Review of ladar: a historic, yet emerging, sensor technology with rich phenomenology

Paul F. McManamon
Review of ladar: a historic, yet emerging, sensor technology with rich phenomenology

Paul F. McManamon
University of Dayton
Ladar and Optical Communications Institute
300 College Park
Dayton, Ohio 45469-2950
E-mail: paul@excitingtechnology.com

Abstract. Ladar is becoming more prominent due to the maturation of its component technologies, especially lasers. There are many forms of ladar. There is simple two-dimensional (2-D) ladar, similar to a passive electro-optic sensor, but with controlled illumination and the ability to see at night even at short wavelengths. There is three-dimensional (3-D) ladar, with angle/angle/range information. 3-D images are very powerful because shape is an invariant. 3-D images can be easily rotated to various perspectives. You can add gray scale or color, just like passive, or 2-D ladar, imaging. You can add precise velocity measurement, including vibrations. Ladar generates orders of magnitude higher frequency change then microwave radar for velocity measurement, because frequency change is proportional to one over the wavelength. Orders of magnitude higher frequency change means you can measure a given velocity orders of magnitude quicker, in many cases making an accurate measurement possible. Polarization can be used. With an active sensor you control both the illumination and the reception, so you can pattern the illumination. Also, because ladar can use narrow band illumination it is easier to easier to coherently combine sub-aperture images to obtain the higher resolution of an array.

Subject terms: ladar; lidar; active imaging; laser radar; remote sensing.

Paper 120004V received Jan. 1, 2012; revised manuscript received Apr. 19, 2012; accepted for publication Apr. 20, 2012; published online June 5, 2012; corrected Aug. 21, 2012.

1 Introduction

Ladar started shortly after the invention of the laser but it is just now emerging as a widespread alternative to passive electro-optic (EO) sensors and microwave radar. Component technologies, especially reliable and affordable lasers, have developed to the point that the extremely rich phenomenology available through ladar can be more easily accessed. One driving factor in component availability is the similarity of many required ladar components to laser communications systems, which is driven by the high bandwidth requirements of the internet, a large and lucrative market. Inexpensive, highly capable, and very reliable active EO components are making ladar competitive compared to alternative sensor technologies. Ladars are being made from the visible through the LWIR. Twenty years ago COCO22 ladar was popular in the LWIR, but it has faded as solid state lasers have become more prominent. Ladar operating near 1.5 μm is becoming widespread.

2 Ladar Range/Signal to Noise

Ladar range and signal to noise calculations can be divided into two parts. The first part is to calculate how much signal is captured by the receiver (or how many photons hit each detector). You can then convert these photons to electrons, based on quantum efficiency. The second part has to do with how many photons you need in each detector to accomplish your objectives, such as object detection, or recognition, or tracking. This depends on the receiver used, and what your information objectives are. The discussion in

Sects. 2.1 and 2.2 uses simplifying assumptions with the objective of bringing out the key dependencies while avoiding the complexity of a fully general representation. The literature cited in the following sections can be used as required to consider more complex situations.

2.1 Calculating the Received Power, or Number, of Received Photons

To calculate the number of photons returned to each detector you start with the transmitted laser power. It can have a shaped beam such as a Gaussian, or you can assume the beam is flat topped. For the computationally simple flat top beam the radiance (Watts/cm²) is the laser power divided by the area of the beam footprint. This is a significant gain over radiating throughout a sphere since lasers have small beam divergence. Beam divergence can be smaller for large aperture transmitters and for shorter wavelength transmitters. We then create a fictitious area we call the cross section. This is not a physical area, but will be related to the physical area. For area targets if you have a flash imaging sensor, using many detectors, you can only count the cross section seen by each detector. Higher spatial resolution will mean each detector sees a smaller area and a smaller cross section. Therefore, if the target illumination area is fixed, an increased imaging resolution (e.g., increased number of detector pixels) results in decreased signal to noise per detector. The signal to noise can be increased by increasing the transmitter power. High range resolution will also reduce the effective cross section if there are scatterers at multiple ranges within a detector angular sub-tense (DAS). Surfaces with high reflectivity in the backward direction (toward the
ladar receiver) have higher cross section. Corner cubes have a lot higher cross section, because light reflected from a corner cube is returned in a small angular cone. The accepted definition of cross section is different for ladar than for microwave radars. For ladar, when specifying cross section it is usually assumed scattering is near Lambertian, and reflected light is reflected into π steradians. We arrive at π steradians as the effective solid angle of reflected light by assuming a cosine distribution of reflected light over a hemisphere (2π steradians). This is for a Lambertian scattering of light from a rough surface. In microwave radar the cross section definition usually involves scattering of light over 4π steradians from a small round gold ball. This makes sense for radar, where often the radar wavelength is longer than the size of the diameter of the round gold ball. For EO it does not make as much sense because the ball would be much larger than the wavelength, so it would block forward radiation. Another thing to consider is the shape of the target. A point target is smaller than the DAS. A line target, like a wire, is smaller than the DAS in one dimension, but larger in the other dimension. Area target cross section can be limited by the DAS or by the illuminated area. Often today we have arrays of detectors, and what we call “flash imaging”, where an area much larger than a given DAS is illuminated so you can see many pixels, or voxels, at one time. Once light is reflected from the object of interest some of it is captured by the receiver aperture. Obviously a larger receiver aperture captures more light. We also have efficiency terms to consider. There are losses in the ladar system, and only so much light makes it through the two-way atmospheric path to and from the target. The total optical power received at the detector is given by:

\[ P_R = P_T \times \frac{\sigma}{A_{\text{illum}}} \times \frac{A_{\text{rec}}}{\pi R^2} \times \eta_{\text{illum}} \times \eta_{\text{sys}}, \]  

(1)

where, \( P_T \) is the power transmitted, \( \sigma \) is the cross section in square meters, \( A_{\text{illum}} \) is the area illuminated, \( A_{\text{rec}} \) is the area of the receiver, \( R \) is the range, \( \eta_{\text{illum}} \) is the transmission efficiency through the atmosphere, and \( \eta_{\text{sys}} \) is the receiver system optical systems efficiency. You can see the power received is the power transmitted times two ratios of areas, times appropriate efficiency terms. The first ratio of areas is the cross section divided by the illuminated area at the object plane. The second ratio of areas is the receiver aperture area divided by the effective average area illuminated by Lambertion reflection.

