Performance of a Time-of-Flight Range Camera for Intelligent Vehicle Safety Applications

S. Hsu, S. Acharya, A. Rafii and R. New, Canesta, Inc.

Abstract

A variety of safety-enhancing automobile features can be enabled by microsystems that can sense and analyze the dynamic 3D environment inside and outside the vehicle. It is desirable to directly sense the 3D shape of the scene, since the appearance of objects in a 2D image is confounded by illumination conditions, surface materials, and object orientation. To overcome the disadvantages of 3D sensing methods such as stereovision, radar, ultrasound, or scanning LADAR, we present Electronic Perception Technology, an advanced range camera module based on measuring the time delay of modulated infrared light from an active emitter, using a single detector chip fabricated on standard CMOS process. This paper overviews several safety applications and their sensor performance requirements, describes the principles of operation of the range camera, and characterizes its performance as configured for airbag deployment occupant sensing and backup obstacle warning applications.

1 Introduction

A variety of safety-enhancing automobile features can be enabled by microsystems that can sense and analyze the dynamic 3D environment inside and outside the vehicle. Safety features may include collision warning and avoidance, smart airbag deployment, obstacle detection such as backup warning, and parking assistance. Common to these applications is the need to detect, isolate, measure, locate, recognize, and track objects such as people, traffic, and roadside features.

It is often proposed to perform these tasks using conventional 2D imaging sensors and analysis software, but achieving cost-effective and reliable performance during all vehicular usage scenarios is a formidable challenge. The appearance of objects in a 2D image varies greatly, depending on illumination conditions, surface materials, and object orientation. These variations in the image complicate the task of software that must interpret the scene. On the other hand, the 3D shape of objects is invariant to those confounding effects. Stereovision based 3D recovery is computationally complex and fails on un-patterned surfaces. RADAR, ultrasonic, scanning LADAR, and other ranging technologies are similarly proposed, but they have difficulty discriminating objects due to limited temporal or angular resolution; moreover, the need for specialized sensors for each safety function poses system integration challenges. A single high frame rate focal-plane-array 3D sensor is desirable because it can serve multiple safety and convenience functions simultaneously, allowing applications to jointly exploit shape and appearance information in a dynamic scene. The output of the sensor should be a sequence of 2D arrays of pixel values, where each pixel value describes the brightness and Cartesian X,Y,Z coordinates of a 3D point on the surface of the scene.

Canesta has developed Electronic Perception Technology (EPT), an advanced range camera module based on pixels that measure the time delay of modulated infrared light from an active emitter. Electronic Perception technology permits machines, consumer and electronic devices, or virtually any other class of modern product to perceive and react to objects and individuals in the nearby environment in real time, particularly through the medium of "sight," utilizing low-cost, high-performance, embedded sensors and software.

The camera module contains a light source constructed from a bank of infrared LEDs (or laser diodes), a lens system for the detector chip, a detector chip with 160×120 phase-sensitive pixels fabricated on standard CMOS process, and an embedded CPU for application processing.

This paper first surveys several automotive safety applications. Next, quantitative performance metrics are defined for range cameras, which are necessarily evaluated on a different basis from 2D cameras. It explains geometric uncertainty, accuracy, and dynamic range as important factors in the performance of range sensors. Then it describes how the automotive applications requirements map to the sensor performance metrics. Next, the paper describes the principles of operation of the EPT range camera, highlighting design features that overcome challenges to this sensing method, including resilience to full-sunlight illumination. Finally, the paper presents two application-specific configurations of the range camera and characterizes its performance for those applications.

2 Safety Applications and Sensor Requirements

Growing government legislation, increasing liability concerns, and the inevitable consumer desire for improved safety make the introduction of new safety features a high priority for automakers. Today, various sensing technologies play a key role in delivering these features, detecting conditions both inside and outside of the vehicle - in applications like parking assistance, adaptive cruise control, and pre-crash collision mitigation. Each of these applications is characterized by a unique customized technology (e.g. ultrasonic, RADAR, LADAR, digital image sensing, etc.), which generally provides either a ranging function or an object recognition function.

The need for investment in multiple disparate technologies makes it challenging to deploy individual safety features as quickly or as broadly as desired.

Future applications pose even more difficulties, as multiple features must be provided in a single vehicle. Plus, virtually all of the new sensing applications on automakers' roadmaps (e.g. pedestrian detection being planned in Europe and Japan) require both ranging and object recognition functions. Combining two incongruent technologies to accomplish this task (such as RADAR and digital image sensing) is expensive, difficult to implement, and poses the additional problem of inefficient development.

