ANALYSIS OF THERMAL CAMERAS FOR SELF-DRIVING CARS

by

Jacob Wilson

A Thesis Submitted to the Faculty of the

DEPARTMENT OF OPTICAL SCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2024

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by Jacob Wilson, titled ANALYSIS OF THERMAL CAMERAS FOR SELF-DRIVING CARS and recommend that it be accepted as fulfilling the dissertation requirement for the Master's Degree.

_____ Date: _____

Dr. Ronald Driggers

_____ Date: _____

Dr. Florian Willomitzer

_____ Date: _____

Dr. Rongguang Liang

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master's requirement.

_____ Date: _____

Dr. Ronald Driggers

Master's Thesis Committee Chair

Department of Optical Sciences

Table of Contents

List of Figures

List of Tables

Abstract

In 2019, a paper called *Leveraging Thermal Imaging for Autonomous Driving* (Miethig et al., 2019) examined the use of thermal cameras for the purpose of aiding self-driving cars. In this article, an overview of different scenarios in which a thermal camera could add to the safe detection of objects was given. Of these listed objects, humans in the road pose arguably the greatest detection risk of all. While *Leveraging Thermal Imaging for Autonomous Driving* gives a great overview of thermal sensing and the use for automated driving, this report will dive deeper into the detection of a defined critical safety target, and how thermal camera performance plays a role in this detection to support Miethig's claim on the benefits of thermal cameras. Parameters such as system Modulation Transfer Function (MTF), Noise Equivalent Temperature Difference (NETD), Signal to Noise Ratio (SNR) will be assessed to validate that a specific thermal imager will work for its intended purpose of human detection.

Introduction

Self-driving cars are a rapidly developing technology that has the potential to revolutionize transportation. One of the key challenges in developing self-driving cars is the ability to perceive the environment safely and reliably. This has already proven difficult for automated driving, as there have been several notable accidents in the industry which involve humans on the road. Miethig et al. lists incidents involving a Tesla striking stationary vehicles, and a retrofitted vehicle operated by Uber striking a pedestrian. Additionally, since the publication of Miethig's paper, other high-profile cases such as a vehicle operated by Cruise dragging a pedestrian have been in headlines. While thermal cameras are certainly not the fix-all solution to these incidents, they can provide critical context to a scene where other sensors may fall short.

Most self-driving cars to date come equipped with a wide array of sensors which help the vehicle perceive the environment in different ways. Each sensor has its own respective role to play in perception but must also fuse with other sensor information to provide full context of the scene. *Figure 1* below is taken from *Leveraging Thermal Imaging for Autonomous Driving* (Miethig et al., 2019) which shows the typical sensors found on an autonomous car and the functions they serve. Different sensors in the table can have the same function as other modalities since better performance is achieved through sensor fusion.

Camera	high resolution, low cost, passive, high sampling rate, light and weather sensitive, 2D imaging RGB measurement
	Lane detection, ground detection, object detection, object tracking, pedestrian detection
Stereo Camera	high resolution, low cost, passive, high sampling rate, light and weather sensitive, 3D imaging RGBD measurement
	Lane detection, ground detection, object detection, object tracking, pedestrian detection
RADAR	low resolution, high cost, active, high sampling rate, light and weather insensitive, 2D scan, velocity and reflectivity measurement, cluster objects
	Ground detection, object detection, object tracking
LiDAR	high resolution, high cost, active, high sampling rate, light insensitive, weather sensitive, 3D scan, velocity measurement
	Lane detection, Ground detection, object detection, object tracking, pedestrian detection
Ultrasonic	low resolution, low cost, active, low sampling rate, short range, light and weather insensitive, low range, 1D scan, distance measurement, cluster objects
	Object detection, object tracking

Figure 1: Different Sensors and Their Benefits (Miethig et al., 2019)

Long wave Infrared cameras can add to these already existing modalities by providing better object and pedestrian detection in adverse conditions such as night-time driving and certain weather conditions which may make sensors such as LiDAR not usable. However, to know how much benefit an LWIR camera can add to the vehicle, it is important to characterize its individual performance to verify that it can aid in the detection of objects and humans.

This paper discusses the use of thermal cameras for self-driving cars and outlines the performance of current available automotive grade sensors to support the claims made in Miethig's 2019 paper. This paper begins by defining a use case for automotive thermal cameras, and then uses this as a baseline metric to compare against actual camera performance. After a use

case has been defined, a commercially available automotive-grade thermal imager will be selected for analysis to determine whether it can meet the needs of a self-driving car. The analysis will be broken down into four separate sections: system specifications (lens, sensor, electronics), radiometry (resolution and sensitivity), performance modeling in Night Vision's Integrated Performance Model (NV-IPM), and image processing techniques. Finally, the paper will conclude if the selected camera meets a self-driving car's needs and reviews the challenges and limitations of using thermal cameras for automotive purposes.

Automotive Use Case

Visible cameras depend on some external light source which reflects off the target and shines on to the sensor. During the night, this light source either needs to be some street lighting or the vehicle's headlights. Thermal cameras, on the other hand, are ideal for self-driving cars during dusk, night and dawn as they can provide clear images of people and objects in complete darkness.

For this paper, the vehicle will operate on a highway or higher-speed road outside of a city. This is because in most cities, the speed limit is much lower which can allow for other modalities such as LiDAR and visible cameras to have more time for detection of humans. This, combined with more visible light due to streetlamps, makes thermal imagers less necessary inside cities. The operating speed will be set to 65 miles per hour, or 29 meters per second.

The most critical detection for an automotive thermal camera will be a small human which is on the road. The human will be defined as 1.52 meters tall and 0.53 meters wide. Other critical features are defined in *Table 1*:

Target Variable	Value
Height	1.52 m
Width	0.53 m
Characteristic Dimension	0.89 m
Target Temperature	37 C
Background Temperature	15 C

Table 1: Target Variables

Given the target and operating speed, a minimum distance for detection can be defined. This will be the shortest distance to the target where the Automated Vehicle (AV) still has enough time to brake and avoid a collision. The worst-case braking scenario for a small car is 132 feet, or ~40 meters at 26 m/s (Consumer Reports 2023). Using simple kinematic equations, the braking deceleration can be found:

$$v_f^2 = v_i^2 + 2 * a * d$$

$$(0\ m/s)^2 = (26\ m/s)^2 + 2 * a * (40\ m)$$

$$a = \frac{-676 \frac{m^2}{s^2}}{80\ m}$$

$$a = -8.45\ \frac{m}{s^2}$$

With this value for vehicle deceleration, the minimum distance to prevent collision at the defined operating speed can be found using the same kinematic equation:

$$(0\ m/s)^2\ =\ (26\ m/s)^2\ +\ 2 * (-8.45\ \frac{m}{s^2}) * d$$

$$d\ =\ \frac{-841\ \frac{m^2}{s^2}}{-8.45\ \frac{m}{s^2}}$$

$$d\ \approx 50\ m$$

Using a worst-case reaction time (the time it takes from the capture of the image to target detection through analysis) of 0.5 s, this puts the minimum distance to prevent collision at 50 m + 14.5 m (reaction distance) = 64.5 m.

Now that a use case has been defined, a thermal imager can be selected to benchmark performance against. This will help to ensure that the thermal imager can meet the specific needs of this application. When selecting a thermal imager, it is important to consider factors such as the resolution, sensitivity, and frame rate. These factors all directly contribute to the camera's ability to detect the defined target and prevent a collision.

Analysis

Currently, there are only a few vendors that manufacture automotive grade cameras. The two main contenders are FLIR and Magna (formerly known as Veoneer). Both vendors offer competitive pricing and performance, however FLIR's Boson 640 automotive camera was selected for this study because it offers several advantages over other cameras on the market. First, it has a high resolution of 640 x 512 pixels, which allows it to capture better resolution images even in low-light conditions. The array format combined with their standard 24-degree field of view lens makes it optimal for human detection. Finally, FLIR has a great track record

with reliability and weather proofing which makes it the desired manufacturer of LWIR camera systems.

