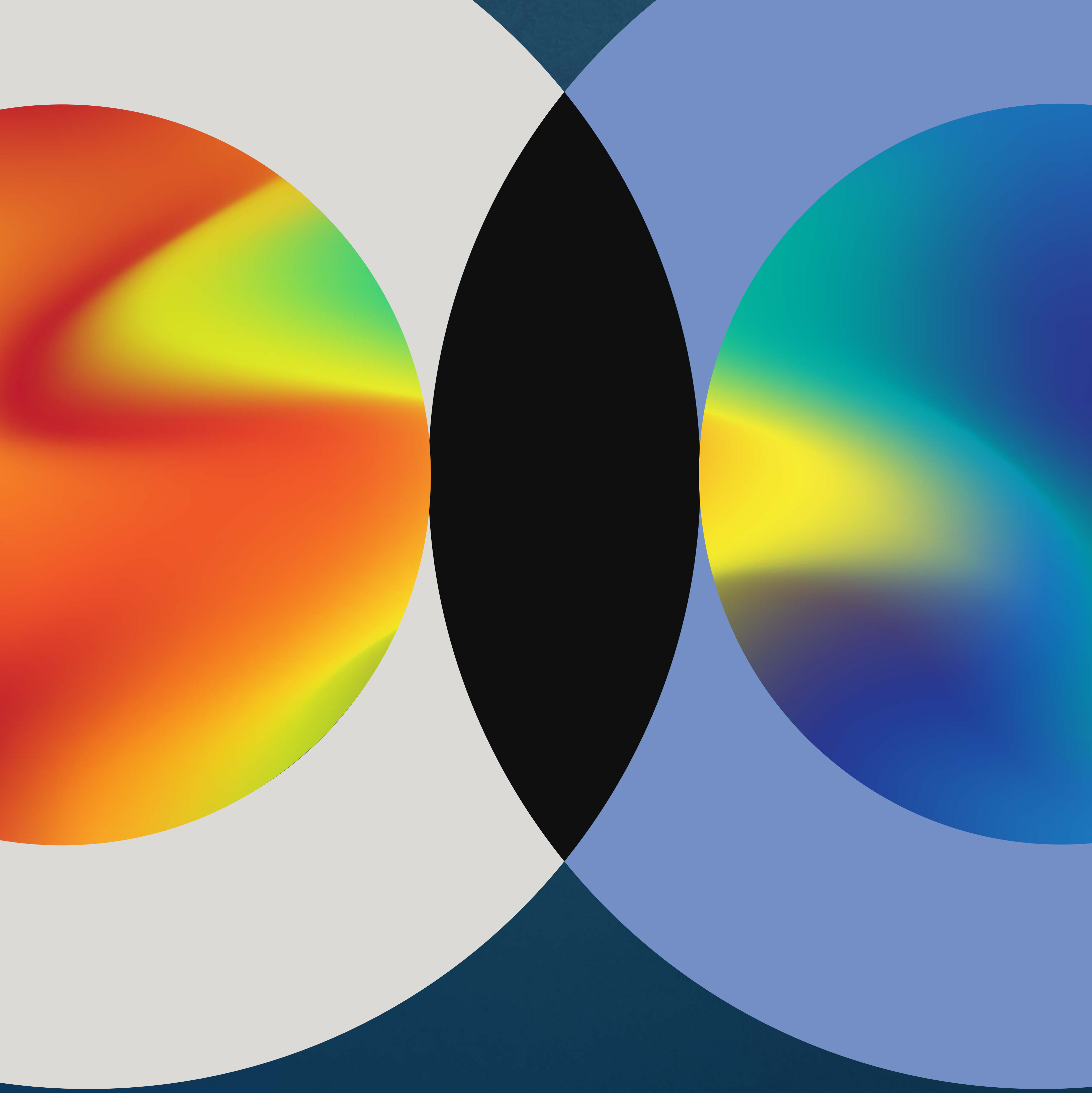


OWL AUTONOMOUS IMAGING • 2023

REQUIREMENTS for EFFECTIVE PEDESTRIAN IMAGING

OWL

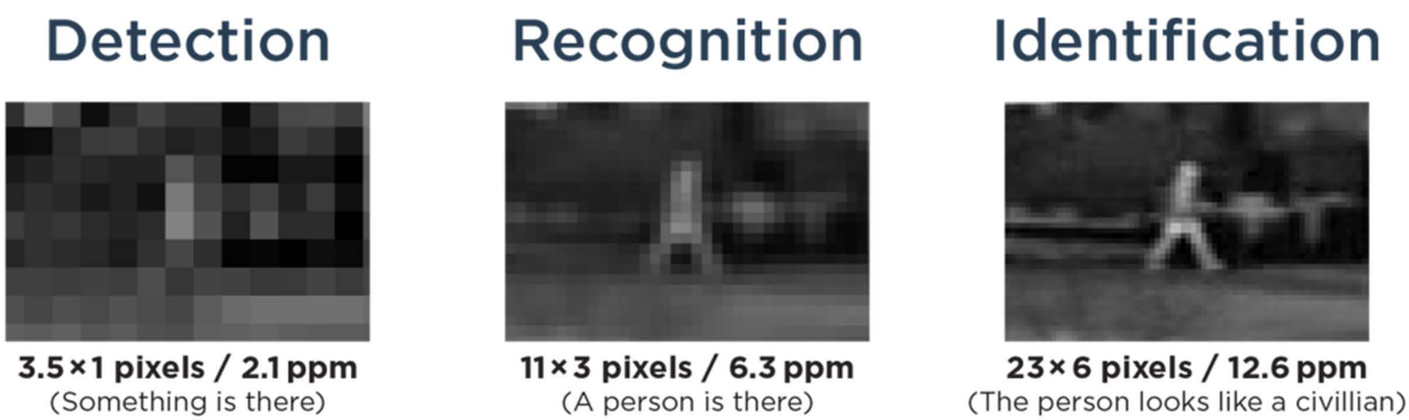
Vol. 6
5/2023

REQUIREMENTS FOR EFFECTIVE PEDESTRIAN IMAGING

When imaging pedestrians from a vehicle, two requirements contribute significantly to the likelihood of successful detection—sufficient detail to separate pedestrians from the background and a broad enough field of view to assure that all vulnerable pedestrians are seen. Combining these requirements defines the resolution required in the camera viewing the scene. This calculation is worth reviewing to facilitate selection of suitable cameras and lenses appropriate to pedestrian detection.

THREE LEVELS OF PERCEPTION

In a previous white paper, we briefly introduced the connection between sensor resolution and the perception task assigned to an observer using the Johnson criteria as shown here [1]:



While these three levels of perception are sufficient for the original defense applications, the variety of objects present in traffic scenes prompts one more level—Classification—where the system needs to know more detail than Recognition provides but does not need all the detail required for Identification.