The use of vision gives added levels of discernment to the air bag systems by providing static or dynamic occupant classification and position sensing. Further, the addition of a vision system inside the cabin enables other value-added applications such as abandoned baby/pet detection, personalization, and security. Applications for vision-based sensing outside the car are blind spot detection, vehicle lane departure, safety in rear vision, proximity of other vehicles around the vehicle, and off road and heavy equipment proximity sensing. The benefits of vision sensors are two fold. They provide enhanced visual feedback to assist the driver in operating the vehicle. But more importantly, when vision sensors also provide range data, they provide the necessary information for advanced algorithms to achieve higher level of discernment and more accurate analysis of object motion dynamic. With such sensors, for instance, the system can use the shape differences between a person and a large box sitting in the front seat to deploy the air bag or not.

We consider the requirements of some of these applications after defining the relevant range sensor performance metrics.

2.1 Range Sensor Performance Metrics

An imaging sensor’s lateral resolution can be measured in pixels per degree (pixels/°), but field distortion in typical lenses tends to reduce the resolution in the periphery of the image.

The geometric measurement performance of a range camera can be characterized by several statistics. For any quantity repeatedly measured over time, we define accuracy as the RMS error between the time average of that measurement and ground truth, uncertainty as the RMS error between instantaneous measurements and the time average, and total uncertainty as the 3-sigma confidence bound between the instantaneous measurements and ground truth. The spatiotemporal extent of the averaging regions for these statistics should be chosen based on the application. For occupant sensing, head-sized regions spanning 20cm X 20cm over 40 frames are appropriate.

Specifically, the range camera produces a 3D measurement P=(X,Y,Z) for each image pixel i at each frame t. By convention, the coordinate system is defined with respect to the camera, with the Z axis in the viewing direction, and X,Y being lateral directions. If Z’ denotes the time average of Z measurements, and Z* denotes the ground truth, then we may compute the statistics as follows:

- Accuracy in Z is the RMS error $\surd\Sigma_i\, n_i(Z_i'-Z^*_i)^2/N$.
- Uncertainty in Z is the RMS error $\surd\Sigma_i\, \Sigma_t\, (Z_{it}-Z'_i)^2/N$.
- Total uncertainty is $|\mu|+3\sigma$, where $\mu=\Sigma_i\, \Sigma_t\, (Z_{it}-Z^*_i)/N$ and $\sigma=\surd\Sigma_i\, \Sigma_t\, (Z_{it}-Z^*_i-\mu)^2/N$.

Performance statistics for X and Y are analogous.

For a conventional imaging camera, dynamic range is defined as the ratio of the largest non-saturating (i_{max}) input signal to the smallest detectable input signal (i_{min}), and is reported as $20\log_{10}(i_{max}/i_{min})$ dB. The objective of a high dynamic range imaging camera is to reproduce brightest and darkest segments of the scene. The ambient light works in favor of a conventional camera as it illuminates the scene for better imaging. On the other hand, the goal of a range camera is different. A range camera is evaluated based on its ability to detect range regardless of the ambient light condition. Therefore, the dynamic range of a range camera is defined as the ratio of the largest non-saturating input signal for detecting a range to the smallest detectable range signal. The dynamic range of a range camera can be further refined as the value of the usable dynamic range that satisfies certain Z uncertainty requirements.

Ambient illumination from the sun and artificial lights is pervasive in automotive applications. A critical requirement, therefore, is for the sensor to meet performance specifications even when viewing sunlit surfaces. Indeed, the sun at noon will project roughly $1KW/m^2$ on sunlit surfaces. The EPT camera module's optics contains an IR filter to pass only light whose wavelength matches the active emitter (e.g. 870nm), reducing the amount of surface sunlight that reaches the sensor to about $50\ W/m^2$. Still, compared to roughly $2\ W/m^2$ projected by the active light source of the camera, a significant amount of sunlight reaches the sensor. The range camera must provide the technology to cancel the effect of the remaining ambient light that reaches the sensor.

Another challenge in automotive environments is operational temperature range, at least -40°C to +85°C. The sensor performance must remain within the specification required by each application for the range of automotive temperatures. Finally, for exterior applications, the sensing modality must be robust to degraded viewing conditions of rain, mud and fog. The mud and rain effect can be mitigated by either placing the sensor inside the car (e.g. behind the window glass), or by carefully protecting the lens by designing appropriate enclosures, or providing a washing and cleaning system.

