Characterization of Time-of-Flight Cameras for use in Diffuse Optical Tomography

Jeremy S. Johnson REU program, College of William and Mary
August 30, 2016

Abstract

Diffuse Optical Tomography (DOT) is a noninvasive, portable technology used for imaging through tissue. It works by using near-infrared light emitted from an array of sources to illuminate tissue and observing the scattered light that emerges with an array of detectors. Using models to approximate unknown scattering and absorption coefficients, one can extract blurred images of structures within the tissue. The three primary absorbers of the near-infrared light are water and both oxygenated and de-oxygenated hemoglobin. Focusing on the two types of hemoglobin and using different wavelengths of light, we can attempt to localize the scattering and absorption of either type to look for areas of greater blood supply or to map the concentration of oxygen in the blood. Today, the most widely used applications of DOT are detecting tumors in the breast and imaging the brain. Our research aims to improve on existing technology by implementing time-of-flight cameras which will introduce far more detectors and, therefore, increase the resolution of images. Work done so far has been in characterizing possible time-of-flight cameras.

Supported by NSF REU grant PHY-1359364

1 Introduction

1.1 Motivation

Time-of-Flight (ToF) cameras are an emerging technology that produce depth maps of an imaged scene by measuring the phase-shift of reflected, near-infrared light. Every pixel on the camera is able to detect a phase-shift, allowing each to represent a point in the scene. These cameras are being developed and tested for computer vision, gesture recognition, and 3D reconstruction (Hansard et al., 2012). The advantage of using ToF cameras for DOT is the greater number of detectors per camera, resulting in greater resolution of reconstructed images. The cameras do, however, exhibit systematic errors that must be resolved before attempting to use them in DOT.

1.2 Theory of Operation

Time-of-Flight cameras determine the distance to an object by illuminating a scene with modulated, near-infrared light and observing the phase shift of the returning signal, figure 1. Light reaching the array of detectors generates an electron current via the photoelectric effect and the resulting voltage is analyzed for each detector (Hansard et al., 2012). The specific method used to calculate light intensity and phase shift are different from camera to camera, but, in general, the phase
information is contained within the modulated portion of the signal so that ambient light can be largely ignored. The range of the device is restricted by the modulation frequency used, since, for example, a phase shift of 30 degrees and 390 degrees would appear identical.

2 CamBoard Pico

2.1 Characterization of Errors

The CamBoard Pico operates by emitting modulated near-infrared light in pulses such that there are four out of phase signals per frame, or image. The length of each pulse can be controlled but is generally 1ms. Each pixel on the camera lens samples the four pulses, from which the phase shift, intensity, and distance are calculated as:

\[ \phi = \omega \tau = \arctan \left( \frac{S_3 - S_1}{S_4 - S_2} \right) \]  

\[ A = \frac{1}{2} \sqrt{(S_3 - S_1)^2 + (S_4 - S_2)^2} \]  

\[ d = \frac{1}{2} c \tau = \frac{c \phi}{2 \omega} \]

where \( S_i \) is the signal corresponding to the \( i \)-th pulse, \( \omega \) is the modulation frequency of the signal, and \( \tau \) is the time the light takes to reach an object and reflect back onto the camera lens. The modulation frequency is 30MHz so that the light will travel 10m before completing one cycle, allowing the camera to see objects at total of 5m away.

By taking successive images of a flat wall and comparing individual pulses, the noise was isolate and its root mean square (RMS) was determined to be 5.24 counts. It was noted that the pixel-to-pixel variation in each pulse was much larger than the noise level. This systematic error was addressed by recording data while the camera was covered, allowing no light to reach the lens. The RMS of pixel-to-pixel error was calculated to be 32.0 counts. Subtracting the error from a single pulse cleaned the signal significantly, see figure 2. The RMS of the difference of two pulses in the same frame was found to be 4.25 counts. Other combinations of pulses in the same frame were similar. This indicates that the errors for the four pulses are the same to within the level of noise and, therefore, subtract out in further calculations.

Intensity and phase information was calculated for several frames. By subtracting two successive frames, the RMS noise for both intensity and phase were found to be 3.74 counts and 1.27° respectively. There was found to be significant systematic errors, different than those described above, when calculating the phase information. Figure 3 shows an example of the error. Subtracting the phase calculated from data taken when the camera lens was covered but allowing some light to reach the lens, figure 4, corrects the error, figure 5. However, the method used does not separate entirely the systematic error from the signal needed to observe the error, therefore, it is likely that some systematic effects are still present.

2.2 Averaging Phase Maps

In order to reduce pixel-to-pixel variation in the phase map, an averaging scheme was performed. Squares of pixels were averaged to a single point over the map, the data was then fitted and the
standard deviation calculated. This was done by various degrees in order to determine a sufficient averaging size, see figure 6. Based on this, we will likely average over $4 \times 4$ squares of pixels as anything greater provides little improvement.

