Canesta 101

Introduction to 3D Vision in CMOS

This paper provides an introduction to the Canesta technology. The paper seeks to be less technical, and in some cases trades absolute accuracy for ease of reading, to address a wider audience.

March, 2008

Table of Contents

Basic Topics

Canesta has a pixel-based, depth vision system. The pixel based, depth vision system is fabricated in CMOS and shares some features with traditional CMOS camera sensors. But in many ways, the Canesta sensor is very different. The Canesta pixel-based, vision sensor features:

- **Active light** - An integral part of the Canesta solution is an LED light source. We typically use invisible, infrared light. The active light source flashes very rapidly, in our example at 44 MHz or 44 million times per second, and only when the pixel is active.
- **Unique pixel structure –** meaning that the Canesta pixel is very different from a traditional CMOS camera chip. We sometimes call the Canesta pixel a "magic pixel," as it is so much more advanced than traditional color pixels. The Canesta pixel has more than one receptor, it features very precise timing, and it features an accumulator function. In some cases, the pixel has additional circuitry to prevent saturation, such as might occur in direct sunlight.

Conceptually like PinPressions

At each pixel, the Canesta sensor measures the distance from the camera to the element in the scene. One model we use to illustrate the concept of this system is an artistic piece known as PinPressions.

Figure 1 - PinPressions

The image above shows a hand. The hand is a literal impression into the “bed of pins” that make up the PinPressions. And each pin is conceptually like a pixel on a Canesta sensor, measuring distance.

Basic Structure of the Sensor

The best way to explore the Canesta technology is to conceptually power up a real sensor and follow its operation. For this example, we are using Canesta’s industrial sensor, a 120 x 160 pixel sensor.

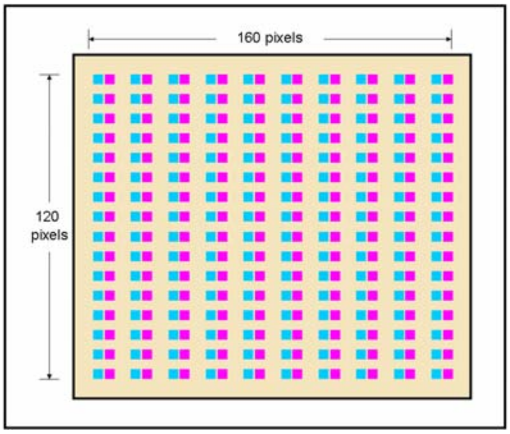


Figure 2 - Industrial Sensor

Each pixel contains two receptors, an In-Phase receptor and an Out-of-Phase receptor. By In-Phase, we mean that the receptor is active exactly when the LED light source is illuminated.

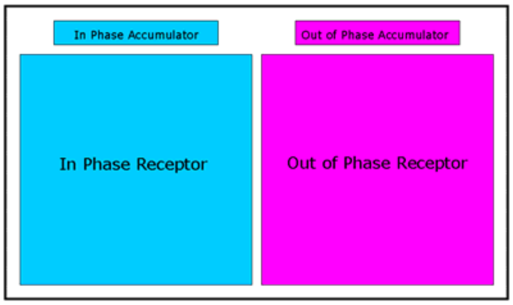


Figure 3 - Pixel Structure

At a simplistic level, if more light is absorbed by the In-Phase receptor, than the Out-of-Phase receptor then the object is closer. We will explore the physics of the system in more detail shortly.

Capturing a Frame

As we start up the sensor, we select a frame rate, the number of times per second that we capture and output an image. For this example, we choose 60 (FPS) frames per second . At 60 FPS, a single frame takes 17 milliseconds.

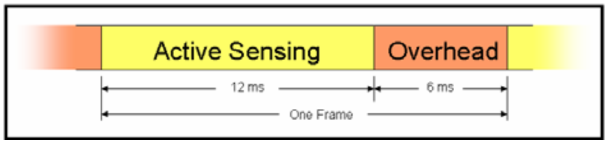


Figure 4 - Structure of a Frame

A frame is comprised of a period of active sensing, sometimes referred to as the “shutter being open” followed by an overhead period, used for sequentially reading all of the accumulators and then zeroing out all of the accumulators.