2.2 Occupant Sensing for Advanced Airbag Deployment

The functionality of high volume production air bag systems was once limited to detecting a collision and then deploying the air bag. Improper seating of children in the front passenger seat, small stature adults, and out-of-position occupants were therefore major considerations. Of concern is the firing of an air bag when a person's head is too close to any of the vehicle air bags or there is something blocking the effective deployment of the air bag. The critical out-of-position (COOP) zone near the dashboard is an area where a person is likely to be injured by airbag deployment.

Government and Consumers are pushing the Automotive Industry for improvements to the air bag systems in vehicles on the road today. The U.S Department of Transportation's (USDOT) National Highway Traffic Safety Administration (NHTSA) has a legislative mandate under Code of Federal Regulations (CFR) Title 49, Chapter 301, Motor Vehicle Safety, to issue Federal Motor Vehicle Safety Standards (FMVSS) and Regulations to which manufacturers of motor vehicle and equipment items must conform and certify compliance. Occupant crash protection (571.208) lists the requirements for the air bag systems [Dot04]. This document is the basis for the vehicle manufactures to improve the air bag systems.

Advanced air bag systems are requiring additional input information. A smart sensor system would be desired that can discern presence, size, location within the passenger space, and category of object. Through the real time monitoring of the space, the air bag system will have the ability of making more informed decisions on when, if, and with how much force the air bag will deploy. An air bag system with smart sensors has also the potential of warning the passenger of unsafe seating posture during the normal operation of the vehicle.

Today's "advanced airbag systems" are primarily based on weight sensors located in the seat cushion or floor. However, weight sensors alone cannot provide all the necessary information for intelligent deployment and furthermore pose implementation challenges for the automaker. They are difficult to integrate with the vehicle, e.g. requiring designing the seat around the sensor, and are different for each vehicle model, contrary to a common components strategy across the product line.

In contrast, Electronic Perception Technology is ideally suited for passenger localization [Gok04a] and classification [Gok05]. The EPT range camera updates three dimensional data to the air bag system every 16 msec. This gives the air bag system time to calculate the passenger's velocity and distance to the COOP boundary. With this information the air bag system can determine the air bag deployment strategy that is required by the situation, based on FMVSS requirements.

For the front passenger occupant sensing application, the volume that should be monitored is the space from floor to roof, from seatback to dashboard. With the range camera mounted in the center of the headliner, the required field of view is typically 110° vertical X 80° horizontal, and the farthest object may be up to 100cm from the sensor. Within this volume, the geometric uncertainty should be 0.5cm or better, and commensurately the image lateral resolution should be 4 pixels/°. The geometric accuracy should be 2cm or better in the region likely to be occupied by the passenger's head. These specifications are needed by intelligent algorithms that predict the location of the head from the dashboard after a crash is detected. Furthermore, the distance of head from the side window pillars is useful for the decision on the deployment of the side air bags. The range camera should provide 3D measurements at least 60 times per second, with latency not exceeding 16ms, to give the airbag system sufficient time to calculate the person's velocity and distance to critical zone boundaries. Lower latency values are possible by reducing the window of interest size of the camera frame and obtain a rapid burst of data for a subset of field of view.

2.3 Backup obstacle warning

The current trends in car styling and vehicle configuration, as in SUVs and Minivans, have reduced backup visibility of the driver. Video and ultra-sound backup warning systems are now fairly common in vehicles. A backup system provides visual and/or audio feedback to the driver when the object appears too close in the trajectory of the car. A backup system can warn the driver if an object is in the blind spot of the driver, or assist the driver during the parking. Backup systems are presently implemented by placing multiple ultra-sound sensors in the bumper of a car.

For the backup sensor to be useful for the driver, it has to correctly estimate the distance of the car to an object that is in the trajectory of the car, even if close to the ground. It should not be triggered by roadside objects or the road itself. Otherwise, the sensor can become a nuisance rather than being a useful tool for the driver. The 2D video sensors have the advantage of giving a good view of the back of the car. However, estimating the real-world distances from a video image is a difficult task for the driver. The 3D image sensors have the advantage of providing both the visual and positional information in this application.