Overall, FLIR's Boson 640 automotive camera is a high-quality camera that offers several advantages over other cameras on the market. It is a good choice for businesses that are looking for a reliable and affordable camera for use in automotive applications.

System Specification

The Boson 640 has standard specifications for consumer Long-Wave Infrared (LWIR) cameras such as 12 µm pixel pitch, F/1 lens, with an uncooled microbolometer sensor. However, what sets FLIR apart is their consistent performance and manufacturing which other vendors end up comparing their product to. *Table 2* below outlines the useful specifications of the Boson 640 which will be used later for analysis.

Specification	Value
Format	640 x 512
Pixel Pitch	12 µm
FOV	24°
F/#	1
Focal Length	21.5 mm (Estimated)
Bandwidth	8 µm - 14 µm
Transmission	60% (Estimated)
Quantum Efficiency	0.7 (Estimated)
NETD	50 mK

Frame Rate	30 Hz

Table 2: FLIR Boson 640 Parameters

With the necessary camera parameters defined above, it is now possible to assess the radiometry of this AV problem and determine whether this camera system is suitable for self-driving cars. The camera must be able to capture enough signal in low-light conditions and withstand certain atmospheric attenuations to detect and identify humans on the road. To test this, the FLIR Boson 640 camera will be evaluated based on several metrics such as resolution, sensitivity, SNR, MTF, and NETD.

Radiometry

Radiometry is the measurement of radiant energy and its interaction with the systems we design. It is a branch of optics that deals with the transfer of radiant energy between objects. This field of optics is important because it provides an understanding of how radiant energy interacts with matter (target and camera system). This understanding is essential for measuring how well a given system can perform under a variety of conditions. Radiometry, for this paper, will be broken down into two main sections – resolution and sensitivity.

Resolution

Resolution is a measure of how well a sensor can see small spatial details on an object (Driggers et al., 2012). The first step in assessing resolution is to determine how many pixels are on the target. If the target's area is larger than the area of one pixel projected at the target distance, then the target is considered resolved. If the camera system passes the test for

resolvability, then system MTF can be looked at to determine how well this system can distinguish the defined target.

For the purposes of a self-driving car, a resolved target (i.e. more than one pixel on target) is not enough of a qualifier when looking at a camera system. An explanation as to why more than one or two pixels is needed will be given in a later section, but for now, the number of pixels will be calculated. To start, the individual pixels need to be projected out towards the target location. This can be done in the following equation:

$$A_{projected} = A_{Pixel} * \frac{R^2}{f^2}$$

where R is the range to the target and f is the focal length of the camera. Plugging in the known values results in the following:

$$A_{projected} = (0.012\ mm)^2 * \frac{(64{,}500\ mm)^2}{(21.5\ mm)^2}$$

$$A_{projected} = 1{,}296\ mm^2$$

This represents the area of one pixel at the location of the target. Next, the area of the target will be calculated and the number of pixels on said target can be found:

$$A_{Target} = (1{,}520\ mm)(530\ mm) = 805{,}600\ mm^2$$

$$n_{pixels} = \frac{805{,}600\ mm^2}{1{,}296\ mm^2} = 621\ pixels$$

For this camera system, at the desired target location there are 621 pixels on the target. This is more than enough to be considered resolved, and give enough buffer to assess recognition and identification, which will be discussed in a later section. Next, the system MTF will be calculated to determine how well it can resolve the target.

Many different attributes in a camera system will all contribute to system MTF, which is a measure of how well a camera can resolve certain spatial frequencies. For this paper, three main contributing factors will be assessed: optics, detector, and atmosphere. There are, of course, other contributors, such as electronics, display, human eye (still required for labeling and training data), etc. Optics MTF, for this analysis, will assume diffraction-limited optics, which is the fundamental maximum MTF an optic can have. Detector MTF is a function of pixel shape and size. Assuming a square pixel shape, smaller pixel sizes will result in better MTF over spatial frequencies. Finally, atmosphere MTF is a measure of attenuation as light travels through it. It is a function of range, altitude, and atmospheric turbulence (referred to as Cn^2). For this paper, a low Cn^2 value – which is indicative of a turbulent condition – will be used as a worst-case scenario. The following chart (generated from an MTF calculator given in OPTI 613) shows how each individual MTF, as well as total overall MTF, performs over spatial frequencies.

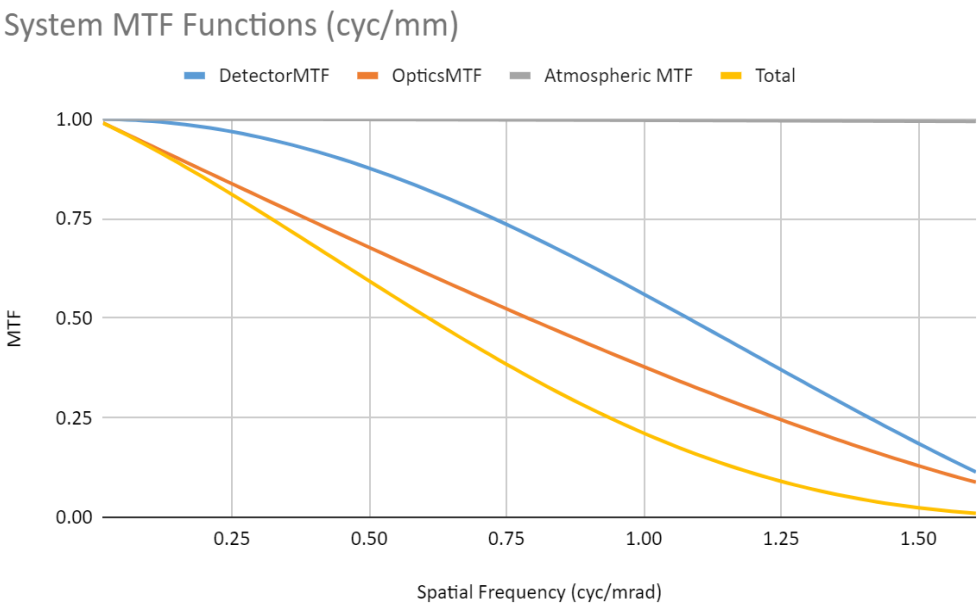

Figure 2: System MTF Functions

The component with the lowest MTF is considered the limiting MTF of the system. In the case for this camera, the limiting MTF is shown to be from the optics of the camera. Since the optics were modeled to be diffraction-limited, the only ways to improve MTF is to either change wavelength or increase the aperture size. For this paper, these variables will be considered fixed, as changing them for mass production will be costly. Additionally, since the limiting component is the optics, any aberrations or alignment issues with the lens will worsen MTF and have a direct impact on system MTF. In a later section of this paper, a Boson 640 camera will be used to measure MTF and gauge how close to diffraction-limited the optics really are.

The final check for resolution is to determine if the system is sampling limited. A system is undersampled if:

$$\frac{f_{\#} * \lambda}{D_{pitch}} < 2$$

If the wavelength is taken to be 11 μm (The average of the band limits), this equation comes out to be:

$$\frac{1 * 11\ \mu m}{12\ \mu m} = 0.92$$

This shows that the system is undersampled and will need to move in space for supersampling. For LWIR cameras on a self-driving car, this will not be a problem as the system is only needed when the car is in motion.

Sensitivity

Sensitivity encompasses a wide variety of metrics such as NETD, SNR, and Minimum Resolvable Temperature (MRT), which is sensitivity as a function of resolution. Overall, sensitivity is a measure of how well a camera system can measure changes in radiant energy. NETD is already given in the specifications table above as 50 mK, so there is no need to calculate this value. However, in a later section, NETD will be measured on a sample camera to see how close to 50 mK a real sample gets.

