able to operate with any existing 2D camera. Our research has shown that the method can yield surprisingly accurate relative-depth information, as there is a significant quantity of depth information that can be extracted from 2D scene appearance alone. MDE accuracy appears best in image regions where the scene is a closer match to the training data but can be relatively inaccurate in regions where novel or unusual scene structure is present.

In summary, MDE faces some significant challenges in the context of in-cabin monitoring:

- **Reliability Concerns**: Like all known generative AI systems, MDE models are susceptible to periodic instability in the form of "hallucinations," potentially generating inaccurate depth estimates at moments in time where the cabin scene is complex or visually ambiguous. This core reliability issue means MDE may be unsuitable for safety applications where consistent and dependable performance is paramount.
- **Dependence on Training:** The accuracy and performance of MDE relies heavily on the quality and diversity of the training data. Ensuring robust performance across the wide range of scenarios encountered in a vehicle cabin can require extensive and costly data collection and annotation efforts, with an unknown upper-limit on the quantity of training data required to eliminate all unstable phenomena. Whilst Seeing Machines has world-leading synthetic data facilities which we can use to drive down the costs of MDE training, a key uncertainty remains; how much data will be enough for an MDE model to be trusted in all driving scenarios? MDE is operating with the disadvantage of not having any physical measurement of range to “anchor” its inferences of depth.
- **Limited Generalizability**: MDE models trained on specific datasets may not generalize well to new environments or scenarios, thus like most AI models, MDE models will require retraining (or fine-tuning) for different vehicle models, or whenever the camera locations or cabin configuration change significantly. These issues may again be greatly reduced with synthetic data technology, but there will remain a per-vehicle-model tuning cost that will be incurred anytime the camera or cabin changes.

A New Paradigm: Transmissive Diffractive Masks and Deep Learning

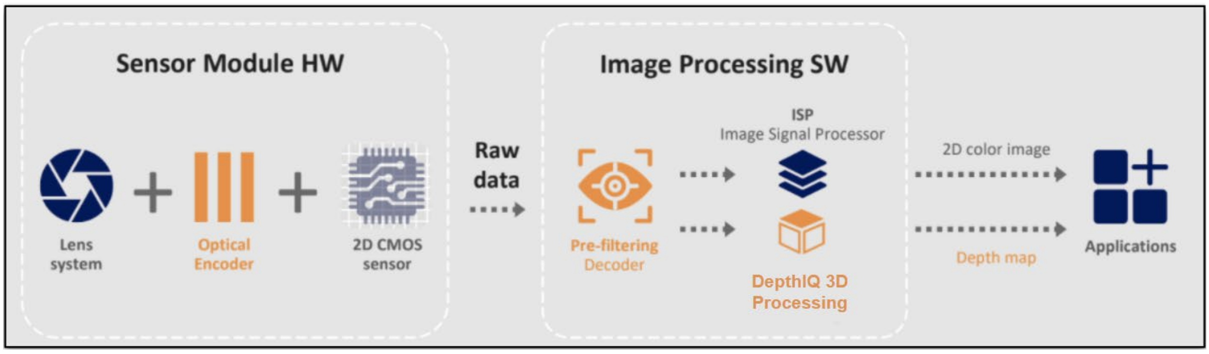
As discussed, all known existing 3D sensing technologies have serious limitations that ultimately hinder their widespread adoption for in-cabin applications, and so the search for a low-cost, reliable 3D sensor suitable for vehicle cabins continues…

Fortunately, Seeing Machines has identified a technology that we feel can overcome these challenges and deliver robust, dependable, accurate, and cost-effective 3D perception within the unique constraints of the automotive cabin environment.