For an area target with illumination area larger than the DAS we can assume that the cross section for a given receiver pixel is limited by the area of a pixel. For square receiver pixels we have:

\[ \sigma = \rho \frac{A_p}{d^2}, \]  

(5)

where \( d \) is cross range resolution, \( \rho \) is the reflectance of the area, and \( A_p \) is the area of the pixel at the target location, which for a square pixel is equal to \( d^2 \).

For a point target, or line target with dimensions smaller than the pixel size at the target location, the cross section will be smaller, due to a smaller area that is reflecting.

The cross range resolution of a pixel cannot be better than the diffraction limit, or

\[ d \geq \frac{R \lambda}{D_{\text{rec}}}. \]  

(6)

In Eq. (4) we have used the full width, half power, diffraction definition. Often other values are used, such has half width, or the width at zero power. For a circular aperture, we have

\[ A_{\text{rec}} = \pi \left( \frac{D_{\text{rec}}}{2} \right)^2, \]  

(7)

where \( D_{\text{rec}} \) is the diameter of the receiver. Equation (7) is for a single receive aperture. If you have multiple sub-apertures then calculating receive area can be more complex. The area illuminated can be no smaller than the diffraction limit as given in Eq. (8) below:

\[ A_{\text{illum}} \approx \pi \left( \frac{\lambda R}{2D} \right)^2. \]  

(8)

The area illuminated can be larger than given in Eq. (8) if the transmit beam is spoiled, or if the transmit beam is not diffraction limited. We can then invert Eq. (4) to obtain the required laser transmit power for a given ladar range

\[ P_T = \frac{\pi}{\sigma} \frac{P_R \times A_{\text{illum}}}{A_{\text{rec}} \eta_{\text{illum}} \eta_{\text{sys}}}. \]  

(9)

Terms in Eqs. (4) or (9) can be expanded if desired. When using these equations care must be taken to properly evaluate...
the cross section per detector pixel for a given target, such as area targets, line targets, or point targets. Assuming enough is known about the target scattering properties and its shape, calculation of the effective cross section is straightforward but must be done with care.

### 2.2 Calculating the Detection/Recognition Threshold

In order to use Eq. (9) to determine the laser power that must be transmitted, \( P_R \), to achieve detection at a given ladar range, we must first determine how much optical signal power, \( P_R \), we need to receive in order to meet the desired probability of detection and false alarm requirements.

The \( P_R \) term in Eq. (9) is received optical power, which converts on hitting the detector to photons per second. Photons per second in turn convert, with some quantum efficiency, to current, depending on the detector. For heterodyne detection the conversion from received optical power to electronic power will be linear because of the local oscillator, as will be discussed later. Received optical power converts to current, which has to be squared for direct detection to get electronic power as used in the signal to noise calculations.

The received optical power is related to the rate of arrival of the received photons as given by Eq. (10):

\[
P_R = \frac{Nhc}{2T_m},
\]

where \( T_m \) is the period of time over which the measurement is made and \( N \) is the number of photons per pixel received during that measurement time. For pulsed ladar systems \( T_m \) will often be the pulse width.

When optical signals hit the detector they create currents in the receiver electronics which have associated mean levels and noise fluctuations. There are also current fluctuations in the receiver even when the detector is not illuminated by any optical signals (i.e., dark noise). When the signal and noise sources are considered, the resulting electrical signal to noise is given by:

\[
\text{SNR} = \frac{\langle i_s^2 \rangle}{\sigma_n^2} = \frac{\langle i_s^2 \rangle}{\sigma_{ns}^2 + \sigma_{nBK}^2 + \sigma_{nDK}^2 + \sigma_{nTH}^2} = \frac{2eB[R_D P_R + R_D P_{BK}]}{2eB[R_D P_R + R_D P_{BK}] + 2eB_iDK + 4ktB/R_{TH}}.
\]

where \( R_D \) is the detector current responsivity, \( e \) is the electron charge, \( B \) is the detection bandwidth, \( P_R \) is the received optical power at the detector as noted in the equations above, \( P_{BK} \) is the received background optical power at the detector, \( i_{DK} \) is the detector dark current, \( k \) is Boltzmann’s constant, \( T \) is the temperature in Kelvin, and \( R_{TH} \) is the effective load resistance that creates the same thermal noise spectral density as the receiver electronics. The first through fourth terms in the denominator represent the signal shot noise, the background light shot noise, the dark current shot noise, and the thermal (or Johnson) noise, respectively. By using the relation between the responsivity and quantum efficiency of the detector, \( \rho_D = \eta_D e/ hf \), the SNR can also be written as

\[
\text{SNR} = \frac{\eta_D P_R^2}{2hfBP_{BK}}.
\]

In the limit where the signal-shot noise dominates the other noise terms the SNR is given by

\[
\text{SNR}_{\text{signal}} = \frac{\eta_D P_R T_m}{hf} = \frac{\eta_D E_R}{hf} = \eta_D N.
\]

where \( E_R \) is the optical energy received, and as previously defined, \( N \) is the number of received photons incident upon the detector during the measurement time \( T_m \). In the limit that the signal shot noise dominates other noise terms the SNR is directly proportional to the number of photons received.