### 2.3 Calibration of Pixel Counts

To find the number of photons per count, light from the diode was reflected off a flat mirror 96 cm away from the camera and directed back to the camera lens. A pinhole with a known diameter of 0.60 mm was placed in front of the lens to restrict its active area to $\pi (0.060/2 \text{ cm})^2 = 0.0028 \text{ cm}^2$. Similarly, an aperture with a diameter of 0.115 cm was placed in front of the mirror, corresponding to an active area of $\pi (0.115/2 \text{ cm})^2 = 0.0104 \text{ cm}^2$. With the mirror adjusted to maximize the number of photons reaching the lens, the total number of counts was recorded for various integration times. We found that the camera became over saturated even for the lowest integration time setting. In response, attenuator plates were placed between the mirror and camera to reduce the intensity. With attenuators in place, the counts increased linearly with integration time until becoming over saturated, see figure 7. A power meter revealed that the chosen combination of attenuators reduced the signal’s intensity by a factor of 0.27, see figure 8. Light reflecting off the mirror and back to the camera lens passes through the attenuators twice so that the resulting attenuation is $0.27^2 = 0.075$. Removing the attenuator plates and replacing the mirror with a power meter with a slow time constant, set to 850 nm, and an attenuator cap, we measured the power density produced by the diode. Several readings indicated that the power increases linearly with integration time, see figure 9.

With the camera off, the power reads 0.0355 $\mu W$. Setting the integration time to 1000 $\mu s$, the power meter reads 0.1118 $\mu W$. Subtracting the background signal gives 0.0763 $\mu W$, so that the average power density is $\frac{0.0763}{0.0104 \text{ cm}^2} = 7.34 \mu W/cm^2$.

An oscilloscope trace of the output of a fast photo diode showed the 30 MHz modulation and confirmed that the duration of each pulse is equivalent to the integration time. Each frame consists of four pulses, and, for an integration time of 1000 $\mu s$, the time between frames is 22.25 ms for a total duty cycle of $\frac{4(1.000 \text{ ms})}{22.25 \text{ ms}} = 17.98\%$. The average power density at the aperture while the diode was on is therefore $\frac{7.34 \mu W/cm^2}{0.1798} = 40.8 \mu W/cm^2$.

Now, the power density at the pinhole should be a fourth of the power density at the aperture, since the light must now travel twice the distance to reach the pinhole. The power density at the pinhole is therefore $10.2 \mu W/cm^2$. Converting this to number of photons per second area yields:

$$I_{\text{pinhole}} = \frac{10.2 \mu W/cm^2}{(1.46eV/\text{photon})(1.6 \times 10^{-19} J/eV)} = 4.37 \times 10^{13} \text{ photons}/(s \cdot \text{cm}^2).$$

With an integration time of 1000 $\mu s$, the diode delivers $4.37 \times 10^{10} \text{ photons/cm}^2$ during each pulse to the pinhole. The number of photons reaching the lens when the mirror and attenuators are in place is:

$$N_{\text{detect}} = 0.075 \cdot 4.37 \times 10^{10} \text{ photons/cm}^2 \cdot 0.0028 \text{ cm}^2 = 9.2 \times 10^6 \text{ photons}.$$

It should be noted that only the modulated portion of the signal contributes to the number of counts because of the method used to calculated the intensity, as described in the Characterization of Errors subsection. The modulation depth was calculated by:

$$\text{modulation depth} = \frac{V_{\text{max}} - V_{\text{min}}}{V_{\text{max}} + V_{\text{min}}}$$

(4)
where $V_{max}$ and $V_{min}$ are the maximum and minimum voltage of the modulated signal when viewed on an oscilloscope. The modulation depth was found to be 0.416.

The total number of counts registered by all pixels when the integration time was set to 200 $\mu$s was 525. This leads to the number of photons per count:

$$N_{\text{photons}} = \frac{0.416 \cdot 9.2 \times 10^6 \cdot 0.2}{525} = 1500 \text{ photons per count.}$$

3 Kinect 2

The Kinect 2 emits infrared light from three identical sources, each operating identically when viewed on an oscilloscope. To confirm this, we imaged a flat wall at a distance of approximately 1.2 m while covering a combination of emitters for total of six combination. Comparing the various combinations revealed a standard deviation of 3.405 mm. The intensity maps did not vary by a significant amount among combinations where the same number of sources were covered.