It is estimated that 90% of backup collisions with pedestrians occurred at a vehicle velocity of 8 km/h or more. At that speed, it is necessary to monitor 4m behind the car just to achieve 95% successful accident aversion [Gla05]. Accordingly, for the backup warning system, the volume that should be monitored is a 4m wide lane from the back of car to 10m behind the car, from ground to 1m. The geometric accuracy should be 10% of the radial distance or better and the uncertainty should be 1.5% of the radial distance or better. This suffices to detect a 15cm thick person or object lying on the ground at 4m distance. The image lateral resolution should be 4 pixels/° in order to detect a standing person at 10m range. Per ISO recommendation, the backup detection system including range camera should provide an audible alert to the driver within 350ms of the manifestation of a hazard.

3 Principles of Operation

A time-of-flight sensor module measures distance by observing the time delay ΔT between emission and detection of light that travels from an active light source, bounces off a surface in the scene, and returns to the camera. For a light source modulated at frequency f and surface at radial distance r, the delay can be expressed as a phase $\phi = 4\pi r f / c$; hence, distance can be recovered by measuring phase.

This section describes the principles of operation of the EPT range camera, highlighting design features to overcome challenges to this sensing modality, including resilience to full-sunlight illumination [Gok04b]. First we describe the operation of the detector pixels in the sensor chip, and then we detail the calculations performed by the CPU to recover phase and 3D coordinates.

3.1 Detector Pixels

The light source in a typical EPT range camera module is a bank of IR LEDs at 870nm wavelengths, switched on and off with 50% duty cycle at a frequency on the order of 44 MHz. Each pixel in the detector array contains a pair of charge accumulating gates, called A and B, which are alternately enabled in-phase and out-of-phase with the light source. The gates are reset to 0 at the start of each frame and integrate over the frame exposure time S, typically 1 to 32ms. As the surface distance and phase of received light varies, the amount of integrated charge in gates A and B will vary (Figure 1).

Even in the absence of modulated light, the charge in A and B would be biased from 0 due to dark current and ambient scene illumination. The offset is independent of the received phase, and the two pixels are enabled for the same amount of time, so the biases could be canceled by measuring only the differential signal D=A-B. The latter signal is digitized by an ADC and stored in a frame memory accessible to the CPU.

While ambient illumination does not theoretically affect the D signal, it could potentially cause the gates A and B to saturate. To prevent this effect, the sensor implements a patented Common Mode Reset (CMR) process that removes the same amount of charge from A and B gates, leaving the differential signal unchanged. Typically, 20 resets are applied during each exposure. The camera detector has also the ability to prevent the saturation of the differential signal. These measures to avoid saturation of the detectors are among key innovations that enable the EPT range camera to operate in high ambient illumination conditions typical of automotive applications.