For a coherent ladar, the SNR is given by:

\[
\text{SNR} = \frac{\langle i_s^2 \rangle}{\sigma_n^2} = \frac{\langle i_s^2 \rangle}{\sigma_{ns}^2 + \sigma_{nBK}^2 + \sigma_{nDK}^2 + \sigma_{nTH}^2} = \frac{2eB[R_D P_{LO} + P_R + P_{BK}]}{2eB[R_D P_{LO} + P_R + P_{BK}] + 2eB_iDK + 4ktB/R_{TH}}.
\]

\[
\text{SNR} = \frac{\eta_D P_{LO}^2}{2hfPR_{LO}} = \frac{\eta_D \bar{P}_{LO}^2}{2hfPR_{LO}}.
\]

One way people eliminate most of the noise terms in direct detection ladar is by using gain on receive. You can use an avalanche photodiode to amplify the signal. This is now very common, as will be discussed later. You could also use a fiber amplifier to amplify the signal. Historically people have used photo multiplier tubes. Although receiver optical or electronic avalanche gain is not included in Eqs. (10) or (12), introduction of gain effectively minimizes the some of the noise terms. In ladar, each photon contains much more energy than in microwave radar, making shot noise more important for ladar than it is for radar. The shot noise comes from the quantum nature of electromagnetic radiation. For a background limited direct detection receiver we have:

\[
\text{SNR} = \frac{\eta_D P_R^2}{2hfBP_{BK}}.
\]

In the limit where the signal-shot noise dominates other noise terms the SNR is given by

\[
\text{SNR}_{\text{signal}} = \frac{\eta_D P_R T_m}{hf} = \frac{\eta_D E_R}{hf} = \eta_D N.
\]
where $\eta_{\text{het}}$ is the heterodyne efficiency, which depends on how well the return signal and LO fields are matched on the detector, and $P_{\text{LO}}$ is the LO power. The new noise term in the denominator (compared to the direct detection SNR) is due to the shot noise of the local oscillator. It should be noted that this is usually referred to as CNR in the literature, meaning carrier to noise ratio, referring to the mean electrical power at the IF carrier frequency (difference frequency) compared to the noise power.

When a coherent receiver is utilized the return signal is optically combined with a local oscillator (LO) signal. The resulting total optical intensity (and power) has fluctuations near DC, at the difference and sum frequencies of the two fields, and at double the frequency of each of the fields. For optical frequencies, the higher frequency power oscillations are well beyond the maximum frequency response of detectors, so the only power fluctuations that are detectable are those near DC and at the difference frequency. The coherent receiver is usually designed to isolate the difference frequency component from fluctuations and noise at other frequencies. The rms amplitude of the optical power fluctuations at the difference frequency is $\sqrt{2\eta_{\text{het}}P_{\text{LO}}P_{\text{R}}}$, which results in a mean squared signal current of $\langle i_i^2 \rangle = 2\eta_{\text{het}}^2 P_{\text{LO}} P_{\text{R}}$ as seen in the equations. Therefore, the electrical power measured in a heterodyne receiver is linearly proportional to the optical power received. The factor of two is eliminated from the denominator because of the factor of two in the signal times LO power mentioned above. In addition, the LO adds additional shot noise which is accounted for in the denominator with the addition of the $P_{\text{LO}}$ term.

For coherent ladar the local oscillator power can be increased to dominate other noise sources, assuming the detector dynamic range can handle the local oscillator power. In general it is better to use AC coupling with a heterodyne receiver to reduce the impact on the dynamic range resulting from having a high power local oscillator. For a well-designed heterodyne case, the main noise will be shot noise from the local oscillator power, and the resulting SNR, is given by

$$\text{SNR} = \frac{\eta D P_{\text{R}}}{hfB}, \quad (18)$$

where $B$ is the bandwidth. For a measurement duration (or pulse width) of $T_m$, the matched filter receiver for the heterodyne detection IF signal at the difference frequency has bandwidth of $B = 1/T_m$ which yields

$$\text{SNR}_{\text{LO, limit}} = \frac{\eta D P_{\text{R}} T}{hf} = \frac{\eta D E_{\text{R}}}{hf} = \eta D N. \quad (19)$$

Note that except for the $\eta_{\text{het}}$ efficiency factor this is identical to the equation for the signal-shot-noise-limited SNR for direct detection [Eq. (16)]. The difference is that for the well-designed heterodyne detection receiver (with sufficient LO power), the SNR is proportional to the number of photons received even when the signal is very weak, whereas, for the direct detection receiver the signal hitting the detector must be strong enough so that its shot (photon) noise dominates all other noise. In direct detection amplification is often used to enhance the received signal. If a coherent receiver is truly shot noise limited, and if the heterodyne efficiency is unity, then the coherent SNR will always be greater than or equal to the direct detection SNR (assuming detectors having the same quantum efficiency). This is not to say that the probability of detection and probability of false alarm are always better for coherent detection, as those depend on the statistical fluctuations of the signal and noise (primarily signal).