In these images, the unit is pixels per meter (ppm) as measured at the object to be perceived so the requirement is to provide enough pixels at the location of the most distant pedestrian to be perceived so that the convolutional neural network (CNN) will reliably and accurately tag that pedestrian with a position in the image frame and a distance. This constraint produces a simple relationship between finding pedestrians and the width of the camera coverage, called the horizontal field of view (HFOV). Taking the minimum required resolution for Classification (an operation between Recognition and Identification) to be 10 pixels per meter, then a VGA camera with 640 pixels horizontally can cover 64 meters (211 ft).

Compared to the typical street width, somewhere in the 25-meter range, 64 meters seems like it should be sufficient for protection of pedestrians but, where improved safety is the goal, initial impressions should be verified by careful analysis. In this case, the necessary field of view must be carefully defined.

WHERE ARE THE PEDESTRIANS?

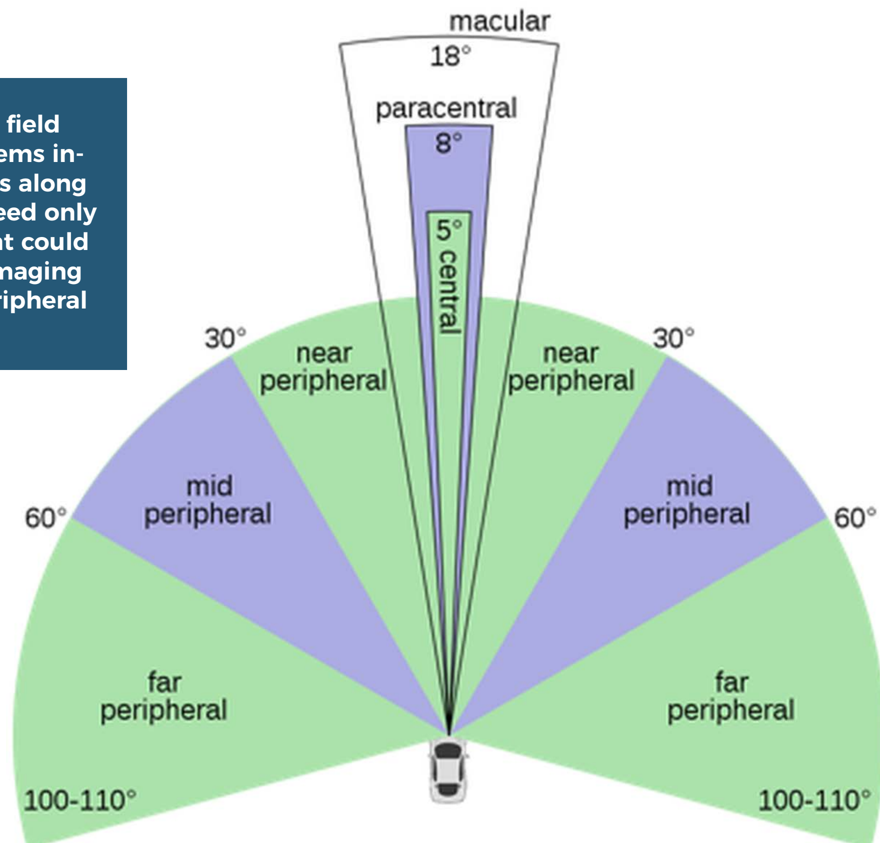
The driver of a car can see, through the windshield and door windows, a scene covering more than 180 degrees horizontally. The center third of this view is most clearly observed [2] but when significant features appear outside that third, the driver's head can rotate to determine the importance of the feature. This behavior is driven by the structure of the retina, where visual receptors are more dense in the center of the field of vision, an area called the fovea. Electronic sensors with similar variation in detector density is designated as "foveated".