To show how sensitivity evolves over distance, power on the detector and corresponding SNR will be plotted below. SNR over distance will be calculated using NV-IPM's SNR probe, which factors in system NETD (in this case 50 mK) and source contrast. As for power on the detector, an equation can be put together to get power as a function of range. This can be done by starting at the camera level and then projecting out towards the target. To start, it will be assumed that all of the power that falls on the aperture will make it to the detector, after accounting for lens transmission.

$$P_{Det} = P_{Apt} * R_{\Omega}$$

However, power on the aperture can be described as the source intensity multiplied by the solid angle subtended by the aperture. Target intensity is a contrast intensity, which is the difference between the source intensity and background intensity at that same distance. Additionally, atmospheric transmission needs to be accounted for which will be found using Beer's Law.

$$P_{Det} = I_{contrast} * \Omega * \tau^{R} * R_{\Omega}$$

While the target is in the resolved domain of this camera system, the target contrast intensity needs to be limited to the area of one pixel. To do this, the target radiance can be used in conjunction with the projected area of one pixel.

$$P_{Det} = L_{contrast} * A_{Projected} * \frac{\pi {D_{Apt}}^2}{4R^2} * \tau^{R} * R_{\Omega}$$

$$P_{Det} = L_{Contrast} * A_{det} * \frac{R^2}{f^2} * \frac{\pi {D_{Apt}}^2}{4R^2} * \tau^{R} * R_{\Omega}$$

$$P_{Det} = \frac{I_{Contrast}}{A_{Target}} * A_{det} * \frac{R^2}{f^2} * \frac{\pi {D_{Apt}}^2}{4R^2} * \tau^{R} * R_{\Omega}$$

If background radiance doesn't change with distance, the above equation represents power on the detector over range. Simplifying this gives the following:

$$P_{Det\ resolved} = \frac{I_{Contrast}}{A_{Target}} * A_{det} * \frac{\pi {D_{Apt}}^2}{4f^2} * \tau^{R} * R_{\Omega}$$

It's important to note that this equation is only valid out until 1.608 km (where $A_{tgt} > A_{Projected\ Pixel}$) because the target is resolved. After the resolved-unresolved crossover, the target contrast radiance is now captured in one pixel and incurs a $1/R^2$ dependence:

$$P_{Det\ unresolved} = I_{contrast} * \Omega * \tau^{R} * R_{\Omega}$$

$$P_{Det\ unresolved} = I_{contrast} * \frac{\pi {D_{Apt}}^{2}}{4R^{2}} * \tau^{R} * R_{\Omega}$$

Using the two equations found above for resolved and unresolved power on the detector, and the calculated values for SNR using NV-IPM, the following plot shows how the sensitivity of the camera system changes over range.

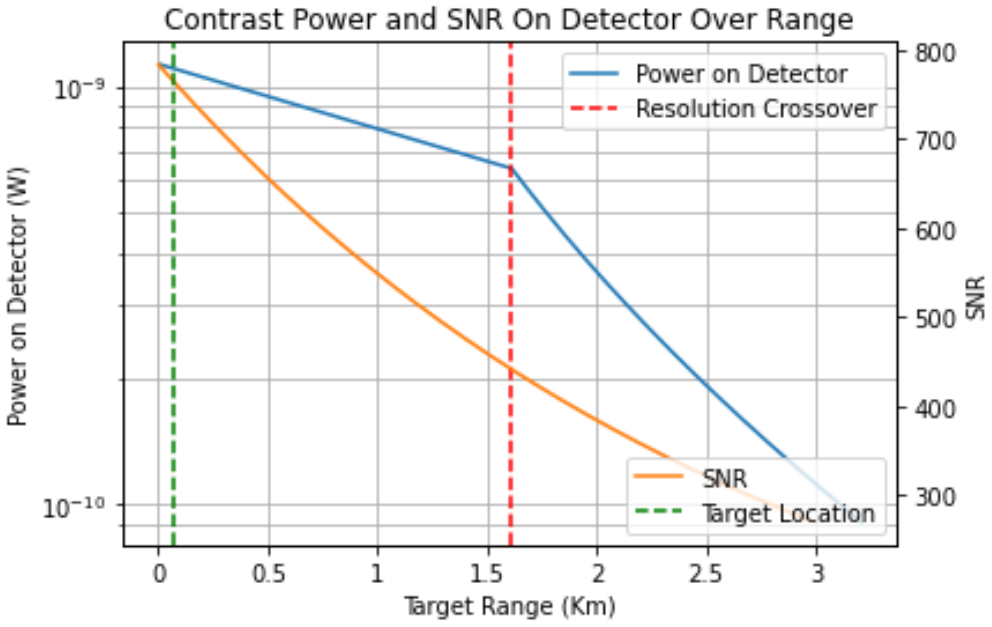

Figure 3: Power and SNR of Detector Over Target Range

The red dotted line on the chart above represents the distance at which the target changes from resolved to unresolved – the Resolution Crossover. For AV purposes, this camera is no longer useful past the resolution crossover as the target is now smaller than one pixel – which can no longer be properly detected. In terms of performance, the Boson 640 appears to have plenty of SNR for this target at ranges far past the critical safety distance (indicated by the green dotted line).

Even with resolution and sensitivity defined for the ideal camera, it's hard to tell just how well this camera performs at the critical task of detecting humans. This is where performance modeling takes over. Data from the radiometric calculations done above can be fed into NV's Integrated Performance Model to show the probability of task performance evolving over distance.

Performance Modeling

NV-IPM is a comprehensive software package that is specifically designed to model a system's ability to complete a task. For this paper, the tasks of detection, recognition, and identification will be modeled. Each level of performance requires a minimum number of pixels on a target to complete the task. For detection, one to two pixels can be used to judge whether some object is in the scene. Recognition requires three to four pixels on the target to determine if the object is human or something else. Finally, identification needs eight to nine pixels on target to find aspects of the human (Driggers et al., 2012). Identification is not necessary for AVs, as the main concern is finding humans in the road and not necessarily what the human is doing. However, it will be kept in the model to show if the camera can go above and beyond the basic need.

To start, a basic model for an uncooled microbolometer (found in NVIPM library) was used as a base. From there, data from Table 1 and Table 2 was input into the model to get statistics on task performance in the defined use case. The Target Task Performance (TTP) model was selected and iterated over a distance out to 1.6 km since this is the distance of the

resolution crossover. The following chart shows how an ideal camera would perform in the three tasks of detection, recognition, and identification.