Figure 5 - Light Pulse

To start the frame, the accumulators are zeroed and the pixels go active for 12 milliseconds. During this active period of 12 ms, the LED light source is oscillated at 44 Mhz, delivering 528,000 light pulses.

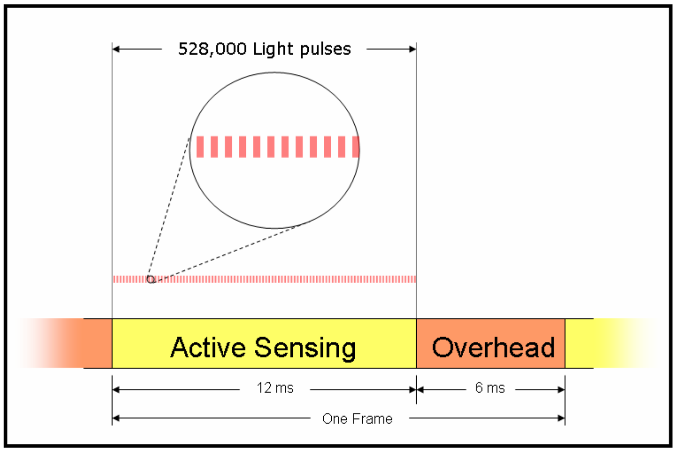


Figure 6 - 528,000 Light Pulses per Frame

Each light pulse has the light on for exactly the same amount of time that it is off, and the transition from off-to-on and from on-to-off is almost instantaneous. The sensor is timed to operate in exact sync with the light source. The first receptor, the In-Phase receptor, goes active in exact sync with the LED pulse. The second receptor, the Out-of – Phase receptor, goes active precisely as the first receptor turns off and is directly out of sync with the first receptor and directly out of sync with the LED.

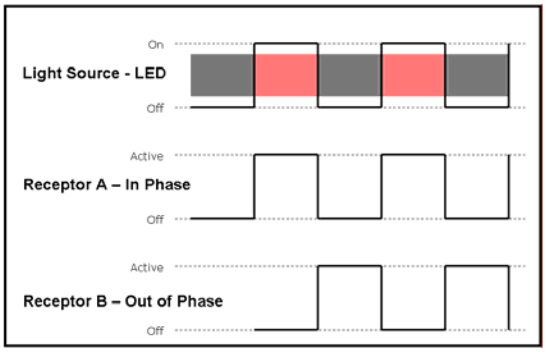
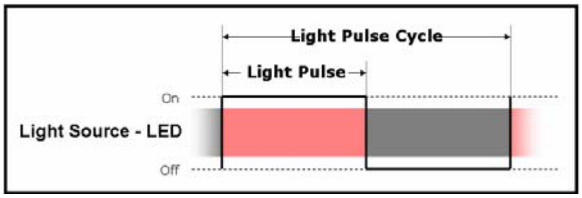


Figure 7 - Sensor Timing

The light pulse projects forward into the scene, hitting objects and part of that light reflects back to the camera. The light reflected back to the camera hits the pixels. Based upon the timing of the reflected light pulse, part of the light pulse is absorbed by the In-Phase receptor and part of the light pulse is absorbed by the Out-of-Phase receptor. The charge received in the In-Phase receptor is added to its accumulator, likewise for the Out-of-Phase receptor and accumulator. These very short light pulses, at 44Mhz or 44 million pulses per second, are about 11 feet long, and are ideal for ranging objects between 0 and 11 feet away. This light pulse is repeated 528,000 times during the next 12 ms, the active period. At the end of the active period, the sensor polls each receptor for its two accumulated values.

Measuring Distance

At Canesta, we typically talk about timing. But, we understand that time is interchangeable with distance when we are discussing light.



For example, we say that the light is oscillating at 44 Mhz, which means that one light pulse cycle lasts 1/44,000,000 of a second. Each light pulse lasts for ½ light pulse cycle, which is 1/88,000,000 of a second. Since the speed of light is a constant, 186,000 miles per second or 980,000,000 feet per second, in 1/88,000,000 of a second, light travels 11.18 feet. So our light pulses are of both a fixed duration in time and a fixed length in space. The light pulse is 11.18 feet in length and the full light pulse cycle (light pulse plus and equal length of no light) is 22.36 feet.