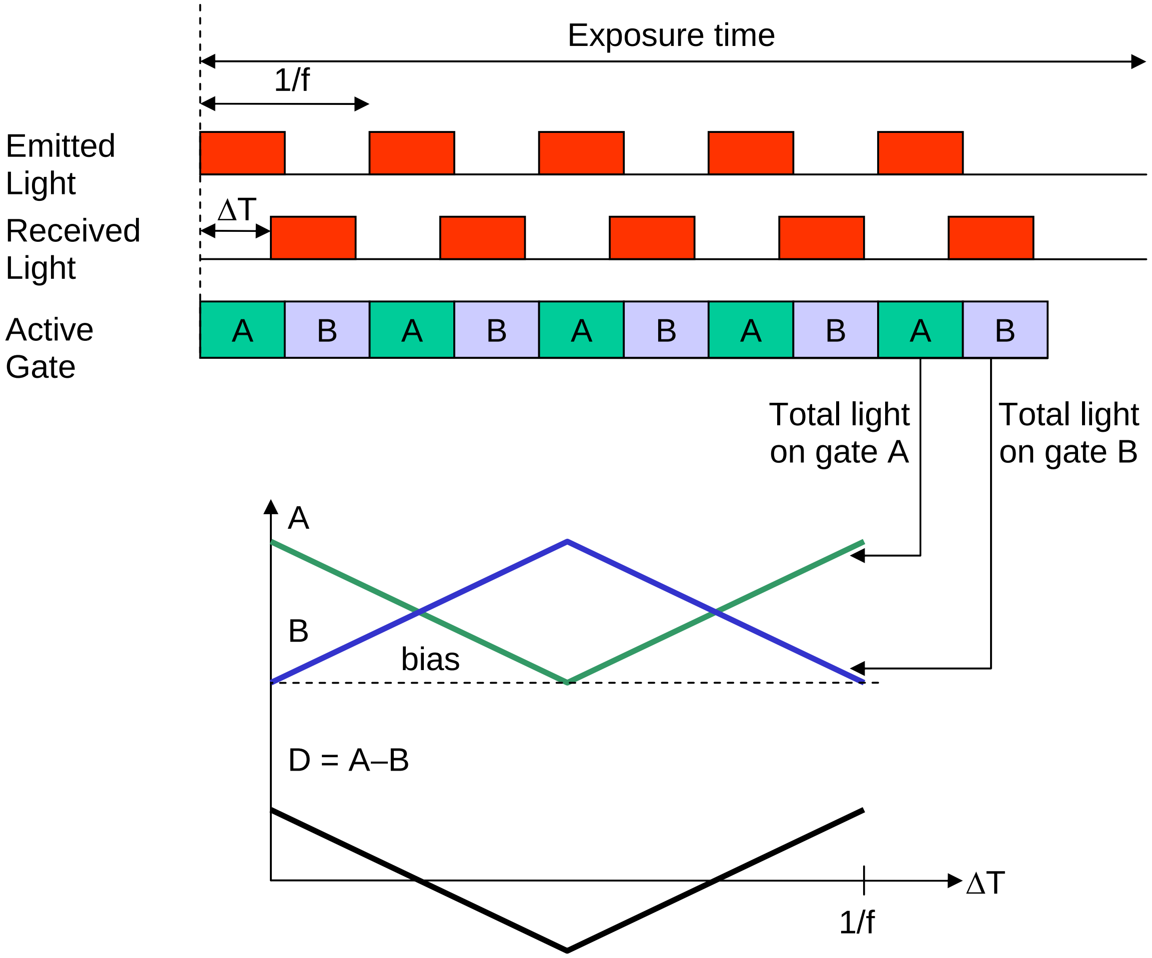

Figure 1: Principle of Operation of the EPT Detector Pixel

3.2 Recovery of Phase and Scene Geometry

While D depends on ϕ, phase cannot be determined from a single observation of D. Accordingly, the same process is repeated, sequentially or in parallel, wherein the light source modulation is delayed by 0°, 90°, 180°, and 270° with respect to the A/B clock. Then phase can be estimated by

$$\phi = \tan^{-1} \frac{D_{90} - D_{270}}{D_0 - D_{180}} .$$

Ideally, $\tan^{-1} D_{90}/D_0$ would suffice to recover phase, except that subtracting pairs of D observations with light phase 180° apart has the benefit of canceling residual biases found in the differential signals. In the sequential phase measurement, multiple exposures each of duration S results in lower uncertainty in depth data, whereas, in parallel operation, one exposure captures all four phases and the motion blur effects are minimized.

In sequential mode, since four exposures are required, it would seem that the EPT range camera can output a result only once every 4S seconds. However, since the light source modulation is cyclically varied in 90° steps, phase can actually be computed using the formula at the end of any one exposure because the last four exposures have four different modulation phases.

In an ideal system with co-located point light source and sensor, with non-distorting optics of known focal length, and perfect square wave signals, it would be possible to convert ϕ directly to radial distance r and then to X,Y,Z coordinates in space by simple geometry. In practice, however, $\phi = 4\pi rf / c$ does not hold exactly, due to pixel-to-pixel variations in gate timing, imperfect square wave signals, and use of a distributed light source. Moreover, a real lens distorts the field of view. Accordingly, a per-pixel calibration function may be used to map from measured ϕ value to Z. This function may be determined empirically for each pixel by measuring ϕ as a planar test target is moved to known Z positions. In addition, a calibration function can be used to map from each pixel (row,column) to a 3D direction ray (X/Z,Y/Z,1). That function is obtained by imaging a grid pattern of known dimensions. Scaling that vector by Z yields the desired 3D coordinates. By these means, the EPT range camera is capable of reporting accurate position information in an absolute sense, not just relative, when required by an application.

Active brightness, a measure of the strength of the received modulated signal, can be defined as $I_A = \sqrt{(D_0 - D_{180})^2 + (D_{90} - D_{270})^2}$. If active brightness in some pixel falls below a threshold, ϕ is deemed unreliable and no coordinates are returned. This feature is important because it informs the application on the confidence level of the depth data that is returned in the pixel. The application can take corrective actions by reconstructing the depth from other good pixels, or take a default action related to the application.