Since rotating a camera is not a practical solution to providing full side-to-side coverage of road situations, a camera must be able to simulate the performance of the eye without moving. The basic requirements for this depend on where pedestrians must first be detected as they come into view from the sides or from directly ahead at safe distances.

COORDINATE SYSTEMS

The goal of a Pedestrian Automatic Emergency Braking (PAEB) system is to bring the car to a stop before collision with a pedestrian. The distance required for this is a straight line between the car and the pedestrian. From the car's view, this distance forms a contour that is a circle with the car in the center and the pedestrian on the circumference, as shown in this illustration [3] with the various segments of the field of view labeled according to human visual function:

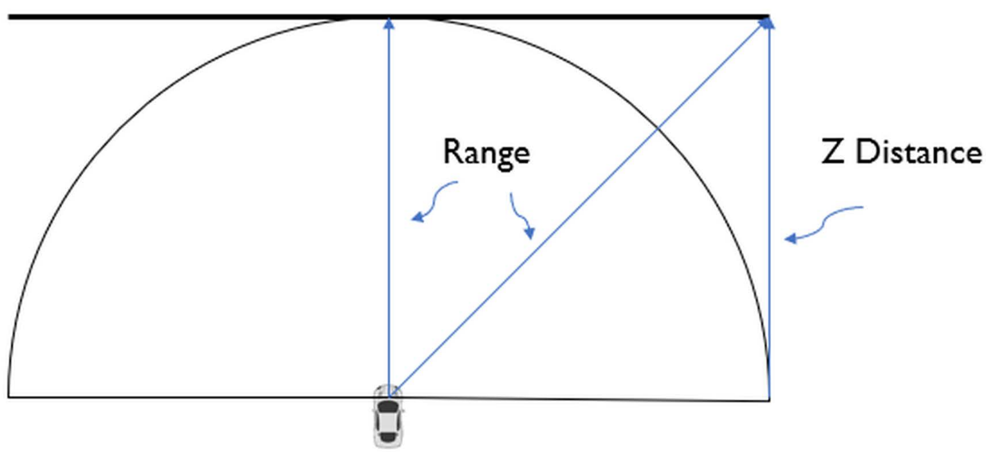
While the human eye can monitor a field of view exceeding 180 degrees, systems intended to avoid contact with objects along the direction of travel of a vehicle need only watch for objects on trajectories that could likely intersect the vehicle's path. Imaging that extends partly into the mid-peripheral zone provides the needed coverage.



COORDINATE SYSTEMS continued

This leads to adoption of a polar coordinate system to define points in the field of view where each point is defined by an angle measured left and right from directly forward called the **azimuth**, a distance from the vehicle called the **range** and a second angle defined vertically above and below a line parallel to the plane of the road called the **elevation**. Note that this is different from the terminology applied to the position and pose of the vehicle itself relative to an external coordinate reference—{x, y, z, yaw, pitch, roll}.