A typical scene has objects in the scene within our range of interest. The range of interest, in our base case, is within the length of the pulse of light, for reasons that will become clear.

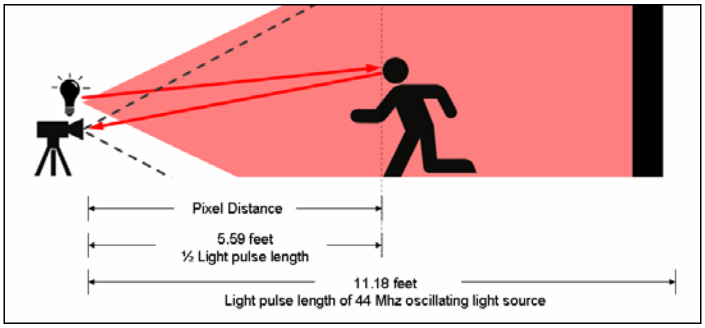


Figure 8 - Typical Scene

Object at Camera

A boundary condition is perhaps the best place to start before examining a real situation. Let's say that the object in the scene is right next to the camera. Even though we show the light traveling a very short distance, we will imagine the distance as 0.

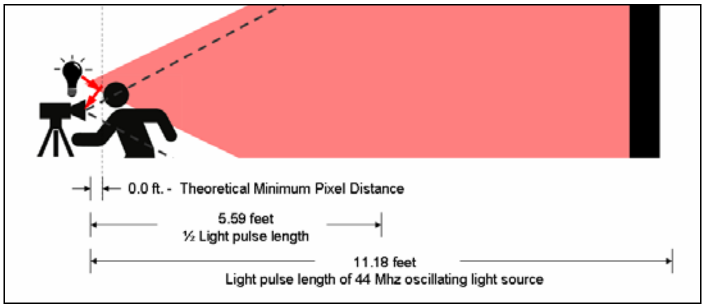


Figure 9 - Boundary Condition, Object at Camera

If we imagine the true boundary condition, the object is exactly at the camera. The light will be reflected back to the camera instantaneously. So, the timing of the reflected light will be exactly the same as the timing of the light source. Since the In-Phase receptor is active exactly at the same time as the light source, then as a percentage, 100%

of the reflected light goes into the In-Phase receptor and none of the light goes into the Out-of-Phase receptor.

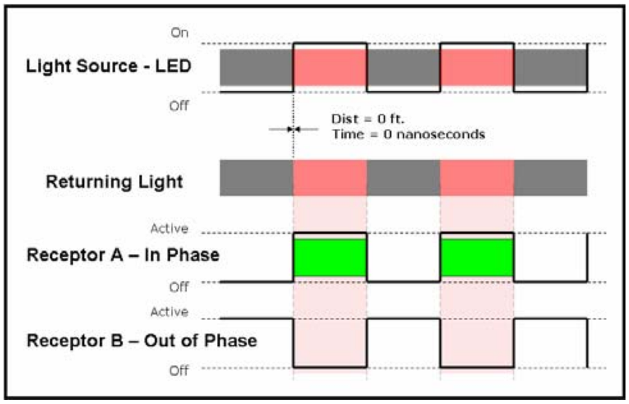


Figure 10 – Timing, Object at Camera

Object One-half Light-pulse-length Away

Next we look at an object exactly at the mid-point in our Range-of-Interest, or 5.5 feet away.