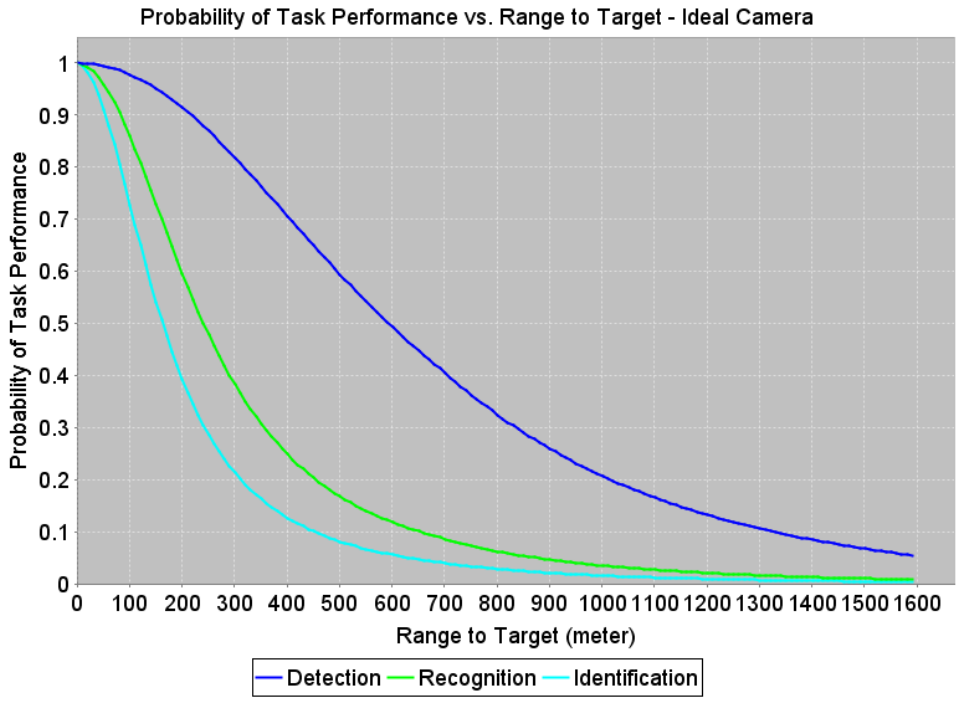

Figure 4: NV-IPM TTP Metric over Distance

The ideal model performs well at the critical safety distance of 64.5 m, showing > 90% probability of task performance for detection and recognition. To make task performance a little more rigorous, and allow for safer modeling, the number of pixels on target was set to 7.5 for recognition (up from 4) and 13 for identification (up from 9). *Table 3* shows the individual task performance probabilities for the three levels of discrimination at the critical safety distance.

Task	Probability of Success
Detection	98%
Recognition	91%
Identification	82%

Table 3: Task Performance for Ideal Camera

A 91% success rate is excellent, especially when combined with other modalities such as long-range LiDAR and RADAR. However, this is just for an ideal camera, whereas most real-world cameras will come with some deficiencies and degradations which will hinder task performance.

Lab Test Results

As a useful exercise, values such as MTF and NETD will be tested in a lab to verify how close a real Boson 640 camera gets to the designed values. If these metrics are, for whatever reason, worse than the ideal case defined in the specification sheet, then these lab-measured values will be fed into the NV-IPM model to compare how metrics such as recognition and identification change.

Both MTF and NETD were measured using a combination of an 8”x8” extended blackbody source and variable target collimator provided by CI Systems. The sample FLIR Boson 640 camera was placed on two Zaber rotation stages to allow for optical alignment and changing field positions as shown in *Figure 5*.

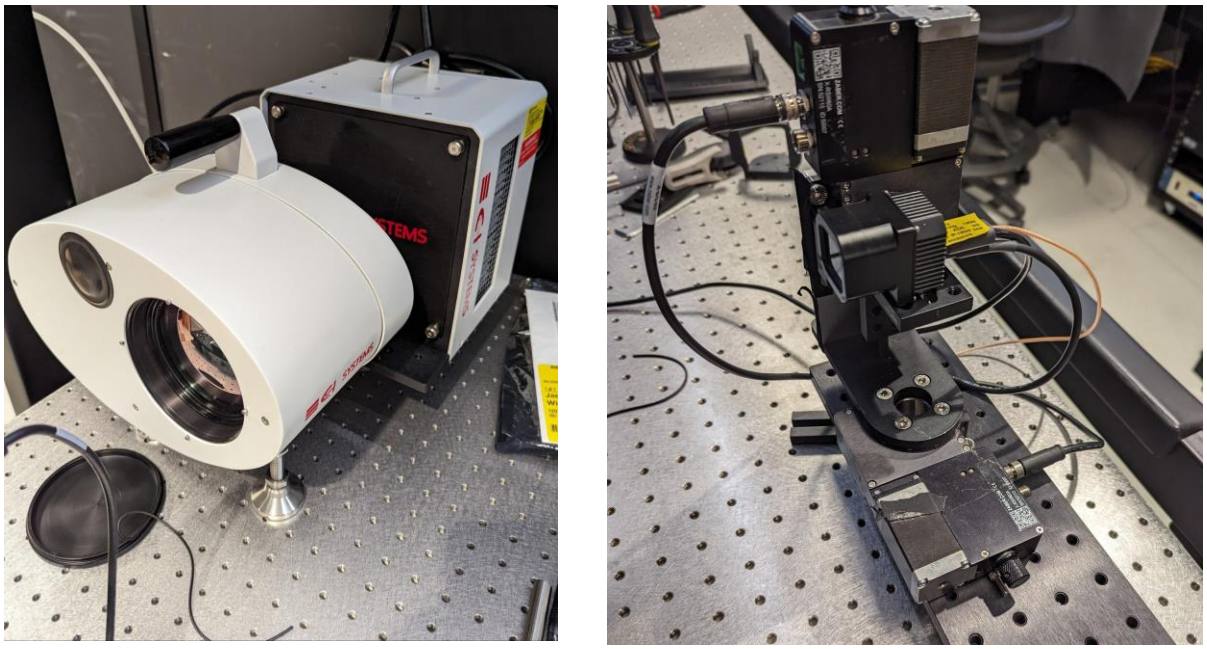

Figure 5: CI Systems Collimator and Blackbody (Left), FLIR Boson 640 Mounting (Right)

For measuring MTF, two separate half-moon targets were used to measure MTF on the horizontal and vertical planes. These half-moons were tilted ~5 degrees off from the horizontal and vertical axes of the sensor to allow supersampling of the reticle edge. Using the collimator's relay lens control, the reticle was projected out to 64.5 m to measure MTF at the critical safety distance. *Figure 6* shows what the reticle looks like through the Boson camera.