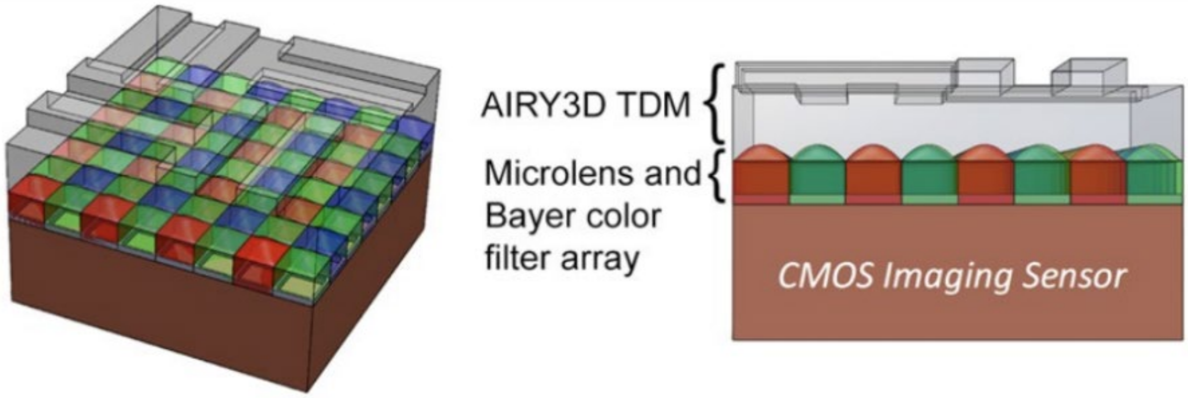
Introducing Airy3D DepthIQ^{TM} Technology

Airy3D is a Canadian optics technology company that has pioneered a truly unique and highly pragmatic approach to 3D sensing. DepthIQ utilizes the phase information in received light, combining it with 2D information from the scene structure to yield the best overall outcome with the lowest optical cost.

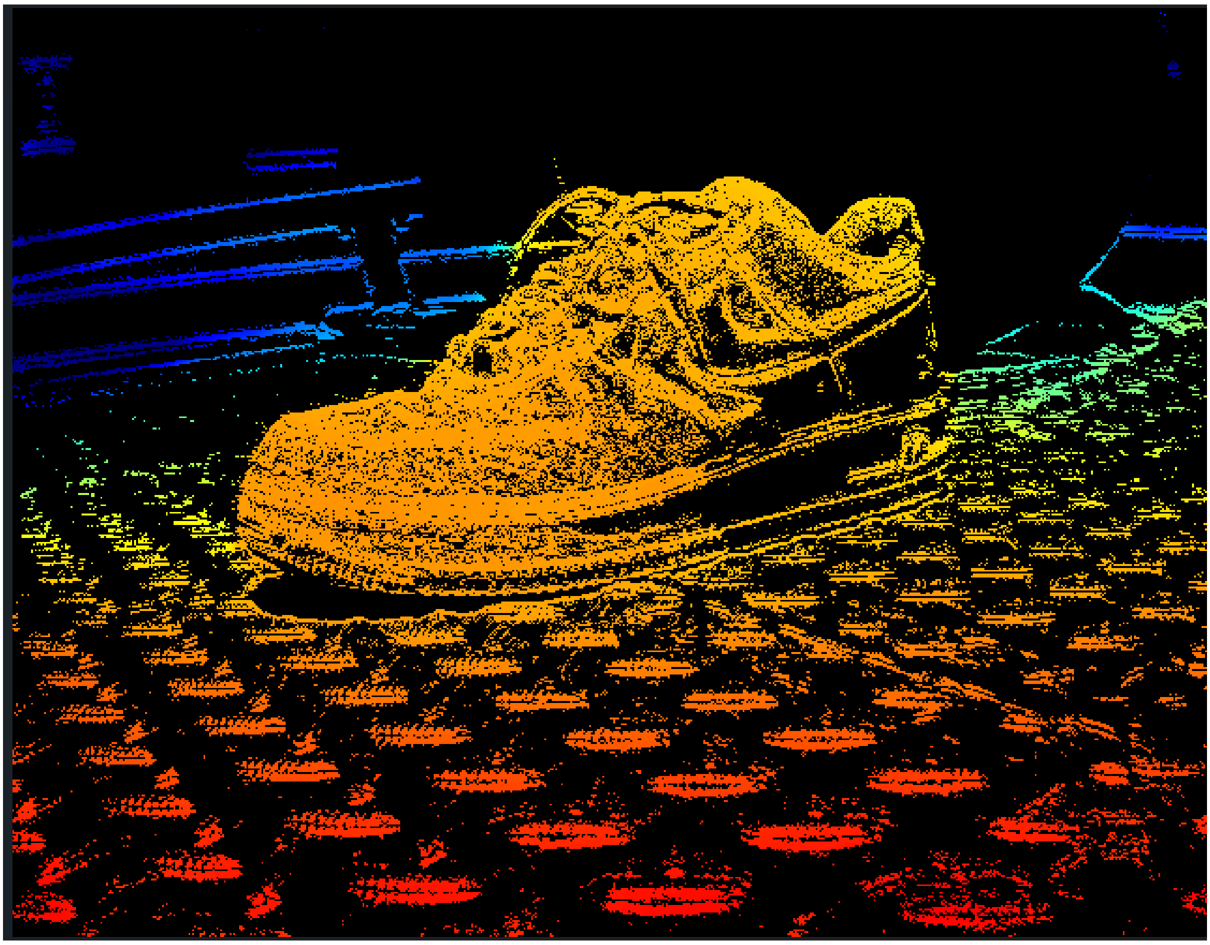
DepthIQ technology uses a combination of an optical encoder, matched to a software decoder, to achieve accurate depth perception without the mechanical and optical cost drawbacks of the other 3D sensing methods.