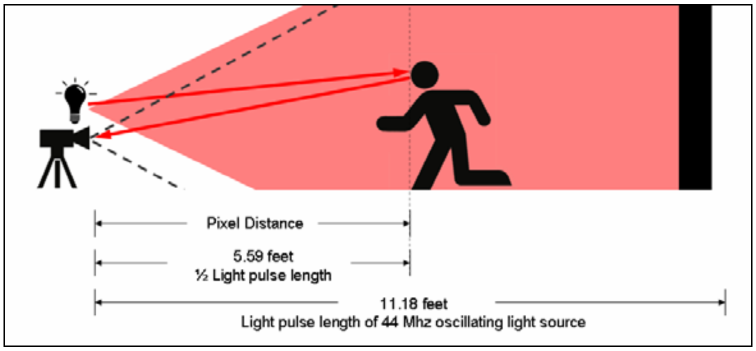


Figure 11 – One half light pulse

The timing of the reflected light, since it has traveled 11.18 feet total (5.59 feet out, and 5.59 feet back for a round trip distance of 11.18 feet), ends up being exactly out of phase with the light source, and exactly in phase with the Out-of-Phase receptor.

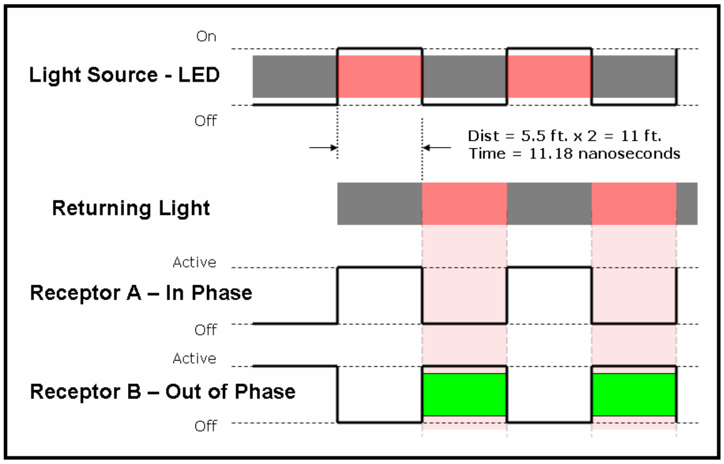


Figure 12 – Timing of Distance = ½ Light Pulse

Typical Example

For a situation where the object of interest is within the camera's range of interest, the actual range is easily computed as a difference between the two receptors.

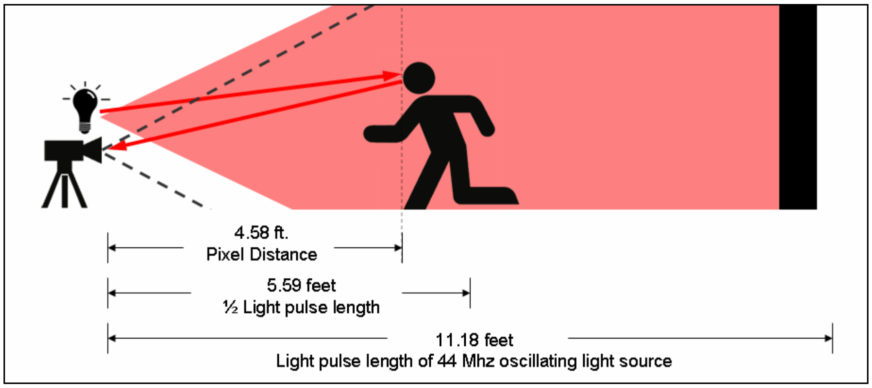


Figure 13- Diagram of Typical Example

In the situation we have chosen, the object is 4.58 feet from the camera or 5/12 of the range of interest which is 11.18 feet. As we look at the timing of the reflected light, we see that part of the light, in fact 1/6 of the light, has returned when the In-Phase receptor is active and 5/6 of the light has returned when the Out-of-Phase receptor is active.

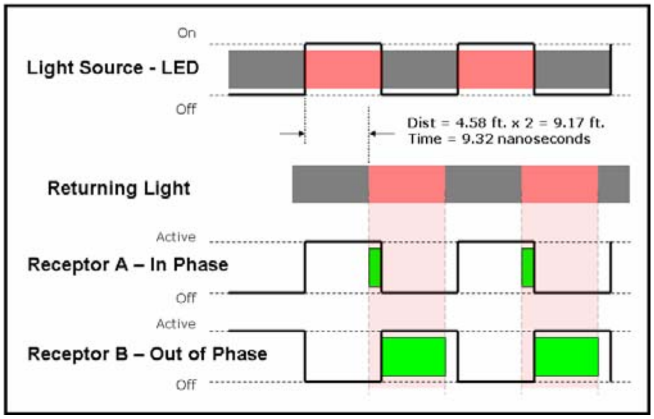


Figure 14 - Timing of Light Reflecting to Receptors

The difference in value between the In-Phase and Out-of-Phase receptors correlates to distance.