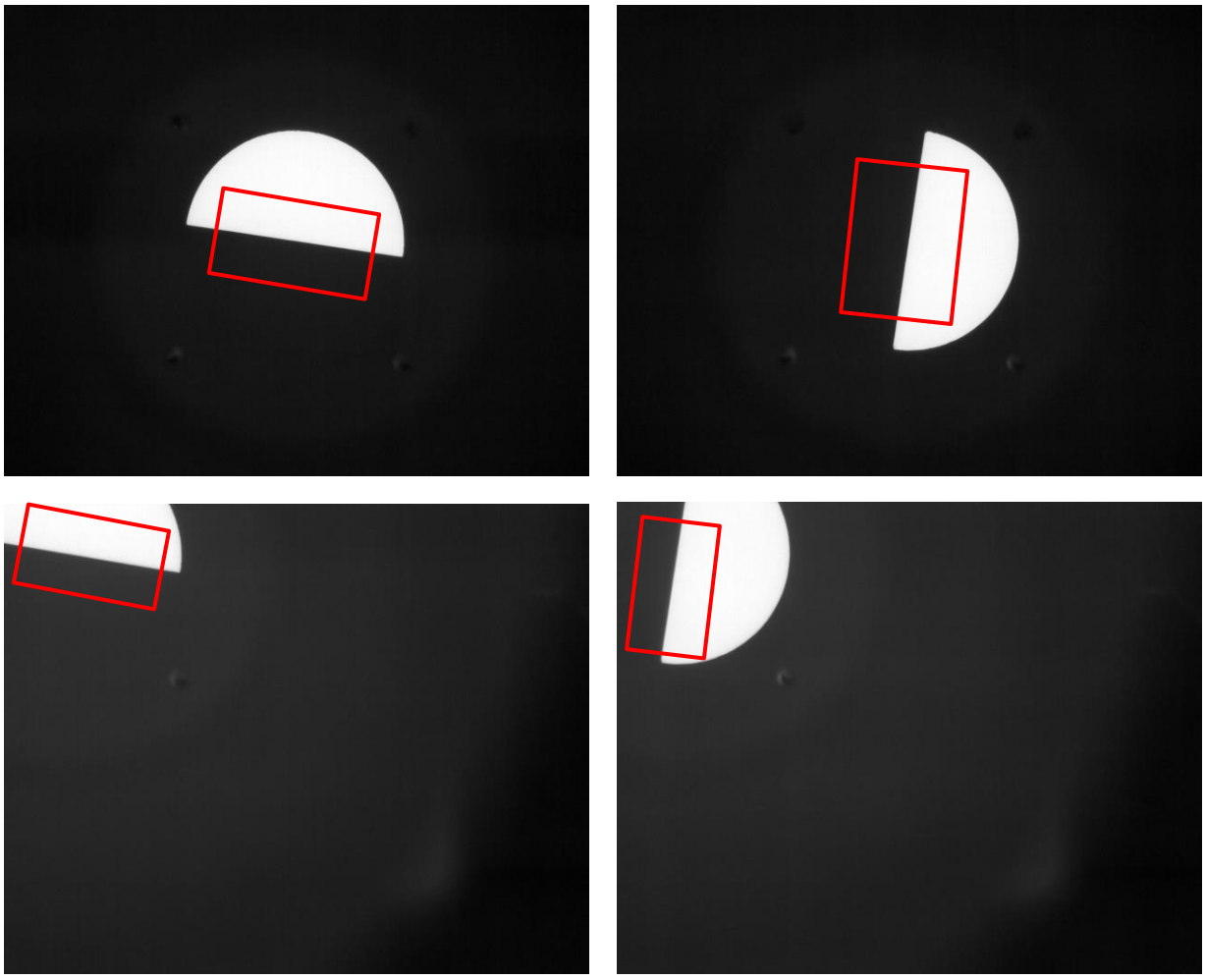

Figure 6: 0F Vertical (top left), 0F Horizontal (top right), 0.8F Vertical (bottom left), and 0.8F Horizontal (bottom right) targets projected at 64.5 meters.

Edge spread functions and subsequent MTF curves were generated using a standard analysis suite called Imatest. Imatest's SFR library was used on each grayscale image to get the most accurate MTF measurement from each image. The red boxes superimposed on each target in *Figure 6* depicts the ROI used in Imatests's MTF calculation. The following graph shows how lab measured MTF compares to the modeled system MTF above.

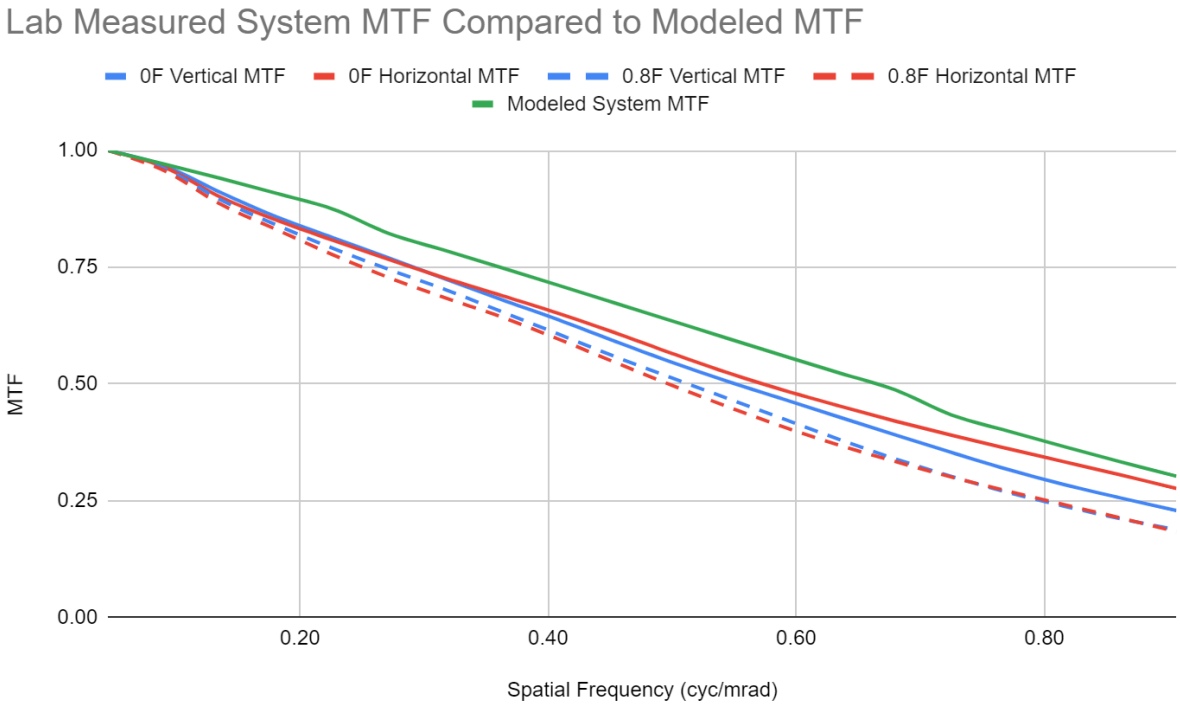

Figure 7: Lab Measured MTF On-Axis and at 0.8F

Figure 7 clearly shows that the sample camera deviates from the ideal camera – which is to be expected. This could either be from the optics not actually being diffraction-limited (either from manufacturing errors or material quality) or due to the system being sampling limited. While this paper primarily delves into finite target locations, it's important to note that a lot of cameras often prioritize focusing at infinity. However, this preference for infinite focus can lead to lower MTF values when capturing images at finite distances. Despite the factors contributing to the reduced MTF, now that resolution has been measured, the sensitivity parameter of NETD will be assessed.

NETD is also measured on the same CI systems setup, except the target used is a square through hole where the temperature of the source (shining through the square) is measured and

the temperature of the aperture blocking the rest of the source is measured. This allows for a differential temperature, between the source and collimator aperture, to be measured. Since the camera reports data in raw ADC counts, a signal transfer function is needed to convert from counts to mK. To do this, differential temperatures between the source and collimator aperture were found at different blackbody temperatures. At each source temperature, frames were averaged together to find mean ADC counts in both ROIs (source and aperture ROI). The slope of this response curve is known as the Signal Transfer Function (SiTF). *Figure 8* shows measured mean ADC counts vs. blackbody temperature.

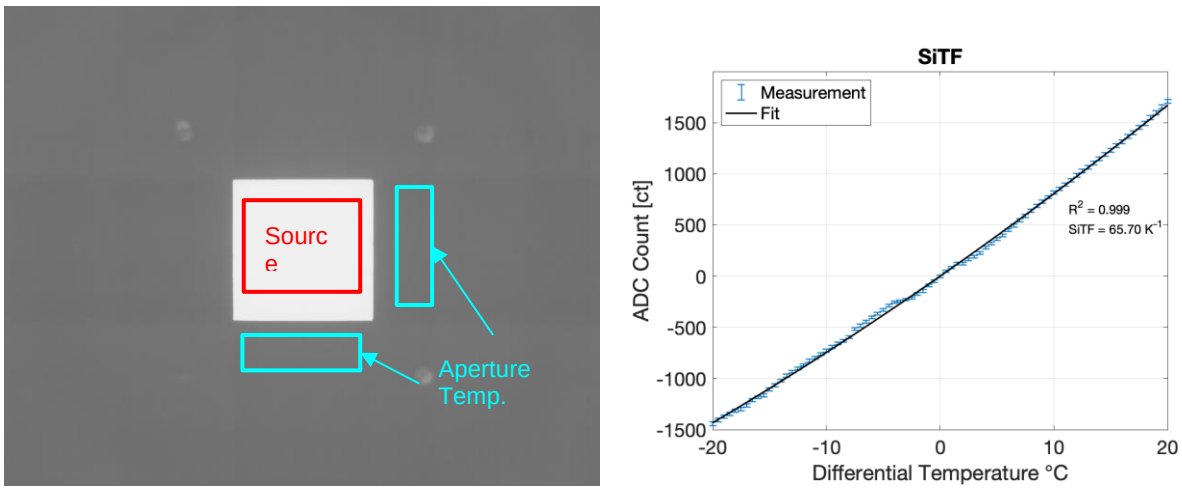

Figure 8: NETD Target (left) and corresponding SiTF graph (right)

With a signal transfer function measured, the random spatiotemporal noise (σ_{tvh}) can be found. This is done by taking consecutive flood filled images of a uniform blackbody source and analyzing the 3D noise present in this dataset. For this paper, an image analysis software called ImageJ was used which has a 3D noise analysis routine. Additionally, if the SiTF slope is

known, ImageJ will output σ_{tvh} in mK. Surprisingly, the resulting σ_{tvh} measured in the lab came out to 42 mK, which is close to the higher tier of cameras offered by FLIR.