Speckle is commonly seen in narrow band laser light that is scattered from a rough surface, such as a wall, as the bright and dark regions in the scattering volume. Speckle fluctuations of the signal can affect the probability of detection and false alarm. Speckle fluctuations of the signal have the biggest impact on narrow laser line width ladar because speckle comes from interference between reflections from various portions of the target. Narrow band signals interfere with each other, whereas that interference is averaged out with a broadband signal. This is why a flashlight on a wall does not produce the same bright and dark pattern on a wall as a narrow band laser. The net speckle interference can range from fully constructive to fully destructive. Speckle can be more easily mitigated in a direct detection ladar because you do not need a narrow line width laser source. Coherent ladar uses a narrow line width laser source so you can measure phase by beating the return signal against a local oscillator, and have the resulting beat frequency be measurable within the detector bandwidth. This is interference with the LO is the same phenomena as interference of the return portions off the wall, which is why a narrow band signal will interfere with itself upon reflection from a rough surface. Broad band light sources average out this interference. Because of the very high carrier frequency of light it is impractical to exactly model the interference between the various reflections, as can be done in the microwave region. Instead we treat speckle as a statistical process. For a full electromagnetic code simulation this process would be deterministic, but for the foreseeable future that is beyond our computational abilities.

The discussion above calculates the SNR. For any given SNR you can pick an operating point that defines the probability of detection and the probability of false alarm. The radar community has worked this issue and usually uses equations derived from a 1947 report by Marcum and a 1954 report by Swerling. Swerling case 2 is for independent pulse to pulse variations. It is often quoted. In ladar you will have pulse to pulse variations in return signal if you have enough change in angle to produce a new speckle pattern. For any given measurement in order to meet a given probability of detection and false alarm criteria, a certain SNR (related to the number of received photons) is must be achieved. References 1 and 8 are good summary of that addresses both direct and coherent detection ladar and performance for general levels of speckle averaging (speckle diversity).

### 3 Two-Dimensional Ladar

Two-dimensional (2-D) ladar is similar to a passive imaging, but with illumination. You can use a gated framing camera to capture photons in an array. The main benefit of this type of ladar compared to passive sensors is that it will work at night while using shorter wavelengths, therefore you can have enhanced resolution. Figure 1(a) and 1(b) shows an 8 to 10 $\mu$m FLIR image and a 1 $\mu$m gated laser image of the same object,
There was an Air Force program called ERASER with the objective of developing a 2-D ladar like this. You can see the significantly enhanced resolution available using a shorter wavelength. Shorter wavelength does not directly cause higher resolution, but the diffraction limit at shorter wavelengths allows more leeway to use longer focal length imaging optics. Also, since you bring your own illumination there are no thermal cross over issues. Both 1.06 μm and ~1.5 μm 2-D active imagers have been developed and tested.

To enhance signal to noise a gated 2-D camera is preferred, which will only gather noise over a short period of time, as compared to continuously gathering noise. This type of ladar will frame at low rep rates, consistent with the framing cameras. 10 to 30 Hz would be a typical rep rate. While nanosecond class laser pulses are not a requirement, it is likely Q switched lasers will be used, resulting in 5 to 15 ns pulse widths.

4 Three-Dimensional Ladar

Three-dimensional (3-D) ladar will measure azimuth, elevation, and range. The last portion of this section discusses range measurement. The initial portions discuss methods of measuring the azimuth and elevation angular positions. Angle/angle information can be measured by scanning an individual, or small number, of detectors, or it can be measured by simultaneously illuminating an array of detectors. In order to measure range accurately a high bandwidth measurement is required. Three methods of “flash” 3-D imaging are discussed. Flash 3-D imaging measures a 2-D array of angles simultaneously.

4.1 Scanning 3-D Ladar

The first 3-D ladars were developed using one, or maybe as many as 8, individual detectors. These were then scanned to obtain an image with a large number of pixels. Obviously high rate beam scanning, and high rep rate lasers, are required. To obtain 3-D imaging usually one uses a high bandwidth detector. The range resolution of the ladar is defined based on

\[ \Delta r = \frac{c}{2B}, \]  

where \( c \) = the speed of light, and \( B \) = bandwidth of the signal. For a pulsed system the pulse width is one over the bandwidth, but alternate modulations can also provide the same range resolution. The high required bandwidth per detector is the reason why initially 3-D ladar used individual detectors, or small arrays of scanned detectors. Obtaining large focal plane arrays with this bandwidth has proven to be a development challenge. As a result all of the initial 3-D ladars were scanning ladars. Obviously with a scanning ladar, larger format images require a higher rep rate laser in order to create entire images at acceptable frame rates. To date scanning ladars all use detectors that can measure gray scale. Ideally you also would like to have detectors that can capture a range profile, if multiple scatterers are in a given angle/angle cell. Figure 2 shows a 3-D image of San Francisco taken with a commercial OPTEC scanning ladar.