To avoid distortion in images acquired by a camera, the lenses used are of the rectilinear type, which map points on a distant plane to points on the sensor, almost always onto a detector array with elements on a square grid. Thus, the points in the image represent equally-spaced positions on a plane at some distance from the camera. These are identified by Cartesian coordinates {x, y, z} with x and y set to zero perhaps in the center or a corner of the detector array while z is the distance from a plane at the front of the camera perpendicular to the camera optical axis. This distance and the range, as defined above, are equal only on the optical axis. Everywhere else, the range is longer than the z distance as shown here:

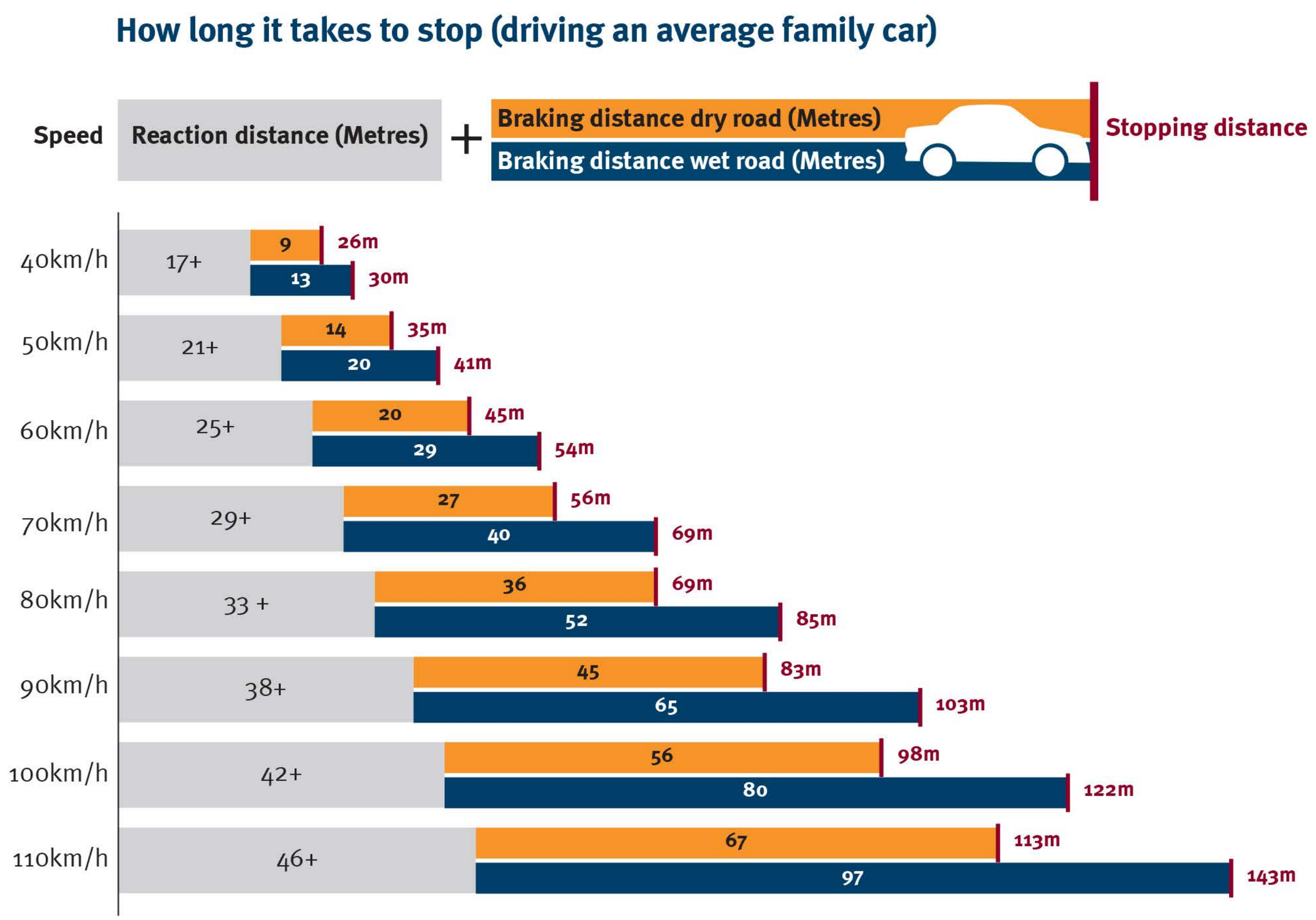


An appropriate field of view is essential in avoiding collisions. If a ring of pedestrians on the circle were all walking to the left at 10% of the speed of the car, those in most danger are in a narrow band to the right of the car's path. The danger extends farther to the right if the VRU is a runner or a cyclist extended to 45 degrees or more if the VRU is a motorcyclist. Somewhere in that range, approaching VRUs of all types will pass behind the vehicle.

This distinction is important because, at any angle, the apparent width of an object will be smaller at the z distance than at a range equal to the z distance that falls on the circle. The math involves application of various trigonometric functions that show that objects on the circle appear to be of constant size when facing the plane intersecting the front of the car. Fortunately, because human bodies can be approximated as cylinders, the direction they face does not affect the size of the image they produce, simplifying the calculation a little. The net effect is that a pedestrian is perceived to be of a constant size anywhere on the circle and progressively smaller in size when located farther and farther in distance from the center of the vehicle in any direction.