The measured values for MTF and NETD outlined above can now be fed back into the system performance model to see whether the deviations found in a real camera affect the probability of performing a task. *Figure 9* outlines the original ideal model and superimposed the new performance model with updated MTF and NETD.

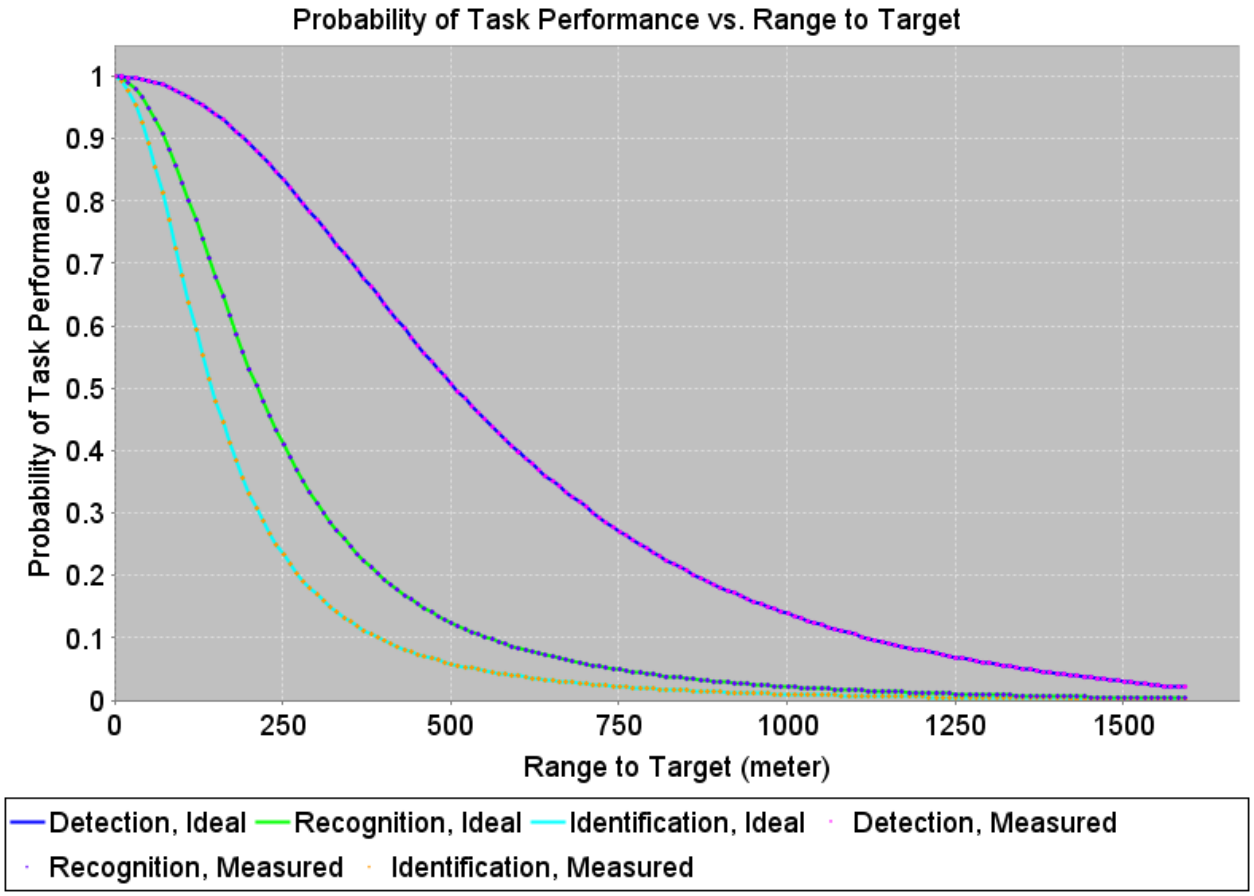

Figure 9: Comparison of Ideal Model and Lab Measured Model

The two models have almost identical results and perfectly overlap in *Figure 9*. It's possible that the slight degradation in MTF has little to no impact on the model, or the improved

NETD value canceled out any impacts from a lower MTF. Table 4 shows how the levels of discrimination for both models perform at 64.5 m.

Task	Ideal Model Probability of Success	Lab Measured Model Probability of Success
Detection	98.69%	98.69%
Recognition	91.26%	91.25%
Identification	82.08%	82.07%

Table 4: Task Performance for Ideal Camera and for Lab Measured Model

Even though measuring resolution and sensitivity in the lab appears to not have an impact compared to the ideal model in this case, it is still an important exercise to always verify parameters in the lab. This ensures that the camera is indeed performing as intended and no other degradations need to be factored into the model.

Now that important camera performance metrics have been defined and analyzed, there is one last step to produce an image for the vehicle. This final step is image processing which takes a raw image (produced given the metrics outlined above) and processes it in some way to better extract the desired features in that image.

Image Processing

There are many ways to process a raw image which can extract or accentuate different desired features. These different methods can also range in complexity from simple spatial filtering to more advanced machine learning algorithms. For the purposes of this report, a bilateral filter will be used to show how features in an image can either be corrected or amplified to further increase target contrast.

One of the main benefits of bilateral filters is that they can smooth out noise and other non-uniformity while preserving edges—it is a filter in both space and intensity. This is an important distinction from the typical gaussian filter, because an automated vehicle's perception model will rely on sharp contrast of objects with respect to the background radiation. If a spatial gaussian filter was used with targets in the image, it would smooth out target edges and make it hard to detect where a target boundary begins and ends.

To define a bilateral filter, it is important to first start with the typical gaussian filter. Just like the name suggests, this filter is used to convolve a gaussian profile centered on some pixel within a defined pixel neighborhood. A Gaussian filter takes on the form shown in the equation below.

$$f(p_{ij} - k_{ij}) = \frac{1}{2\pi\sigma^2}\, e^{\frac{-[(p_i - k_i)^2 + (p_j - k_j)^2]}{\sigma^2}}$$

Where k_i and k_j represent the i and j row/column index of the center pixel, and similarly p_i and p_j represent the row/column index of the pixel under operation. The parameter σ defines the width of the gaussian profile and is important to fine tune for your specific purpose. A large σ will apply a higher weighted average to pixels farther away from the center pixel. This will create a more spatially averaged image, but critical features in the image may be blended with the background or lost all together. Conversely, a σ which is too small will leave undesired noise in the image because pixels further away from the center pixel will be unaffected.

Next, the bilateral filter comes into play by creating a gaussian profile in the intensity domain of a pixel neighborhood and multiplying that by the spatial domain filter, this relation is shown below (Durand et al., 2002):

$$J = f(p_{ij} - k_{ij})g(I_{ij} - K_{ij})$$

The intensity gaussian filter looks the same as the spatial filter, but instead of applying a weighted average to a pixel based on the distance away in space, it applies a weighted average based on the difference in intensity (in reference to the center pixel). Additionally, the σ now changes the width of the gaussian profile in intensity, which means a higher σ will "smooth" intensities which are further away from the center pixel's intensity value. A σ parameter which is too high will not preserve sharp changes in contrast, resulting in loss of boundaries along targets. The overall function of a bilateral filter can be seen in Figure 10, where a spatial kernel is cut off along the boundary of intensity values which preserves the sharp edge but smooths out noise.