4.2 Flash 3-D Ladar

4.2.1 Geiger mode APD based flash 3-D ladar

Geiger mode APD based flash ladar has been pioneered by MIT/LL. MIT/LL has made Geiger mode APD cameras with up to \( 64 \times 64 \) × \( 256 \times 256 \) detectors. They started with a \( 32 \times 32 \) array. They initially made Silicon based Geiger mode APDs that work in the visible and in the lower portion of the near IR. More recently they made Geiger mode APDs that can operate at 1.06 μm, and then ones that can operate at 1.55 μm. Dark current is higher for longer wavelengths. MIT/LL is working on pushing the wavelength even further. More recently two companies have commercialized Geiger mode APD arrays, Princeton Lightwave and Boeing Spectralab. Both of these companies have \( 32 \times 32 \) array based cameras available at 1.06 μm, and at 1.55 μm. They are both developing \( 32 \times 128 \) format cameras.
Geiger mode APDs can have a very large avalanche gain for any photon hitting a detector. People refer to this as photon counting because the signal amplitude is sufficiently large from a single photon that it can be detected. One disadvantage of a Geiger mode APD is there is a dead time after each triggered event. During the dead time the detector will not detect any received photons. Also, there is no ability to measure the signal intensity per pixel (or gray scale image) on each pulse. 100 photons create the same signal as one photon, so you cannot inherently see gray scale. Also, there may be some cross talk between detectors. Dark current can be an issue. These attributes are discussed in the various references previously provided. Geiger mode APD flash imagers tend to run at high rep rates because you do not need a lot of energy per pulse to obtain a response. You keep energy per pulse, and probability of detection, low on a single pulse, and then integrate a number of pulses. This low energy requirement can allow us to use 1.06 μm radiation without eye hazard because of low single pulse intensities. Initial Geiger mode APD cameras operated at a rep rate of about 20 kHz. Now Princeton Lightwave has a camera that can operate up to 180 KHz. You can develop an effective gray scale by using multiple pulses. If you have multiple pulses return from a given location, and keep the probability of detection low, then a higher reflectance area will have a higher probability of return, causing more events to trigger where you have high reflectance. Accounting for the total number of counts (signal trigger events) from each pixela cross multiple pulses results in an effective gray scale. When you use multiple pulses to create an effective gray scale you effectively lower the frame rate. For Geiger mode APDs it takes more energy to 3-D map an area with gray scale than to simply 3-D map the area, because of the requirement for multiple pulses. Also, if you have pixels with mixed range returns you can play essentially the same trick as used with gray scale to map the returns as a function of range. If the probability of triggering is low for any event then you will get events triggering at various ranges. There is a slight bias toward nearer ranges because of dead time after triggering, but this bias is slight if probability is low for triggering an individual event. With Geiger mode APDs both effective range profile as well as an effective gray scale require more photons than a simple 3-D mapping. For simple mapping Gieger mode APDs are however very sensitive because they are single photon counting. Dynamic range of the emitted laser power is another issue for Geiger mode APDs. For a given range target, and given transmit and receive aperture diameters, you need to set the emitted laser energy at the right level, or you will not achieve the right probability of an avalanche. Geiger mode APD based flash imaging is a highly efficient method of 3-D mapping an area. It becomes less efficient as you require gray scale or range profiles. A key advantage of Geiger mode imaging ladar is the relative simplicity of the receiver and data acquisition electronics compared to wide-bandwidth linear mode receivers. The primary disadvantage is that in cases where the target has range depth and/or when gray scale information is needed the total energy required to map an area can be significantly higher than for a linear-mode photon counting receiver. Of course it is only recent that linear mode APDs are approaching photon counting sensitivity.

Figure 3 shows early Gieger mode APD images compared to a low light level TV. Notice the significant reduction in required number of photons compared to a low light level TV, and the fact you can gate out the camouflage when processed appropriately.

MIT/LL has done an interesting experiment with Geiger mode APDs, allowing them to be used in an unusual heterodyne mixing approach. Normally you might not think of a Geiger mode APD as being capable of doing heterodyne detection ladar. MIT/LL took the approach of using a low power local oscillator. This means you do not increase sensitivity by doing heterodyne, but Geiger mode APDs are already sensitive. Low LO power is required to avoid saturation of the Geiger mode APD. In this paper MIT/LL gangs together a 44 × 44 array of detectors into a super pixel. If the LO power is on the order of the signal power, and if the probability of detection is low, then the beat frequency between the LO and the returned signal will result in more and then less detections. The beat frequency must be kept < half of the frame rate of the camera. For the MIT/LL experiment in this paper the frame rate was 20 KHz, meaning the beat frequency had to be under 10 KHz.

4.2.2 Linear mode APD based flash 3-D ladar

Linear mode APD cameras have also become available. ASC is a company that has pioneered this approach, especially for commercial applications. They sell a 128128 × 128128 pixel 1570 nm flash 3-D imaging camera. It is a linear mode APD. The camera will frame at 1 to 20 Hz, or at 30 Hz in burst mode. The ASC receiver arrays have a noise floor that is significantly higher than a single photon. Therefore, the energy required to image a given area at some range is higher for the ASC receiver than for a Geiger-mode receiver. At short ranges this is not an issue. The commercial products tend toward relatively short range operation, say <11 km/km. When you go to longer range operation you will require higher pulse energy or fewer pixels (lower area coverage rate). Pulse widths should be 2 ns to obtain 1 foot range resolution if the full pulse width is used. A sharp rise time can provide better resolution than provided when using the full pulse width. This type of camera will measure gray scale on a single pulse, since the output is proportional to the reflected light. A camera like this can provide a range profile from a single pulse, so long as range profile
storage is built into the ROIC of the device. Building in this storage can make the ROIC physically larger. Sensitivity for 3-D mapping will currently tend to not be single photon, but there is development in that direction.\textsuperscript{26–31} Figure 4 shows gated imagery through a sand cloud, from an ASC 3-D imager.\textsuperscript{32}

Raytheon and DRS have made significant progress in developing high sensitivity linear mode APD arrays. High gain in the APD will reduce the effect of any noise introduced after the amplification gain. As sensitivity of linear mode arrays increase the main advantage of Geiger mode APDs becomes less important.