4 Performance Characterization

In this paper we describe range cameras for two applications that mate the 160×120 pixel detector chip with appropriate optics and light source. The geometric measurement performance of a range camera for occupant sensing has been measured. The uncertainty of the backup obstacle warning camera has been predicted using a system model.

Ongoing work includes characterizing the measurement performance of the EPT range camera over the full range of sunlight illumination and temperature conditions expected in an automotive environment, as well as verifying the frame rate and latency.

4.1 Occupant Sensing System

For the front passenger occupant sensing application, a lens was chosen to cover 104° vertical × 81° horizontal. Due to field distortion, the relative resolution is raised in the image center to 1.6 pixels/°. A light source was constructed from forty IR LEDs, radiating 5W peak optical power into a volume matching the lens field of view.

In order to measure the geometric performance of this EPT camera module, a planar test target with IR reflectivity of 0.5 was placed at six different distances from 15cm to 105cm. For each placement, sequences of 40 frames were acquired by the range camera for different values of exposure time S up to 32ms.

For analytical purposes, forty-three 20cm×20cm regions were chosen within the occupant detection volume at the six distances. Within each region, the metrics defined in §2.1 are calculated for the best S for that region. (The best S is selected as the exposure time minimizing the Z uncertainty metric. Typically S is 32ms, except for brightly lit regions that could saturate with such a long exposure.)

Figure 2 shows histograms of the performance metrics. Z uncertainty varies from 0.3cm to 1.5cm for different analysis regions, due to non-uniformity of the light-gathering power of the lens, non-uniform radiation from the active light source, and $1/r^2$ fall-off of power with distance. The median of each metric in each dimension meets the goals stated in §2.2.

In the sensor chip, the maximum non-saturating signal with 50 common mode resets is 2×10^5 ADC units, while the noise level in the signal is around 5, yielding a raw dynamic range of 92 dB. On the other hand, for range cameras a more relevant metric is the range of signal levels that satisfies a measurement performance goal. A regression model can be fit to the relationship between active brightness and Z uncertainty, predicting that I_A=480 suffices to achieve uncertainty of 1cm. On that basis, the dynamic range for satisfying the performance specification is 52 dB.