Above: *Airy3D's breakthrough approach offers the lowest cost of optical hardware for any 3D sensing approach, requiring only a single additional lithographic fabrication step applied to the CMOS image wafer. Cost is instead pushing into the computational domain, where it is optimized away with embedded AI techniques.*



Above: *Illustration of the Transmissive Diffractive Mask, as a lithographically etched layer that sits atop the micropixel structure. The mask is manufactured from optical material that is compatible with the extreme temperature ranges required by automotive image sensors. Because the TDM mask can be applied to any 2D sensor as an end-of-line coating, Airy3D of a "drop-in" compatibility which is unparalleled by any other 3D sensing method and represents a major advantage of the technology.*



Above: *Airy3D and Teledyne E2V partnered to develop an industrial 2MP sensor, the Topaz5D, released in 2024. The sensor and software solution offers sparse phase-based depth information, which appears in the image wherever there are strong edges. This sensor targets robotics, AR/VR headsets and other non-automotive markets.*

Key Details

For details about the TDM optical technique, refer to this 2023 publication by Pascal Grégoire, Niloufar Faghihi, Alexandre Favron and Gil Summy, where they present a mathematical model for the depth sensitivity of the optical design.

Generally:

- **Optical Encoder - Transmissive Diffractive Mask (TDM):** At the heart of Airy3D's technology lies the optical TDM encoder. This is a precisely engineered transparent diffraction grating which is lithographically deposited as the final layer on the image sensor wafer. The diffractive mask pattern is uniquely tailored to the sensor pixel design. The mask is etched using standard low-cost lithographic techniques, adding only a trivial additional cost to sensor fabrication. The TDM modulates incoming light based on angle of incidence, effectively mapping phase-difference into an intensity response across the image array.
- **Pre-filtering Decoder.** While the TDM encodes phase-difference, the *a-priori* known TDM mask geometry and pixel response allows for precise recovery of the 2D colour and intensity information without any significant losses. This is similar in nature to fixed-pattern noise removal and is low-compute step that allows any TDM-modified sensor to be fully compatible with any pre-existing 2D software stack. Our measurements have confirmed a negligible MTF system loss in the RBGIR 5MP sensor supplied by ST Microelectronics.

- **Depth Decoder – DepthIQ AI Model**: DepthIQ extracts depth information using an AI model that processes the raw sensor data (both intensity and phase-difference) to generate depth information. The native resolution of the phase-derived depth information is ¼ of the sensor native resolution and the phase-derived depth information is spatially distributed around edges in the image. However, when phase information is mixed with intensity, like an MDE model, the DepthIQ model can yield depth at any desired resolution and be fully dense if so desired. The key here is the flexibility offered by this software approach; by being able to train different AI models to best fit the requirements of the system including available computational resources.