4.2.3 Polarization based flash 3-D ladar using framing cameras

In the early 90’s the Air Force had a program called LIMARS, Laser Imaging and Ranging System. Multiple patents were awarded using this technology.\textsuperscript{33,34} The main idea is to replace a high speed camera with a Pockels cell and a couple low frame rate cameras\textsuperscript{23–37} for flash 3-D imaging. One of the challenges of flash imaging is having a large enough focal plane array to detect an area based object with a single pulse. In the 90’s we did not have area based detector arrays with high enough bandwidth to measure range to nanosecond precision. In the LIMARS receiver, high bandwidth cameras are not required. Temporal (range) resolution is provided by a high-speed Pockels cell, as described in the following. Figure 5 shows a diagram of the LIMARS receiver.

Light enters the receiver. A single polarization of return light is isolated. Alternately you could use twice as many cameras and detect both polarizations. A ramp is placed on a Pockels cell to switch polarization as a function of time. Two standard framing cameras are used. In any given detector the ratio of power in one camera versus the other camera provides range information. A steeper slope provides more accurate range information, but also repeats quicker. In order to expand the unambiguous range there are a number of standard techniques that can be used, such as chirping the length of the ramps. The big advantage of this technique is that you can use a pair of standard framing cameras for high range resolution. The biggest disadvantage is that you need to use a Pockels cell to rotate polarization. Pockels cells traditionally require high voltage and have a narrow field of view. You can use other waveforms on the Pockels cell besides the saw tooth waveform shown, but a saw tooth is a good waveform for this purpose.

In the visible Silicon based TV cameras will work fine for this technique. Visible cameras can have a large number of pixels. Even in the NIR, 1.5 $\mu$m region, you can buy $320\times320$, $640\times512$, or now even $1280\times1024$ pixel, 15-$\mu$m pitch, military-hardened SWIR camera.\textsuperscript{38} These formats are larger than available with high speed cameras for flash imaging that are discussed above. Figure 6 shows two images from the DARPA SPI 3-D effort, which uses this approach.\textsuperscript{39} Figure 7 shows an image from a small company called Tetravue, again using this technique.\textsuperscript{40}

4.3 Range Measurement

Ladars measure range based upon the time it takes for light to travel from the transmitter to the target and then to the receiver. Short pulses are one way to send a time reference from the transmitter to the target and back. Unless extreme accuracy is required the speed of light in vacuum is used. For very long range, and very precise measurements, this may not be accurate enough since there is a slight deviation of the index of refraction of air compared to vacuum, but this can be ignored in almost all cases. A useful rule of thumb is that the speed of light is about 1 foot per nanosecond. Ladar sends light out and back, so 1 ns yields about 6 inches in distance, because the light path is out and back. If a pulse

---

Fig. 4 3-D flash LIDAR shown penetrating through 150 of dust. Visible image is completely obscured while Flash LIDAR can provide 3-D imagery of hazards.

Fig. 5 Diagram of the LIMARS polarization based 3-D flash ladar concept.

Optical Engineering 060901-7 June 2012/Vol. 51(6)
is used as a time reference then a specific trigger level on the rising edge of the pulse or the peak of the pulse can be used. Range resolution is related to the ability to separate two objects in range that are in the same angle-angle bin. Range precision is related to the ability to measure a change in the range of a single-range object. Range accuracy is how accurately the absolute range to the object can be measured. Range precision depends on the signal to noise ratio and is typically better than range resolution. Short pulses are not the only method of measuring the time of flight of light to the target and back. A frequency chirp can be used. Usually an FM chirp will be a saw tooth waveform. A pseudo random code can be used, jumping in frequency, or phase, or amplitude, although amplitude is more difficult to measure due to noise. One over the bandwidth of the waveform is the equivalent to pulse width in a pulsed Ladar.

5 Laser Vibration Detection

Coherent ladar systems have the capability to perform remote sensing with high sensitivity Doppler (velocity) information. As the object moves towards the ladar the frequency of the reflected light will shift upwards due to the Doppler effect, and it will shift downward as the object moves away. Figure 8 shows a transformer, and a ladar vibration measurement showing the 60 Hz vibration of the transformer, that results from 60 Hz AC power running through it. Of course, time-of-flight can be used to determine the distance to the target, which allows constructing a 3-D image of the target. Therefore, a ladar can display a 3-D image of the target and its vibrational modes simultaneously. Figure 9 then shows the time return from a tank behind a tree. You can separate the returns in time, and then analyze the frequency return at each voxel (i.e., range resolved pixels) to obtain a 3-D vibrational spectrum image of the target. Figure 9 is an example excerpted from a 2002 Air Force data collection. A pulse waveform coherent ladar was utilized that allowed range resolved measurements of target vibration. The pulsed coherent ladar waveform used is described in Ref. The 10 ns pulse duration results in a range resolution of 1.5 m. Range resolution is important as it allows the separation of clutter from actual target returns and also improves the identification of vibrational hot spots on the targets. For the data shown in Fig. 9, a running vehicle was placed behind a tree. A cw waveform ladar without range resolution would not be able to reliably detect the vibrating target behind the tree. The pulsed waveform ladar, consisting of coherent 10 ns pulses at a PRF of 1 kHz, is able to range separate the signals scattered from the tree from those scattered from the target. The top panel in Fig. 9 shows the raw temporal heterodyne signal from the tree plus vehicle. By range gating this signal the returns from the tree are separated from those from the vehicle. Precision measurement of the phase shifts between pulses separated by one millisecond allows precision velocity measurements with time. The resulting velocity versus time data can then be spectrally analyzed to show vibrational features. The lower panel in Fig. 9 provides example spectra obtained from the tree (left panel) and the tank (right panel). Note that the tree return has no discernable vibrational tones above the noise floor, whereas, the tank has vibrational features detected at 30, 60 Hz, and other harmonics of 30 Hz. The vertical scale of the spectral plots in the lower panel of Fig. 9 is logarithmic – the vibrational tone at 30 Hz is about a factor of 6 above the mean noise floor. The maximum frequency shown in the plots of 500 Hz (horizontal scale) is the Nyquist
The vibrational image of an internal combustion engine can be used to identify combustion pressure pulses and inertial acceleration of the pistons and drive trains. It can also help identify mechanical imbalances and misfires. Vibrometry can also be used to identify hidden faults in a structure, such as cracks or delamination. Using an external vibrational excitation, internal defects can be identified by the reflections and scattering in the waveforms at the defect sites. An aircraft or tank can be identified by the vibrational signature it emits. A diesel engine can be easily distinguished from a turbine engine. Most velocities due to structural vibrations are in the range of 1 μm/s to 1 mm/s. The Doppler frequency shift on a laser return is