APPROACHING A PEDESTRIAN

Stopping before a collision requires adequate time for deceleration after an object is detected. Charts of stopping distances have been developed to provide guidelines. This one from Australia [3] is representative:

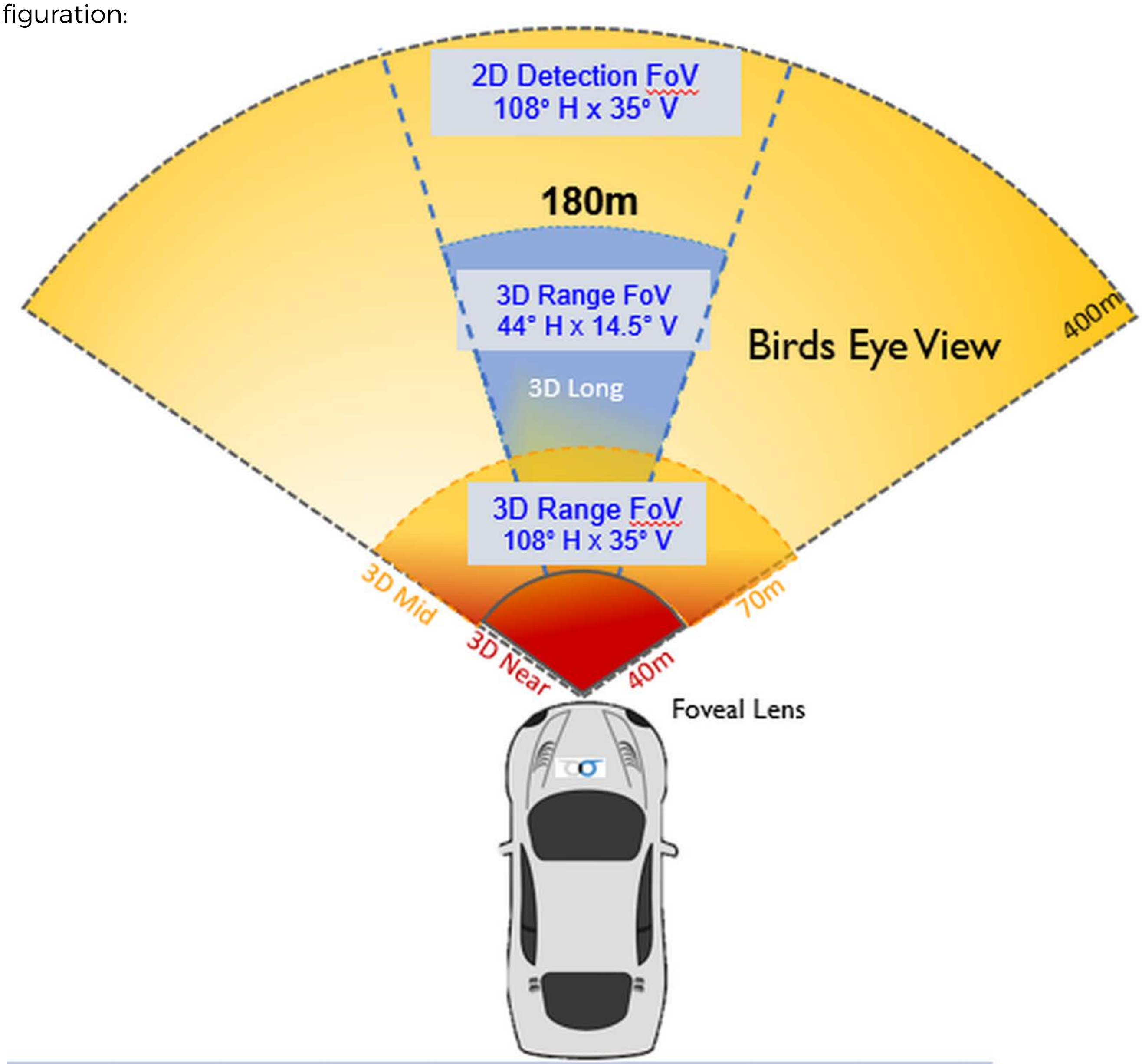


Braking distance increases with speed. Notably, up to 70 km/h most of the braking distance is generated by human reaction time, on the order of 750 ms. The reaction time of autonomous emergency braking systems is much shorter than that, reducing the braking distance significantly.

The total braking distance includes both the distance traveled while the driver is recognizing the hazard and the distance traveled during brake application. While the typical detection time for an electronic system is generally faster than the human reaction time, the traditional values will be used here for conservatism. The conclusions to be drawn from the chart are that pedestrian distances up to at least 150 meters should be measurable on the highway while 70 meters should be sufficient on city streets.