Advanced Topics

Sources of Error

One interesting way to understand our technology is to think through the sources of error and the manner in which we deal with each of those sources of error.

Characteristics of the Light Source

In our theoretical model, we assume that the light source turns on instantly and off instantly. We use the term Modulate to mean turn from on to off and from off to on. In practice, the light source does not modulate instantaneously, and instead takes some time for the signal to rise and fall. The edge sharpness of the light source signal contributes to a term called modulation contrast that determines the ability of the camera to measure the distances with more precision

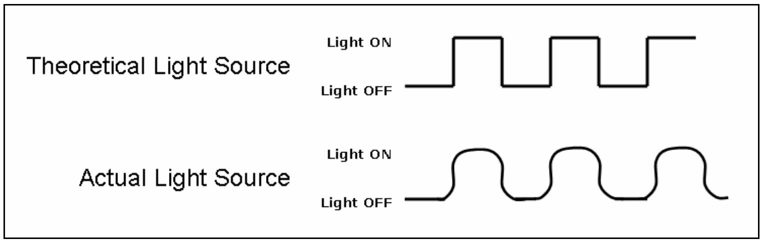


Figure 15 - Modulation Contrast

We spend significant effort to find the best LEDs and to design the light source driver so as to produce the best Modulation Contrast and to minimize this source of error.

Problem - Distance Ambiguity

In practice, more than one distance can deliver the same "difference" between the In-Phase receptor and the Out-of-Phase receptor. Our first distance ambiguity occurs as we cannot tell the difference between objects that are exactly the same distance either closer or father than ½ the light pulse distance. Remember that for objects exactly ½ light pulse distance away, all of the light pulse goes into the Out-of-Phase receptor. For every situation where some small percentage goes into the In-Phase receptor, we cannot tell if that small percentage means that the object is slightly closer, or slightly farther away.

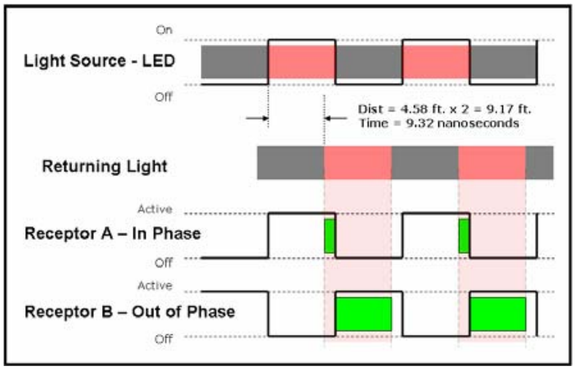


Figure 16 - Closer Ambiguous Distance

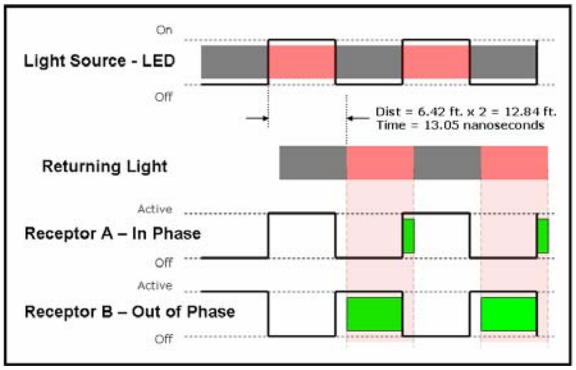


Figure 17 - Farther Ambiguous Distance

In the example above, 1/6 of the light hits the In-Phase receptor, and 5/6 hits the Out-of-Phase receptor. So the difference, A-B, is the same even though the distances are different.