Utilizing Both Light-Phase and Scene Structure: As described, DepthIQ AI models can utilize not only the phase information of the received light but also the scene structure in the 2D image. The key takeaway is that both phase and intensity are related by the scene structure and therefore a depth model that observes only intensity, or phase in isolation, will be inferior to one that can operate on *both phase and intensity at the same time*.

In other words, DepthIQ models are similar MDE models but have the advantage of incorporating an additional dimension of measurement (phase). Practically speaking, the phase information acts to effectively “anchor” and thereby mitigate the generative-AI stability issues witnessed in MDE models.

Our Research

Seeing Machines, Airy3D, ST Microelectronics and Sunex have worked together to develop a world-first Airy3D-enabled automotive sensor and lens solution, specifically designed for in-cabin monitoring systems.

The main technical challenges were to:

(i) Simulate and design the sensor + TDM + lens optical stack to cover the wide field-of-view (WFOV) for a vehicle cabin (160 degrees) while yielding the optimal depth sensitivity.

(ii) Tune the TDM mask to operate at the infrared wavelengths used by interior monitoring systems, matching it to the ST-VT1940’s novel RBGIR pixel design.

(iii) Verify that the technology will operate successfully under automotive lighting and temperature extremes.

The research has yielded a first prototype camera (shown below) and also sharpened our understanding of the technical issues required to further fine-tune the technology.

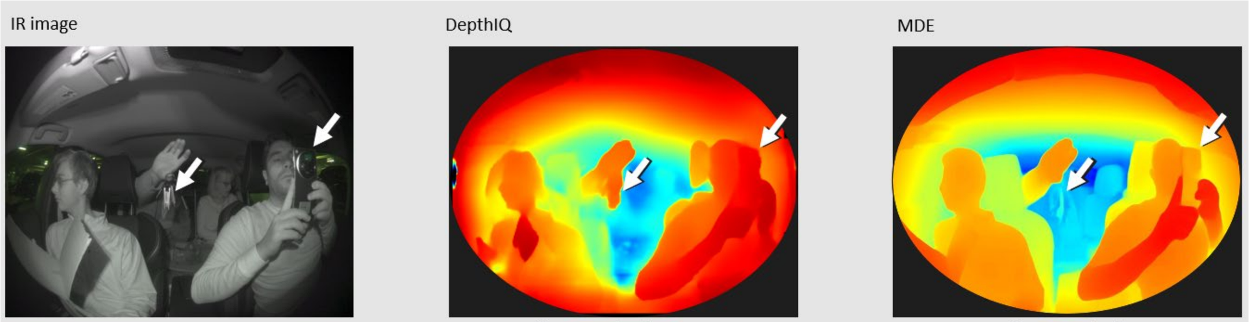
Stability Testing vs MDE

A calibrated Kinect Azur system was used to compare 3D accuracy between a state-of-the-art MDE method (DepthAnythingV2) and Airy3D DepthIQ.

Both models used were general-purpose models. Critically, both were trained on the same datasets; a mixture of random scenery with a fraction of that training data being vehicle cabins.

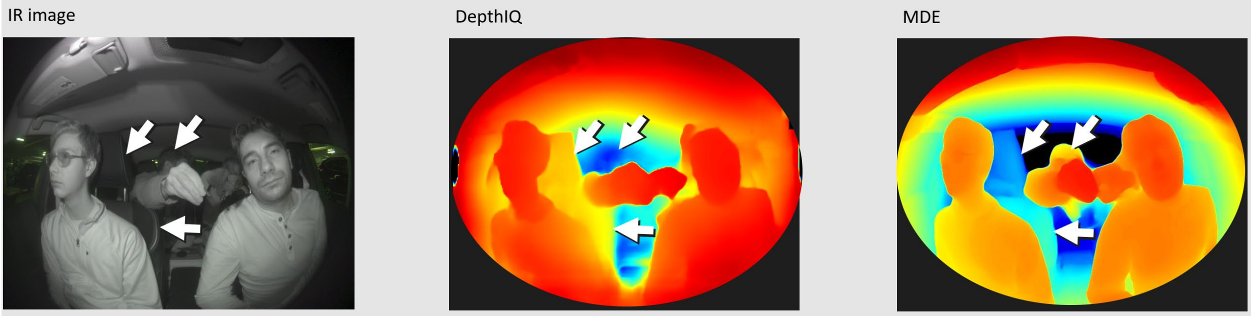
This means the models have **not** been fine-tuned on a vehicle-model-specific cabin, which is a very important distinction when understanding accuracy outcomes.