TECHNICAL REQUIREMENTS

To summarize, the total system field of view should be comparable to that of a human driver and the capture of detail at the locations of pedestrians at risk should meet the Johnson criteria. Combining these requirements with the capabilities for sensor and optical fabrication led to this achievable configuration:



Use of a high-definition sensor with a wide-angle lens provides the curb-to-curb coverage needed in urban driving combined with sufficient resolution for reliable classification across the field of view. Split-field optics increase the range directly forward for highway use while training the CNN to detect objects as soon as possible extends the warning range to 400 meters.

TECHNICAL REQUIREMENTS continued

The camera needs a wide field of view mode, as in the example on page 6, covering 108 degrees horizontally. At 70 meters, the arc length on the circle is 132 meters, indicating that the sensor should have at least 830 pixels per line to meet the Recognition criterion and preferably near 1300 elements to provide some margin at 10 ppm. The sensor resolution was selected to be 1280 pixels per line for technical reasons.

On the highway, the range needs to be at least 150 meters but the angle of approach of pedestrians at risk is much smaller than in the city. In the long-range mode, the camera was designed to have a field of view of 40% of the wide field, 44 degrees. The maximum operating distance for this mode at the same resolution is then 70 meters x 2.5 or approximately 180 meters. In both modes, the resolution is sufficient to meet the Johnson criteria and provide some margin for the convolutional neural network (CNN) to determine the range of recognized pedestrians.

Like the eye, the CNN can continue to detect objects to the lower level of the Johnson criteria, 2.1 ppm. In the wide field mode, this corresponds to an arc length of 610 meters and a range of over 300 meters. Testing has shown that detection of pedestrians with a thermal imager at this resolution is possible out to 400 meters. In the long-range mode, the detection range calculation indicates that the detection range might extend to 750 meters but practical considerations in optical design lead to limiting the range to 400 meters to match the wide-field mode.

ELIMINATING MODE SWITCHING

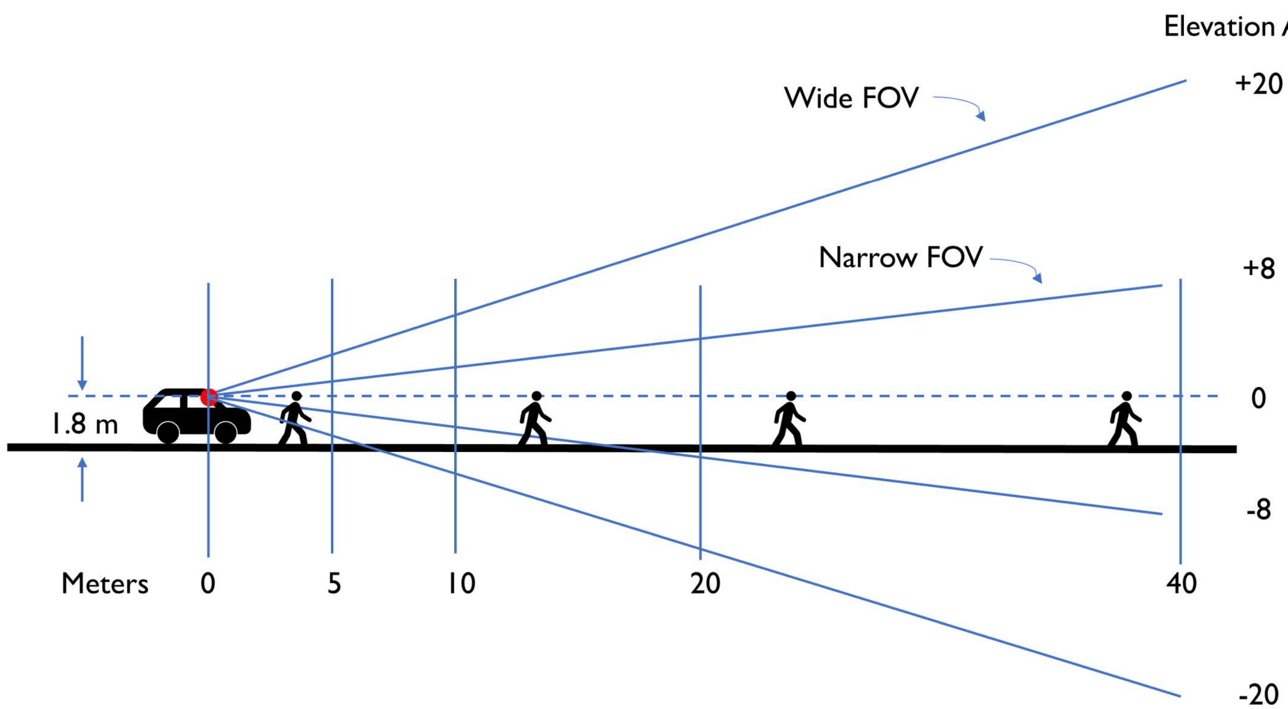
Changing camera field of view while maintaining resolution requires the use of zoom lenses. While these are common, they pose problems in ruggedness, complexity, and cost and they require the CNN to adjust to the changing image size on the fly. All of these ultimately affect the robustness of the detection system and are better avoided. To accomplish this, a different optical strategy can be used.