Solution - 90 degree Phase Change

This ambiguity is easy to resolve, by taking a second measurement. We have always talked about Receptor A as exactly In-Phase with the light source. We can also change the phase of the receptor in increments of 90 degrees. By taking the same reading 90 degrees out of phase with the original, we disambiguate between these two situations.

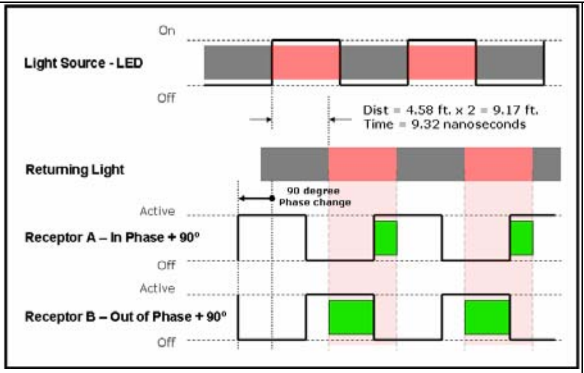


Figure 18 - 90° Phase Change, Closer

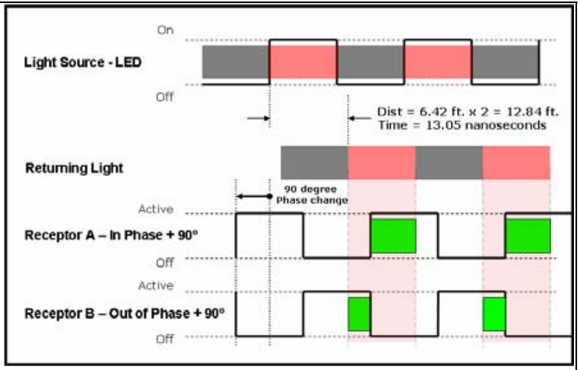


Figure 19 - 90° Phase Change, Farther

For clarity, we use the exact same situation as was ambiguous above. For the slightly closer example, when the phase of the receptor is changed by 90 degrees, the In-Phase receptor gets 2/3 of the light. For the slightly farther example, the In-Phase receptor gets 1/3 of the light. As such, this second measurement resolves the ambiguity identified above.

Note: we are using this simplistic distance ambiguity as the way to introduce the phase change timing of our sensor. This simplistic example, rather than posing a real problem, is the way we are introducing the phase change capability of our sensor, which is essential to our advanced distance measurement algorithms.

Problem: Distance Ambiguity – Wrap around Problem

Another ambiguity occurs for objects which are father than one light pulse away. In our basic, theoretical system, we cannot tell the difference between distance X and distance X + 1 light pulse. In both cases the difference between the two receptors is identical. The light travels farther, the brightness is lower, but our key measurement, the difference between receptor A and receptor B, is identical.

Solution: FarSight™

Canesta FarSight™ solves the wrap around distance ambiguity problem by taking a second distance reading, where the entire system, both light source LEDs and receptors are modulated at another frequency. The second modulation frequency should not be a multiple of the first frequency, such that the distance ambiguities do not match up or overlap. Combining the two independent distance readings, each with potential wrap around ambiguities, leads to a single, unambiguous distance reading.

Problem: Pixel Saturation

In some cases too much light enters the pixel. If one of the receptors gets too much light, it will not be able to accumulate any more and the difference between the two receptors is corrupted. Canesta has designed sensors to work on a car dashboard or car tailgate, in direct sunlight, where pixel saturation is a very real potential problem.

Solution: SunShield™

On a regular basis, the Canesta pixel reduces the charges in both receptors by an equal amount to eliminate the effect of the ambient light. This method preserves the charges that are used to determine the depth while preventing the pixel saturation due to strong sun light. In practice, our SunShield™ works so well that our distance measurements are just as accurate in direct sunlight, indoor lighting and in complete darkness.

Problem: Clothing Reflectivity

Most clothing is good at reflecting infrared light. Most clothing reflects 50% or more of the light that hits it. Occasionally, we will find a cloth that seems to absorb almost all light that hits it. As an example, we recently found a black Rayon fabric with a reflectivity of 3%. To get an accurate depth image of a person wearing that fabric, we need to image reflected light, which means we need to push more light at the person, get the person closer to our sensor, or increase the shutter time of the sensor.