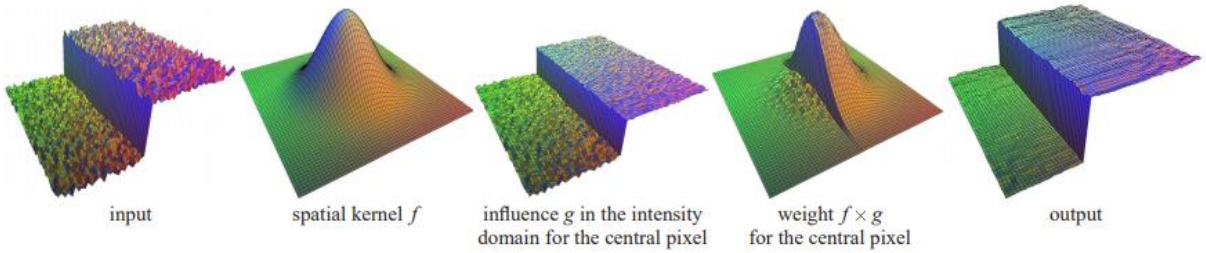

Figure 10: Bilateral Filtering in Space and Intensity (Durand etl. al, 2002)

This filter can be used effectively with thermal images to smooth out issues like residual fixed pattern noise or shot noise while possibly sharpening contrast along the boundaries of a target. To show this effect, Figure 11 contains two photos of the same scene, visible and thermal.

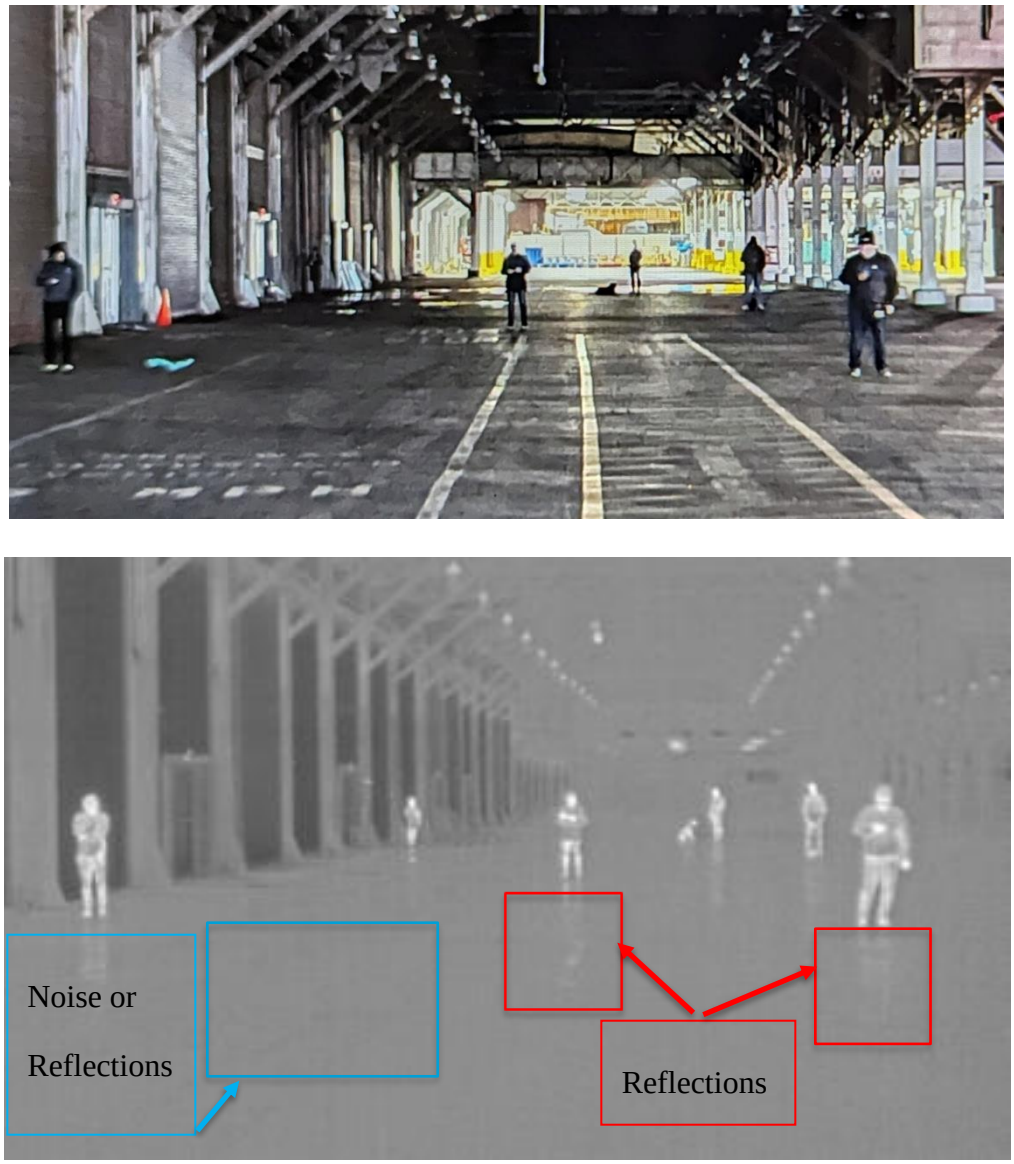

Figure 11: (Top) Visible Spectrum Image of Targets in a Scene, (Bottom) Thermal Image of Targets in a Scene

The thermal image shown in Figure 11 already has some processing such as non-uniformity correction done to it as this is performed before the image is saved out. However, the red boxes outline reflections from the targets off the smooth concrete. Additionally, the blue box highlights possible fixed pattern noise or additional reflections which are hard to make out. These features in this raw image are undesirable to a vehicle's perception system, as it may mistakenly track an object that is not there, which can create safety concerns.

A bilateral filter was then applied to this raw thermal image with specific spatial and intensity gaussian widths. Figure 12 shows the results of this process with the bottom image containing little to no reflections, and less overall variation in the background.

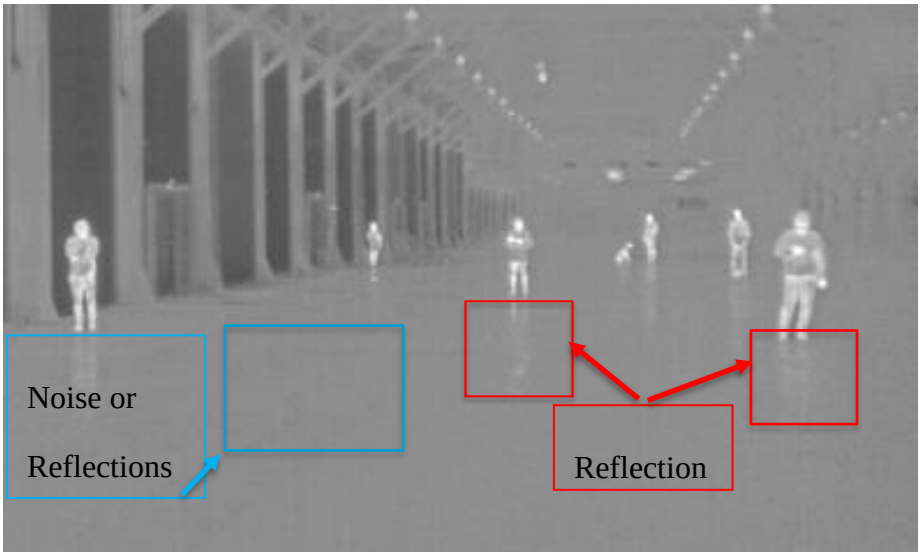

Figure 12: (Top) RAW Thermal Image of Targets in a Scene, (Bottom) Processed Thermal Image of Targets in a Scene

With sharper target contrast and less overall background variance, this processed image can boost a perception model's probability of detection. It's hard to say by how much exactly without running an entire training set and analyzing the results, but it is reasonable to expect a couple percentage points increase in probability of task completion.