Segmented optical systems can supply different views of the scene to different parts of the sensor. The most common application of this technique projects a stereo pair of images on a sensor so the stereo image can be captured as a single scene already aligned. A similar technique allows two images to be projected on the sensor with two different magnifications but also already aligned. In the case of automotive cameras, the scene characteristics match better with a top-bottom split of the sensor rather than left-right as in stereo imaging.

The top-bottom split results in two long thin stacked images ready for simultaneous processing by the CNN. An analysis of the vertical coverage requirement results in an imager in which each of these images is 400 pixels high and, of course, 1280 pixels wide.

USEFUL VERTICAL COVERAGE

An image that is more than three times as wide as it is high is reminiscent of wide-screen movies. These images are useful because the field of view of the eye is much wider than it is high and because typical action occurs more horizontally then vertically. This is certainly true in traffic. In fact, as this drawing shows, much of the vertical field even in an image that is 1280 x 400 (3.2:1) seems wasted.



It would be nice if roads were flat, but hills and surface irregularities like dips and potholes assure that they aren't. Further, as a vehicle brakes, typically its front end dips. In these situations, some vertical coverage is needed so the system does not lose the image of the VRU it is trying to avoid. Vertical coverage of usual video systems is not necessary because this would image the sky and the hood of the vehicle, but coverage necessary to avoid image loss is essential.

Above, the scene is mostly sky and below, mostly pavement, but this coverage has a purpose. These drawings show static cars on flat roads. As soon as the car starts moving everything changes. Fifteen degrees of sideways tilt (roll) moves the pavement line to a diagonal. Twenty degrees of front-back tilt (pitch) moves the pavement line to the top or bottom. Driving over a rail crossing is enough to generate both of these deviations. Other causes abound—potholes and pavement tilt from resurfacing, speed bumps and drainage channels, not to mention tilt towards storm sewer grates, especially at corners.

USEFUL VERTICAL COVERAGE continued

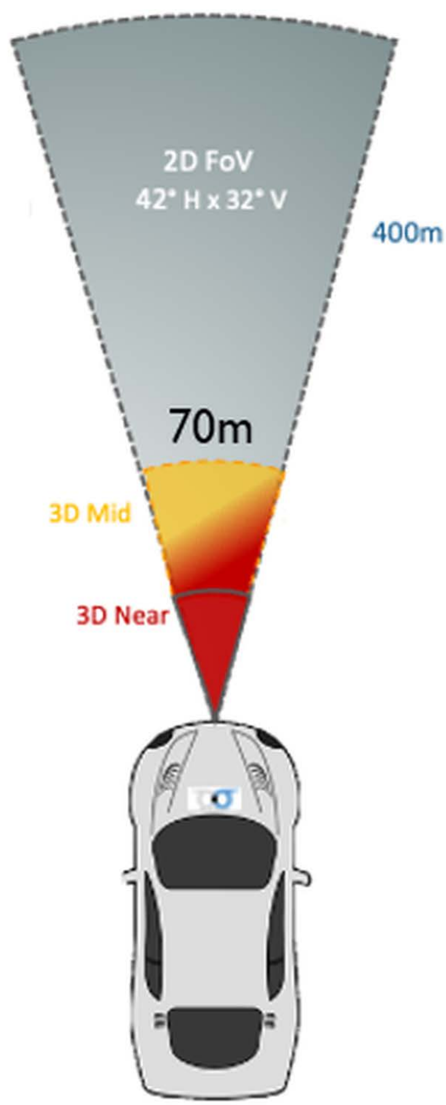
In the narrow field image, large changes in angle are less likely but since the distances monitored are longer, some accommodation must be included for roads that are changing in slope. Twenty degrees of tilt becomes eight degrees but that is well within the typical grade changes on highways.

The goal with a CNN is to try to provide continuous capture of pedestrian images under changing conditions. The correct field of view is critical in maintaining this data flow.

THE CONTINUING ROLE OF OLDER SENSOR CONFIGURATIONS

Analysis of the pedestrian fatality and injury data shows that reducing the collisions in highway settings could produce the largest initial improvement in incidents. This means that while full-function PAEB systems are being developed and deployed, simpler systems using thermal cameras equipped with VGA sensors could bring the reductions in casualties forward.