\[ \Delta \nu = \frac{2 \cdot V}{\lambda}, \]  

(21)

where \( V \) is the object velocity, and \( \lambda \) is the wavelength. Assuming a laser wavelength of 1550 nm, the magnitude of Doppler frequency shifts that need to be detected for various target velocities are shown in Table 1. There can also be gross object velocity due to target and/or source movement. Additionally, the local oscillator can be offset from the outgoing laser frequency. Both of those factors increase the frequency of the received beat frequency signal. A 10 m/s gross velocity (36 km/h) results in a 12.9 MHz frequency shift. If the sensor and the object are stationary there will be no gross Doppler shift. For an aircraft flying over an area however the gross aircraft velocity needs to be taken into account. In addition the angle from one side of the area to the other will result in a different Doppler shift since it is only velocity directly toward or away from the laser that contributes to the Doppler shift.

### 6 Synthetic Aperture Ladar

Synthetic aperture ladar uses the motion of the ladar to develop an effective aperture larger than the size of the real aperture. In concept this is simple. At a given location the ladar emits a pulse, and measures the returned signal from the target. If the return field can be measured at many different locations then those fields can be stitched together in the pupil plane to develop a larger pupil plane image. A real beam aperture just collects a pupil plane image. This larger pupil plane image can be Fourier transformed to obtain a high resolution image. The difficulty with this approach comes in the implementation. You need to precisely measure, or estimate, the fields in each location, and then add them. If you do not know the position of the receiver aperture exactly then you may place a pupil plane field measurement in the wrong location. It is not surprising that synthetic aperture imagers were developed first at microwave frequencies, where the wavelength might be 3 cm, such as with 10 GHz synthetic aperture radar. It is easier at longer wavelengths to align pupil plane images to a fraction of a wavelength. In addition, the real aperture based resolution of ladar is very good, so the need for synthetic aperture ladar was less. Synthetic aperture radar, SAR, has existed for decades. In SAR before you can add the various segments to obtain a large pupil plane image that can be Fourier transformed to obtain a high resolution image you need to transform all the image segments to the same reference system. Usually in SAR the point of closest approach called the phase center, is used as zero for the reference system. Because of the reference frame adjustment, and more fundamentally because both the transmitter and the receiver move, the resolution in the track direction for a SAR is given by:

\[ \Delta \theta = \frac{\lambda}{2L}. \]  

(22)

where \( L \) is the distance flown. This is twice the resolution of a real aperture with the dimension, \( L \). This equation neglects the size of the real aperture, assuming that the real aperture is much smaller than the synthetic aperture baseline, \( L \).

SAL has been demonstrated in the lab and in flight over the last 5 years or so. Some early, more limited, demonstrations were also performed. A very early SAL image is shown in Fig. 10. According to the article this is first true SAL image. On the left is the raw image. It is dominated by speckle. The circle overlaid on this image shows the approximate size of the resolution element that would result from the system’s physical aperture alone. The resolution actually obtained is better by a factor of about 50. The image on the right shows the result of filtering the image to reduce the effect of speckle.

Figure 11 shows another early SAL image, somewhat more recent than the one in Fig. 10. The real-aperture

<table>
<thead>
<tr>
<th>Velocity (μm/s)</th>
<th>Frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.29</td>
</tr>
<tr>
<td>10</td>
<td>12.9</td>
</tr>
<tr>
<td>100</td>
<td>129.03</td>
</tr>
<tr>
<td>1000</td>
<td>1290.32</td>
</tr>
</tbody>
</table>
diffraction-limited illuminating spot size is represented at the right. A picture of the target is shown at the left. A recent article talked about a SAL flight demonstration. An image from that flight is shown in Fig. 12.

Phased Array of Phased Array Based Ladar

7.1 Spatial Heterodyne

Spatial heterodyne captures a pupil plane, or image plane, image in each sub-aperture. An off axis local oscillator is used in each sub-aperture, along with low bandwidth imaging detector array, as shown in Fig. 13. Each pupil plane image contains both image information and spatial phase variation.

A pupil plane image from each detector is Fourier transformed into the image plane and then sharpened, using a cost function. For example we can look at the sum of all pixels squared. This is a traditional cost function, although due to speckle issues we expect to use a lower exponent, such as to the 1.2 power. In the presence of speckle the higher cost function tends to create artificial bright points in the image. In order to sharpen each pupil plane image each sub-aperture must have sufficient signal to noise. The sharpened pupil plane images from each sub-aperture are transformed back into the pupil plane and used to assemble a more complete pupil plane image. The phase distortion in each sub-aperture image, and between sub-aperture pupil plane images, can be judged based upon the difference between the captured image and the sharpened image.