Solution: Next Generation LEDs

The next generation of LEDs will help with this problem. We are seeing a great increase in light output per LED and per unit power. In the last two years, we have seen a 10X performance improvement. While this “better than Moore’s law” performance improvement may not go on forever, it has certainly benefited us for the last two years and we are hoping for more in the coming years.

Design Considerations

One of the best ways to understand the Canesta technology is to think through the design of a camera. We have a series of tradeoffs that we consider in each camera design.

Length of the Light Pulse

How does Canesta choose the length of a light pulse? Why 44Mhz, which is a pulse length of 11 feet?

First, one interesting argument for a longer light pulse is that we could reduce our ambiguity, especially the wrap around problem. Let's say we had a light pulse 300 feet long, then in any normal indoor space meaning a space less than 50 feet, there is no wrap around ambiguity. But, we would give up a lot of precision.

The precision of the Canesta distance measurement is related to light pulse length. Shorter light pulses produce more precise distance measurements, while longer light pulses have greater errors. Our 11 foot light pulses resolve to millimeter precision within the first 2 feet. The precision resolves to centimeter accuracy at 8 feet.

We are conscious of the tradeoffs between wrap around ambiguity and distance precision. We might choose a longer or shorter light pulse for a particular application.

Light Wavelength

The Canesta sensor is most sensitive, and most efficient with visible red light, at a wavelength of 658 nm. Most applications specify the use of invisible light, so we use a slightly longer wavelength of infrared light. Our sensor is less sensitive to infrared light, so we need to increase the output power of the light source to compensate. As an example, our sensor is less sensitive to the light produced by our current LEDs at 870 nm. In fact, we have measured its efficiency relative to a baseline at visible red light of 658 nm, and we see a 25% efficiency. This means that we need to drive 4 times as much 870 nm infrared light at the target to get the same results. If we can find applications that would value the use of visible light, we could get away with much lower powered red LEDs. Of note, we are seeing a greater selection of LEDs emerge, and anticipate using wavelengths closer to visible light, which are more efficient, in the near future.

Light Power

Light power is the same as brightness. A light source with more light power is brighter than a light source with less light power. The more light output power (lumens) from our light source that reaches the sensor, the better our pixel performs. So we like to use a very bright LED or more than one LED to make it brighter. Fortunately for us, the makers of LEDs are outperforming Moore's law and we are seeing about a 10x increase in output light power from LEDs of the same input power over the last two years. In consumer webcam and game console accessory applications, we are conscious of the requirement to only use USB power. So we try to limit the power used by the light source, and the overall camera, to make it within the power budget of USB power.

Field of View

We have built a variety of cameras with a variety of fields of view. For a gesture interface to a PC, we typically like a wider field of view, like 90 to 110 degrees. For family room activities, we have settled on a 70 degree field of view, appropriate to imaging a 6 foot adult body at 6 feet. For living room gesture activities from a 10 foot couch, we like a narrower field of view, like 40 to 50 degrees. This narrower field of view presents some challenges especially in terms of pointing the camera.

Diffuser

As we try a variety of LEDs and a variety of fields of view, we run into a problem of the LED light source not illuminating the scene evenly. A relatively simple solution is to put a diffuser on top of the LED light source that is designed for the particular field of view. The diffuser provides an even light distribution over a predefined shape.

Pixel Count

Our current sensor features an array of 160 x 120 pixels. Some applications would benefit from more pixels. The cost of our sensor goes up with pixel count. We are considering a 320x240 sensor. The silicon area, which is directly related to cost, goes up by a factor of 4. We constantly weigh the tradeoffs between cost and resolution.

Some of the leading volume applications, like video game publishers, and PC webcam manufacturers seek very low costs, which keeps us from increasing the pixel count.

Glossary

Diffuser - A diffuser is an optical component which converts light from one shape/pattern to another shape/pattern. A diffuser is typically made from glass or plastic. We use diffusers to take a narrow beam of light from an LED and change that beam into a even pattern of light across a specific field of view.