Here is a useful coverage and range combination that can be achieved using a VGA sensor and a common lens type:



A VGA sensor with a simple lens can provide protection out to 2/3 of the near-peripheral coverage of human vision. This is sufficient to detect most VRUs encountered in highway driving.

THE CONTINUING ROLE OF OLDER SENSOR CONFIGURATIONS continued

The CNN ranging technique works using these sensors, as evidenced by the videos linked at the end of the white paper, so deployment could be rapid. Further, VGA sensors built using the Owl design techniques reduce the cost sufficiently so that these systems would be inexpensive to install, costing even less than current very low resolution handheld thermographic viewers.

With the split optics intended for use with the HD sensor, the VGA sensor could extend its horizontal coverage beyond 90 degrees still with sufficient range to provide additional protection in urban settings.

THE SOLUTION IS THE OWL AI THERMAL RANGER® SYSTEM

Careful analysis of the detailed requirements of imaging for PAEB applications led Owl AI to conclude that a new sensor designed specifically for this purpose was needed. In addition to providing tailored coverage and resolution, this sensor, now nearing release, uses leading-edge technologies to deliver small image size to minimize optics bulk, proprietary on-chip pixel processing to eliminate the need for calibration shuttering and wafer-scale packaging to bring thermal imaging into a cost range that permits use in every vehicle.

REFERENCES

[1] https://www.infinitioptics.com/dri

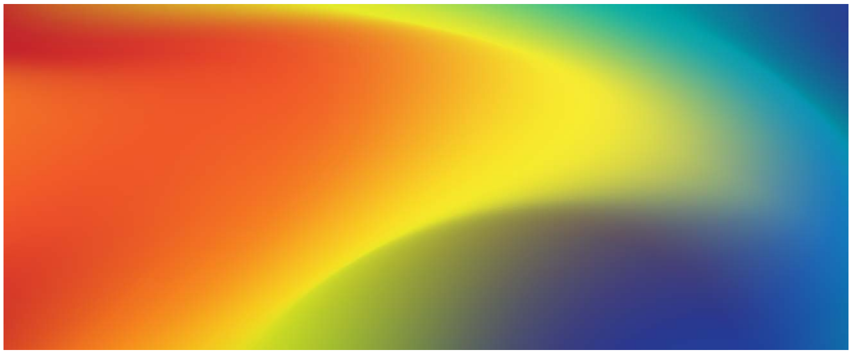
[2] Bhise, Vivek D.,
Ergonomics in the Automotive Design Process. CRC Press, 2011.

[3] https://en.wikipedia.org/wiki/Peripheral_vision#/media/File:Peripheral_vision.svg

[4] https://www.qld.gov.au/transport/safety/road-safety/driving-safely/stopping-distances

INFRARED AT OWL AI

With the goal of locating and identifying pedestrians at night even in weather-degraded conditions, Owl AI studied the spectral band options, evaluated equipment operating in candidate bands, tested the resulting images in an AI recognition system and, based on the results, selected diffraction-limited, high-resolution imaging in the LWIR band as the best way to meet the goal. A new image sensor incorporating the Owl AI research results now becomes the latest embodiment of the OWL AI principle Safety Starts with Perception™.



Safety Starts with Perception

YouTube LINKS

MORE ON THE MYTHS

Thermal Imaging in Winter: https://youtu.be/o2bzGyc6WAg

Hiding from a Thermal Camera: https://youtu.be/redhD3P7xrA

Comparing Images in Various Bands: https://youtu.be/zwZ13NGtlEU

Stupid Thermal Imaging Tricks: https://youtu.be/Fx49t4sv7f0

Getting Started with ROS: https://youtu.be/46TPAKXBOF8

DEMONSTRATIONS OF OWL AI TECHNOLOGIES

November 2022 NHTSA Nighttime Pedestrian Test:
https://youtu.be/wQ5VdMJPOvw

April 2022 Pedestrian Demonstration:
https://youtu.be/JaaTngahlms

January 2022 Pedestrian Comparison with Visible:
https://youtu.be/TmfzYcGRH_Y

November 2021 Pedestrian and Automobile Classification:
https://youtu.be/BMGLgnxNI6M

Example videos of our **THERMAL RANGER** in action can be found on our **YouTube Channel** at this link >>>

“OWL’s THERMAL RANGER® is unique as it delivers rich detail and 3D response day or night.”

Thermal Ranger® is a registered trademark of Owl Autonomous Imaging, Inc.
Safety Starts with Perception™ is a trademark of Owl Autonomous Imaging, Inc.

OWL
AUTONOMOUS IMAGING
The smartest solution for the road ahead

Rochester:
470 Willow Brook Office Park
Building 400, 2nd Floor
Fairport, NY 14450

www.owlai.us