Above: *Example images from validation database, containing 834 images in total.*



Keys and phone:
Correct Depth

Keys and phone:
Incorrect Depth



Headrest &
Rear Seat Passenger,
Vehicle-right seat:
Correct Depth

Headrest &
Rear Seat Passenger,
Vehicle-right seat:
Incorrect Depth

Above: *Examples where MDE makes z-ordering "mistakes" thereby inducing significant errors where DepthIQ does not. These examples are relatively common when inspecting any recording frame-by-frame.*

Ratio of Absolute Relative Errors MDE / DepthIQ (%)	
Face Region	**Whole Cabin**
209%	237%

Above: *DepthIQ reduces the depth error by a significant margin relative to MDE. While MDE depth data appears high-quality to casual observation, it is periodically inconsistent, occasionally producing incoherent depth due to incorrect ordering of objects in the z-dimension. In comparison, DepthIQ data appears noisier and has less precise depth transitions but overall offers a more accurate, consistent and dependable measurement. This dependability will be a critical issue for integration into future passive safety systems as described by the Euro NCAP 2029 roadmap.*

Discussion

The purpose of the testing performed was to reveal the error characteristics of a general purpose state-of-the-art MDE model against DepthIQ as applied to wide field-of-view cabin monitoring in vehicles.

The testing confirmed that DepthIQ depth data is inherently more stable due to the additional information available in the light-phase being utilized by the model, whereas the MDE must try to infer depth from scene-structure alone and is therefore prone to the kinds of instabilities that commonly occur in generative AI models.

Again, it is important to recognize that both the MDE and DepthIQ models used for this study are general-purpose and have not been fine-tuned against a specific vehicle cabin. This is intentional to help reveal gross characteristics in noise stability.

Fine-tuning either MDE or DepthIQ against vehicle-model specific data is also known to greatly improve accuracy outcomes, but exact accuracy figures remain commercially sensitive. Here we perform an apples-to-apples study to examine the ***relative*** *performance difference between the two technologies when trained and validated on identical datasets.*

Similarly, while it is tempting to compare the results to ToF systems which can yield very accurate depth information, it should be remembered that the Airy3D camera used here is able to simultaneously deliver high-quality 5MP RGBIR 2D images and has been confirmed able to support the full 2D ICMS software stack, including precision eye-gaze tracking.

Summary

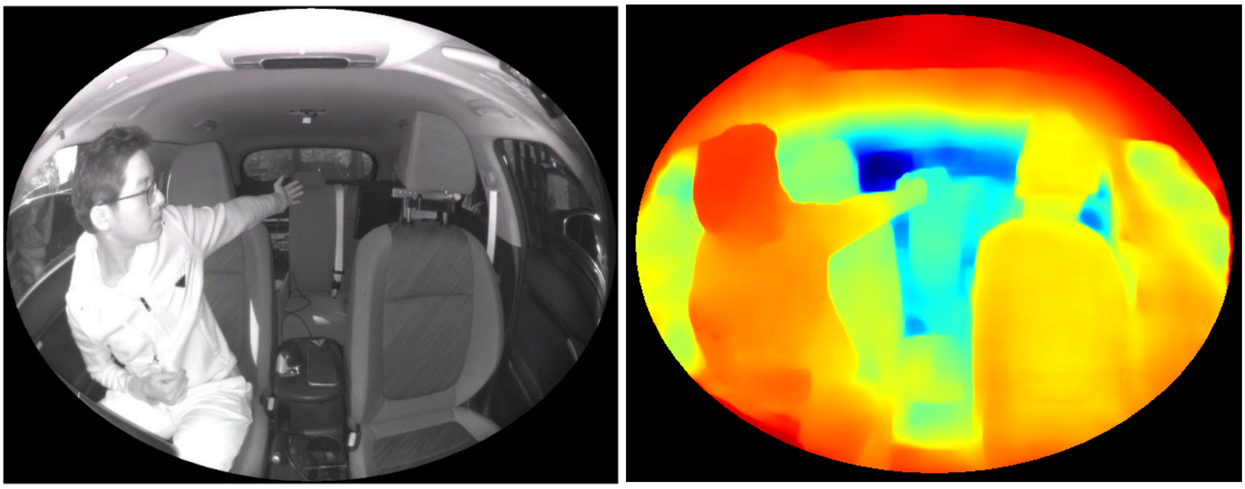
Working with both Airy3D and ST Microelectronics, Seeing Machines has developed a reference camera design for cabin monitoring that offers significant advantages over competing sensing technologies:

- **High-Quality 2D Imaging**: Unlike ToF sensors, which often compromise 2D image quality, the Airy3D approach utilizes standard image sensors, enabling high-resolution, high-dynamic-range imaging suitable for demanding applications like driver monitoring.

- **Reliable 3D Perception**: The combination of phase-difference measurements and deep learning algorithms provides accurate and robust depth estimation, overcoming the limitations of MDE, which can be susceptible to hallucinations and inaccuracies.
- **Robust Performance in Varying Conditions**: The system's robustness to ambient light and its ability to handle complex scenes and dynamic occupants ensure reliable performance in the challenging environment of the vehicle cabin.
- **Cost-Effectiveness**: Through having a very minor optical cost impact leveraging existing 2D sensor technology and through harnessing the AI-compute cost-reduction curve, the Airy3D DepthIQ solution is expected to quickly evolve into the most bang-for-buck sensing method for vehicle cabins.

Our collaborative efforts have resulted in a 3D sensing solution that is uniquely positioned to meet the evolving demands of in-cabin perception, enabling a new era of safety, comfort, and personalized experiences within the vehicle.

Above: *Prototype camera module built around the ST-VB1940 5MP RGBIR sensor with a custom TDM layer applied. The lens is custom designed by Sunex to deliver a 160-degree field-of-view with a carefully tuned depth of field gradient. All optics here are inherently passive; the illumination requirements identical to conventional 2D systems used in today's vehicles and will work with conventional LEDs. Unlike ToF or structured light approaches. The technology promises the lowest hardware cost of any "true" depth sensing technology.*



Above: *High-quality IR data and accompanying depth "heatmap" delivered by the prototype camera. While more accurate depth data can be obtained from other 3D-specific technologies, high-quality 2D IR images remain key for interior-sensing. The camera balances high-quality 2D, with true 3D at minimal cost.*

Future

Through several years of research, Seeing Machines has worked with Airy3D to explore the fundamental limits of their unique technology. Through this work we have convinced ourselves that DepthIQ offers the most plausible 3D sensing solution for automotive vehicle interiors. While others have focused on 3d accuracy, we instead prioritize 2d image quality and cost.

Having completed this first phase of fundamental research, we are now working with our partners to further optimize the technology both on the optical encoder side, and the software decoder side to further improve accuracy and reduce computing costs.

For the optical encoder, improved pixel isolation methods are poised to yield further improvements in light-phase sensitivity, and further innovation by Airy3d is expected to yield more advanced TDM grating designs, which when combined will further improve encoding fidelity.

For the software decoder, Seeing Machines continues to work with Airy3D on AI methods, as well as with silicon partners, to evolve models, and drive the compute into “AI-ISP” designs, targeting the processing cost and latency requirements needed to support automotive safety use-cases.

Expressions of Interest

For more information about this 3D technology for in-cabin automotive applications, please contact: timothy.edwards@seeingmachines.com

For any other enquiries relating to Airy3D technology, contact Airy3D at: info@airy3d.com