This phase distortion can be used to place the sub-aperture pupil plane images in the correct locations. Real geometry can aid in this placement, as a reality check. Piston phase information between sub-apertures must be estimated, also using a sharpness cost function. An image, such as shown in Fig. 14, is generated by Fourier transforming the phase adjusted mosaic of pupil plane images from the individual sub-aperture pupil plane images.

There is a desire to have an approach that is scalable to a large number of sub-apertures. The speed of closure on the “best” high resolution image as a function of the number of sub-apertures needs to be investigated. The addition of a high bandwidth detector in each sub-aperture may allow scaling to a larger numbers of sub-apertures by measuring the piston phase shift between sub-apertures. This could eliminate much of the processing time associated with using a second sharpness metric to estimate piston error between sub-apertures. The main issue with using a high temporal bandwidth detector to measure piston is the difficulty in obtaining a common path to measure. The mosaic image in the pupil plane is based upon the low bandwidth spatial heterodyne detectors.

A good example of the power of spatial heterodyne can be shown in a recent paper by Tippie, from Fineup’s group. Figure 15 is an extracted portion of the figures from that paper. It is obvious a huge gain in resolution is made comparing the set of bottom images to the single sub-aperture top images in Fig. 15.

The first row is images she took using a single sub-aperture. The last row is images she created after processing, using many sub-aperture receiver positions. You can see the dramatic resolution enhancement. More detail is of course available in the reference.
7.2 Flash Aperture Synthesis

Flash aperture synthesis uses multiple coded transmitters, along with multiple receive apertures, to obtain resolution approaching a factor of two better than an aperture array that does not use multiple spatially separated transmitters, without needing aperture motion. As stated in the section on synthetic aperture ladar, microwave SAR radar has since its inception taken advantage of transmitter spatial diversity (in that case movement) to gain a factor of two increase in resolution compared to just using receiver spatial diversity. This has recently been demonstrated in the optical regime by mechanically switching from one transmit location to the next. RF MIMO techniques that use multiple simultaneous phase centers have been developed. The increase in effective aperture diameter $D_{\text{eff}}$ for a synthetic aperture imaging ladar, with the real aperture large enough to be taken into account, is shown in Eq. (23):72

$$D_{\text{eff}} = 2D_{\text{syn}} + D_{\text{real}},$$

where $D_{\text{real}}$ is the real sub-aperture diameter. An experiment was designed to verify the equations derived in Ref. 73. For a discrete set of sub-apertures in one dimension the resolution will increase by a factor of $2N + 1$, where $N$ is the number of sub-apertures. In Fig. 16 an array of static transmit/receive sub-apertures illustrates the distinction between beam resolution and image resolution. In order to obtain the increase in resolution from transmitter diversity, multiple sub-apertures must transmit a unique signal. All pupil plane fields resulting from illumination with different transmitters need to have a common reference before the image is reconstructed.

Figure 17 is extracted from a paper by Rabb et al., at AFRL.70

8 Advanced Object Discrimination Using Ladar

The paper started by highlighting the many diverse discriminates available using ladar. Some of those ladar modes have been discussed in moderate detail, but additional discriminates are available.74 Polarization can for example strengthen the ability to discriminate one object from another, or to detect an object.75–78 This was mentioned in the LIMARs flash imaging section, but this discriminate is available essentially whenever you are willing to double the number of detectors, or focal plane arrays, and to add the required optics to separate the polarizations. Wavelength diversity is available if you are willing to provide laser illuminators at multiple wavelengths, and the associated detectors.79–84

![Fig. 15](https://astronomicaltelescopes.spiedigitallibrary.org/journals/Optical-Engineering) Comparison of the single sub-aperture image (top panel) versus a processed multiple sub-aperture spatial heterodyne image (bottom panel).

![Fig. 16](https://astronomicaltelescopes.spiedigitallibrary.org/journals/Optical-Engineering) Difference between real beam resolution and flash aperture synthesis resolution.

![Fig. 17](https://astronomicaltelescopes.spiedigitallibrary.org/journals/Optical-Engineering) An image formed from (a) 36 averages of an unsharpened, single-aperture image of the quarter and (b) 36 averages of a sharpened single aperture, (c) 12 averages of a synthesized aperture from the three physical apertures and a single transmitter, and lastly an image (d) is shown which is created from the set of 36 field values synthesized into a single, large field.
An additional way you can provide enhanced angular resolution can be obtained by using the Doppler shift across the beam. This is called Range Doppler imaging. The return speckle field from an object can say big how the object is. Other object features can be obtained by analysis of the speckle field.

9 Conclusions and Summary

Ladar is a rich phenomenology that is poised to become much more widespread because required components are becoming more widely available. It can have many modes, many of which are described in this paper. It can also have a wide variety of applications.

Acknowledgments

The author would like to thank Dr. Sammy Henderson for a very thorough review, along with excellent suggestions, which has made this a more valuable article.

References

McManamon: Review of ladar: a historic, yet emerging, sensor technology


Paul F. McManamon owns Exciting Technology LLC and is the technical director of LOCI at the University of Dayton. Until May of 2008 he was chief scientist of the Sensors Directorate, AFRL. He has participated in three Air Force Scientific Advisory Board studies. He was instrumental in the development of laser flash imaging to enhance EO target recognition range by a factor of 4 or 5. McManamon was the 2006 president of SPIE. He was on the SPIE board of directors for 7 years and on the SPIE executive committee for 4 years. McManamon received the WRG Baker award from the IEEE in 1998, for the best paper in any refereed IEEE journal or publication. McManamon is a fellow of SPIE, IEEE, and OSA. AFRL, and MSS. He was vice chairman of the NAS study called “Seeing Photons.” He is co-chair of the NAS “Harnessing Light 2” study.