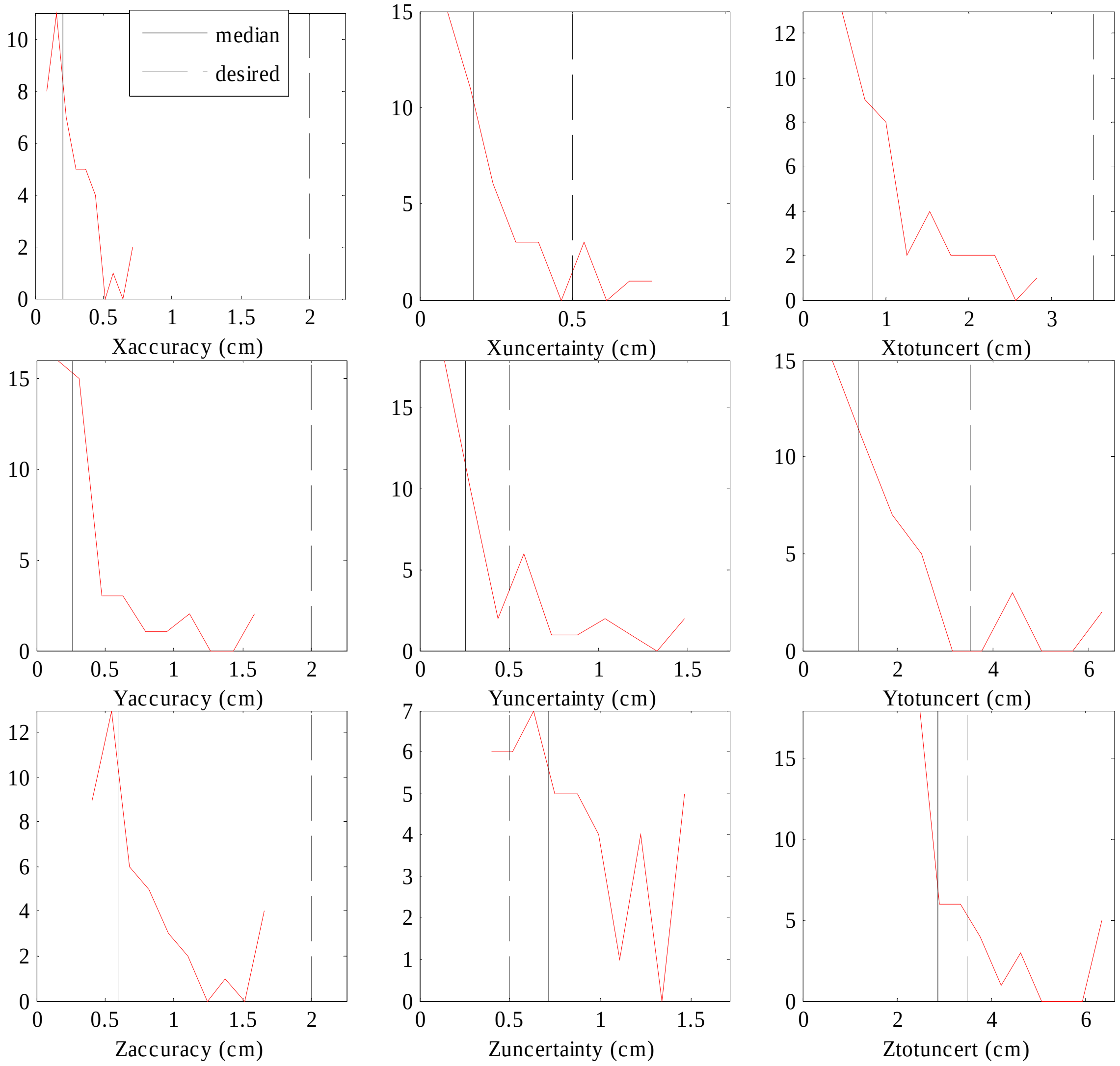

Figure 2: Histograms of Geometric Performance Metrics for Occupant Sensing EPT Range Camera

4.2 Backup Obstacle Warning System

We have developed a system model to predict the depth precision for a given active light source, target object, and optics, or alternatively the locus of 3D points where a specified precision is achieved. The model could be inverted to design the light source to achieve a specified precision on a given locus. The parameters of the model include:

- Light source model: optical wavelength, total power, angular distribution (assumes point source co-located with sensor)
- Object model: reflectivity and radial distance (assumes surface is perpendicular to line of sight)
- Optics model: focal length, F/#, transmission losses of lens and IR filter (assume no vignetting or field distortion)
- Pixel model: pixel area and quantum efficiency

- Analog to digital converter model: quantization step, voltage swing, exposure time
- Range estimator model: modulation frequency

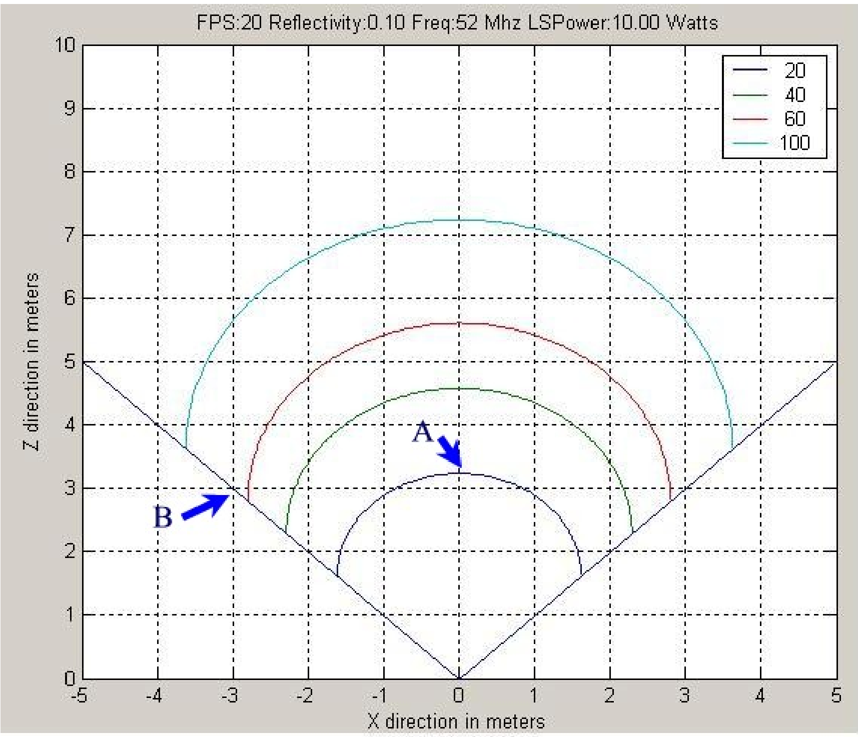

Figure 3: Iso-Uncertainty Contours, showing loci of 20,40,60,100mm uncertainty.