Having established the essential metrics defining a thermal camera and outlining their impact on system performance, this paper can proceed to delve into a detailed discussion on the actual benefits of integrating a thermal camera into a self-driving car. This exploration provides insight into how such technology can enhance the capabilities and safety of autonomous driving systems as described in *Leveraging Thermal Imaging for Autonomous Driving*.

Discussion

This section of the paper revolves around the discussion of results obtained from the radiometric calculations and performance models previously presented, which help inform the investigation into automotive Long-Wave Infrared (LWIR) cameras for self-driving cars as outlined in *Leveraging Thermal Imaging for Autonomous Driving*. One of the key challenges in developing self-driving cars is the ability to perceive the environment safely and reliably.

On top of the automated vehicle collisions mentioned in Miethig's paper, In 2021 alone, the US had nearly 43,000 reported fatal car accidents. Of these 43,000 incidents, 7,388 were fatal to pedestrians (IIHS, 2023). This is the motivation behind the automotive use case defined above and mentioned in Miethig's paper. Night-time driving presents a large risk to vulnerable road users, such as pedestrians, because the main source of information (visible light) is missing. The reliance on other modalities such as LiDAR, RADAR, and LWIR sensors becomes necessary as they either come equipped with their own light source (LiDAR and RADAR) or rely on targets in the scene to be the light source (LWIR). Naturally, each modality has its own problems that need to be designed around.

Thermal cameras are a promising technology for the task of night-time detection, and as the data in the previous section suggests, they are a good candidate for Vision Zero safety (*Vision Zero SF*. 2023) in environments with little to no visible light. Specifically, FLIR's Boson 640 camera appears to be well-qualified to meet the demands of automotive safety. This may come as no surprise to those familiar with FLIR as this company has set the standard for LWIR

cameras over time, so much so that other commercial vendors will package Boson sensors into their own camera system.

As stated previously, two main companies produce LWIR cameras for automotive purposes – FLIR and Magna. This may be because FLIR and Magna have already cornered the market, but it may also be due to the complexity of manufacturing an automotive-hardened camera that doesn't inflate the cost of vehicles. Uncooled microbolometers have been around for decades but have previously been limited to military purposes which usually have much higher budgets for cameras. Whatever the reason for limited suppliers may be, since FLIR is an industry leader, and has readily available information on their cameras, the Boson 640 camera was selected.

As far as basic resolution and sensitivity for this camera system go, FLIR has designed a reliable product that closely matches its outlined specifications. With a 640x512 array, 12 μm pixel pitch, 24-degree lens option, and 30 Hz frame rate, the Boson 640 has more than enough resolution performance to detect vulnerable road users at, and far past, the critical safety distance of 64.5 m. To match this, the Boson sensor also has a good sensitivity of 50 mK NETD and meets the required contrast SNR over the desired distances.

What makes an even better justification for this camera is how close the camera performs against its design parameters, sometimes exceeding them. While there was a slight degradation in MTF, which may be due to the sample being focused on a different object distance, it was only a slight deviation, even off-axis. The surprising result of an improved NETD is a testament to the quality sensors that FLIR produces. The 50 mK specification is, in fact, a maximum value, so the camera only has to have a max 50 mK NETD. This sample camera measures only 2 mK

away from the higher price level of 40 mK NETD cameras. Typically, vendors will break out their cameras into different levels of NETD, since a 10 mK improvement in NETD exponentiates the difficulty of manufacturing the sensor and severely decreases the lot yield. If a large sample of cameras sold to an AV company had a maximum NETD of 50 mK, but mostly measured close to 40 mK, then the company would be getting a great deal on higher quality sensors.

Due to the successful lab verification of this camera's resolution and sensitivity, the resulting performance model closely resembles the ideal case as shown in *Figure 9*. Any potential deviations from a manufactured camera have minimal impact on the model, thereby reinforcing its reliability and robustness. Since finding vulnerable road users in the scene is the most critical function for an LWIR sensor on a self-driving car, a very high probability of task completion is required for it to be a viable solution for low light operation. Particularly, recognition is the most important task, as the car needs to be able to tell that the target is a human (above the threshold of detection) or another creature. Both the ideal model and lab-measured model predict a 91% probability of success in recognition. By itself, this is already a promising threshold, but the vehicle needs to come as close to 100% as possible to avoid any critical safety event. This is where image processing can help boost probability of detection.

The previous section also outlines how a simple bilateral image processing technique can suppress undesired features and bring out human targets in the scene. With a possible increase in target contrast, this could potentially boost a camera’s probability of task completion. This, combined with sensor fusion of modalities such as RADAR or LiDAR can help close any possible gap in human detection and recognition.

While there are a lot of promising aspects of this camera, it is also important to note the challenges that these sensors will face. Because these sensors operate in the 8-14 μm range, their performance can be severely degraded in certain types of weather such as dense fog and humid rain. This is because water is an absorber in this bandwidth, which can drastically cut down on contrast SNR of a target. While *Figure 3* shows that there might be some room for a decrease in SNR at 64.5 m, Figure 13 shows a possibly missed detection at what appears to be a relatively close range.

Figure 13: Missed Detection of Thermal Image due to Weather (Miethig et. Al, 2019)

Since these are fundamental limitations of the system (i.e. cannot be designed out of the camera), LWIR cameras should not be the only solution for night-time driving. Since every camera will have their fundamental limits, an emphasis on sensor fusion – the combination of multiple systems – should be used to counterbalance each unit's limitations.

Conclusion

As self-driving cars start to become more of a reality and accessible to the public, more emphasis needs to be placed on how they can create a safer environment for all road users. In particular, the risk of operating at night needs to be mitigated for any level of vehicle autonomy. This is the point that *Leveraging Thermal Imaging for Autonomous Driving* makes, and this paper attempts to back up. LWIR cameras can really show their usefulness to the full stack of sensors in an autonomous vehicle. The analysis section of this paper clearly shows that these cameras have more than enough performance to help avoid critical safety events with vulnerable pedestrians. As such, any automotive company dealing in self-driving should include these sensors – at least until other modalities can provide the coverage that LWIR does.

References

1. B. Miethig, A. Liu, S. Habibi and M. v. Mohrenschildt, "Leveraging Thermal Imaging for Autonomous Driving," 2019 IEEE Transportation Electrification Conference and Expo (ITEC), Detroit, MI, USA, 2019, pp. 1-5, doi: 10.1109/ITEC.2019.8790493.

2. Driggers, Ronald G., et al. Introduction to Infrared and Electro-Optical Systems, Second Edition. Artech House, 2012.
3. Palmer, J. M. & Grant, B. G. *The art of radiometry*. (SPIE Press, 2010).
4. Grant, Barbara G. *Field Guide to Radiometry*. SPIE, 2011.
5. Bartlett, Jeff S. “Cars, Suvs, and Trucks with the Best and Worst Braking Distances.” Consumer Reports, www.consumerreports.org/car-safety/best-and-worst-braking-distances-a2960086475/. Accessed 14 March 2024.
6. Durand, Fredo, Dorsey, Julie, “Fast bilateral filtering for the display of high-dynamic-range images”, Proceedings of the 29th annual conference on Computer graphics and interactive techniques, July 2002, Pages 257–266, https://doi.org/10.1145/566570.566574
7. “Fatality Facts 2021: State by State.” *IIHS*, May 2023, www.iihs.org/topics/fatality-statistics/detail/state-by-state#deaths-by-road-user.
8. *Vision Zero SF*, https://www.visionzerosf.org/. Accessed 17 March 2024.