Distance Ambiguity - Two different distances, which our sensor have difficulty distinguishing, result in distance ambiguity. In each case of distance ambiguity, we have advanced algorithms which resolve the ambiguity.

FarSight™ – FarSight™ is a Canesta algorithm for sensing objects outside of the Range of Interest. At a high level, the algorithm uses two different frequencies to range the object. In the case of an object outside of the Range of Interest, the second frequency is used to locate any objects beyond the range of Interest.

Frame – A frame is another name for a single image.

Frame Rate – The number of images captured within a given time period. Typically the frame rate is expressed as a number of frames per second. The period of time the sensor is active for each frame is a function of the frame rate and the time for overhead in each frame.

Field-of-View – The field of view is the area which a camera can see. The area is shaped as a pyramid, on its side. Field of view is measured in degrees, either a single figure, which is a diagonal field of view, or else as two figures, the horizontal and vertical fields of view.

In-Phase Receptor – This receptor is in phase with the light source meaning that it is actively absorbing light energy at the same time that the light source is on. The In-Phase receptor absorbs most of the light energy for objects that are very close to the camera and light source.

LED – Light Emitting Diode. LEDs are electronic components that produce light. They are typically inexpensive in terms of both cost and power. LEDs are generally very efficient, producing a relatively lot of light per unit power.

Light Pulse - A single burst of light. Canesta uses very short and exactly timed light pulses, coordinated with precisely timed sensors to measure distance.

Light Pulse Cycle – A light pulse cycle is a period of illumination, the light pulse, and a period of darkness where the light source is off. The current Canesta solution uses light pulses and and periods of darkness that are exactly the same length.

Modulate – We use the term modulate as in “modulate the LED.” We arereferring to turning the LED on and off. We also use the term oscillate

Modulation Contrast – The Modulation Contrast is measure of the sensor's quality to resolve distances. The rapid rise and fall of the light source signal improves the modulation contrast.

Oscillate - We use the term oscillate to describe turning the LED on and off. Circuitry in the sensor is oscillated at the same frequency, to synchronize the sensor with the light source.

Pixel – A pixel is literally derived from the term “Picture Element” and is the smallest logical unit of a digital camera. A Canesta pixel is composed of two receptors, two accumulators, and a variety of other circuitry for additional functions like SunShield™, a technology for reducing saturation from exposure to direct sunlight. A Canesta pixel is significantly different from a traditional CMOS RGB camera pixel.

Out-of-Phase Receptor – The receptor which is active during the time that the LED is not illuminated.

Range of Interest – The range of interest is a range of depth values sensed by a given Canesta technology. The Range of Interest is typically modified by changing the frequency of both the light source and the receptors. Canesta has a technology for extending the Range of Interest, call FarSight, that uses more than one frequency of light.

Receptor – A receptor is a part of a pixel. Each pixel has two receptors: an In-Phase receptor and an Out-of-Phase receptor. The In-Phase receptor is in phase with the light source.

Reflectivity – Reflectivity is a measure of the percentage of light that is bounces off a surface. A white, matte surface can have a reflectivity close to 100% A black, matte surface can have a reflectivity below 10%. The Canesta sensor measures reflected light. Surfaces with low reflectivity require greater illumination. Note that the reflectivity of a surface cannot only be judged by the color of the material. In the case of clothing, fabric, weave and color all play a role.

Sensor – We use the term Sensor interchangeably with computer chip. The Canesta sensors are packaged as single silicon chips. The current sensor is a 160 x 120 pixel ship.

SunShield™ – SunShield™ is a Canesta technology to reduce or eliminate pixel saturation. Periodically both pixel receptors are decremented the same amount. This simultaneous decrement operation preserves the difference, while ensuring that the pixel does not saturate due to strong ambient light.

Wavelength –	The wavelength of light describes both the color of the light and the visibility. For example, light with a wavelength of 650 nm is bright red. Light with a wavelength of 810 is invisible, infrared light.
Webcam –	A camera designed to be connected to a PC for projecting live images or video over the Internet. The webcam is typically powered via the USB port and is typically constrained to the bandwidth of the USB port.