The object model converts the emitted light to received light at the lens, which the optics model converts to light power falling on a pixel. Assuming received and emitted light are in phase, the pixel model converts light power to a voltage, which the ADC model converts to a digital differential signal D_{max}. Depending on actual phase, D will vary from 0 to D_{max} in steps of 1. The range estimator model converts this to the corresponding step in radial distance.

Figure 3 illustrates the result of the model assuming 10W of light source power, 52MHz modulation frequency, 0.1 object reflectivity, 90° field of view and 20 frames per second. The figure shows iso-uncertainty contours for 20, 40, 60 and 100mm values as a function of distance from the sensor and the light source. For instance, the lower curve shows the set of points with uncertainty of 20mm. Clearly, the points that are closer to the sensor and aligned with the axis of the camera reflect more light to the sensor and thus produce better uncertainty values. The shape of the curves is influenced by the profile of the light source and the well-known cosine^4-effect. If necessary, the light source can be designed to compensate for the cosine^4-effect by increasing the concentration of the light on the boundaries of the field-of-view.

For the backup obstacle warning system, this configuration of the camera shows that distances for objects as far as about 3m (see arrow A in Figure 3) in the direction of the center axis of the camera can be measured with 2cm uncertainty. At center axis of the camera, 1.5% uncertainty requirement is easily met (§2.3). In the perimeter of the field-of-view, the model suggests that the present light source only gives about 6cm uncertainty (see arrow B in Figure 3) for objects in a plane that is 3m from the back of the vehicle where the camera may be mounted. The model suggests that an alternative light source with higher concentration of the light in off-center regions would be better suited for this application. We observe the clear advantages of using a system model for sensor performance analysis. The model provides the information that is needed for the camera system designer to tailor the camera and its light source for a vehicle application to achieve cost/performance goals of the application.

5 Conclusions

This paper described the advantages of Electronic Perception Technology, a high frame rate focal plane array 3D sensor, for automotive safety systems. Performance metrics and application-driven requirements were presented for two specific systems, occupant sensing and backup obstacle warning. We detailed the principles of operation of the range camera. We demonstrated the use of actual measurement and the use of analytical modeling to show how the two configurations of the camera module satisfy the requirements.

References

[Dot04] National Highway Traffic Safety Administration, U.S. Department of Transportation, *Federal Motor Vehicle Safety Standards*, 49CFR571.208, pp. 497-580, 2004.

[Gla05] V. Glazduri, “An Investigation of the Potential Safety Benefits of Vehicle Backup Proximity Sensors”, in *Proc. Int’l. Tech. Conf. Enhanced Safety Vehicles*, 2005.

[Gok04a] S.B. Gokturk, C. Tomasi, “3D Head Tracking Based on Recognition and Interpolation Using a Time-Of-Flight Depth Sensor”, in *Proc. Computer Vision Pattern Recog.*, Vol. II, pp. 211-217, 2004.

[Gok04b] S.B. Gokturk, H. Yalcin, C. Bamji, “A Time-of-Flight Depth Sensor—System Description, Issues, and Solutions”, in *Proc. IEEE Workshop Real-Time 3D Sensors, 2004.*

[Gok05] S.B. Gokturk, A. Rafii, “An Occupant Classification System—Eigen Shapes or Knowledge-Based Features”, in *Proc. IEEE Workshop Machine Vision Intell. Vehicles*, 2005.

Dr. Stephen Hsu, Sunil Acharya, Dr. Abbas Rafii and Richard New

Canesta, Inc.
440 N. Wolfe Rd.
Sunnyvale, CA 94085
USA
shsu@canesta.com, acharya@canesta.com, arafii@canesta.com, rnew@canesta.com

Keywords: range camera, time-of-flight camera, range video, lidar, 3D camera, 3D sensor, CMOS sensor, airbag occupant sensing, advanced airbag, dynamic suppression, backup obstacle detection, pedestrian detection, pre-crash sensing, electronic perception technology