<table>
<thead>
<tr>
<th>Combination</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1.210m</td>
</tr>
<tr>
<td>010</td>
<td>1.212m</td>
</tr>
<tr>
<td>001</td>
<td>1.217m</td>
</tr>
<tr>
<td>110</td>
<td>1.212m</td>
</tr>
<tr>
<td>101</td>
<td>1.213m</td>
</tr>
<tr>
<td>011</td>
<td>1.218m</td>
</tr>
</tbody>
</table>

Table 1: Variation in depth measurements for different combinations of emitters: 1 denotes the emitter is operating, 0 denotes that it is covered.

Rather than using a four pulse method to calculate intensity and phase, the camera emits a single pulse whose modulation shifts between three different frequencies, starting at 80 MHz then changing to 16 then 120 MHz. This three-frequency pulse lasts for 12 ms, each stage lasting 4 ms. For each pulse, the camera internally calculates the intensity and the depth from the object.

An assessment of the Kinect’s imaging limitations was done by (Lachat et al., 2015). After preforming the calibrations described in (Lindner et al., 2010), the group concluded that the difference between the measured and true distance varied from 5 to −5 mm for ranges between 0.8 and 4.5 m. Indeed, when we imaged a flat wall at a distance of approximately 1.2 m and moved the camera forward a true 10.2 mm using a translation stage, the measured change was 10.4 mm. They also showed that, over the course of 90 minutes, the measured distance would vary by approximately 8 mm depending how long the camera was operating, relating to the temperature of the device.

No significant pixel-to-pixel variation was described by (Lachat et al., 2015). Indeed, when imaging a flat wall, the noise contained in the uncalibrated data is on the same order as the pixel-to-pixel variation observed for both the intensity and depth data. Restricting the depth map to relativity flat region, the RMS error of the noise is 1.55 mm, while the RMS error of the depth data is 1.65 mm. Due to the systematic shape of the intensity map, we can only compare the pixel-to-pixel intensity variation and the noise visually, figure 10. The depth map does, however, contain structural error, figure 11, assumed to be due to differences in intensity, similar to the reflectivity error described above, since the data is uncalibrated.
4 Nonlinear Error

The literature describes errors exhibited by any ToF device different than the noise and fixed pattern error that can be filtered or subtracted out. So called “wiggling” errors cause the measured distance to be shifted from the true distance depending, non-linearly, on the true distance. It is speculated that this is caused by inhomogeneity in the signal shape, which, in practice, cannot be a true sinusoid (Lindner et al., 2010). Generally, wiggling errors are fixed using lookup tables or B-splines to approximate the error function, both requiring large amounts of reference data and precise distance measuring tools (Lindner et al., 2010). The reflectivity of the imaged object also leads to errors in the measured distance, but can be addressed by incorporating the intensity data and using a similar method as that used for the wiggling errors. Lastly, defects in the camera lens also lead to errors. The intrinsic parameters of the lens are often determined by imaging a checkerboard pattern in many different poses, as described by (Hansard et al., 2012). It is assumed that both the CamBoard and Kinect exhibit these errors.

5 Conclusion

The goal of this research was to characterize and remove systematic effects in ToF cameras before imaging objects through synthetic tissues. This was delayed due to our primary camera, the CamBoard, breaking early in the summer. It was demonstrated, however, that fixed pattern error could be isolated and removed from single-pulse data for the CamBoard, improving the signal approximately by a factor of 10. Furthermore, structural error in the phase data was also largely removed. It was also found that the CamBoard requires approximately 1500 photons to be incident on the lens to produce a count.

In future work, one could characterize several different cameras and select one that is best for DOT. Regardless of the camera selected, a calibration procedure must be performed to reduce time-invariant, nonlinear errors exhibited in all ToF cameras. After this, new diodes and fiber-optics must be made in order to achieve the desired geometry for DOT. Once this is complete, tests can be performed with synthetic tissues.

References


Figure 1: Representation of phase shift between emitted and received light.

Figure 2: Left: Single pulse data. Right: The same pulse with pixel-to-pixel error subtracted.

Figure 3: Phase data obtained by imaging a flat wall. There is a noticeable distortion on the right side of the map. Axes specify pixel number. The data has been smoothed with a moving average filter to better highlight the structure of the error.
Figure 4: Possible systematic error obtained by blocking the lens with stock paper, but allowing some infrared light to be reflected onto the paper and into the lens. There must be some light incident on the lens in order to observe this effect, therefore, a different method must be used to separate the systematic error from the signal needed to observe it.

Figure 5: The same data as figure 3, but with the systematic error subtracted. This shows a noticeable improvement as the phase now has radial symmetry and no significant distortions.
Figure 6: Standard Deviation for various averaging sizes.

Figure 7: Total number of counts registered for various integration times. The linear fit shown is only for points ranging from 100$\mu$s to 800$\mu$s as the camera begins to over saturate near 900$\mu$s.
Figure 8: Integration Time vs. Power with and without the attenuators. The resulting attenuation was found to be 0.27 by comparing the slope of the curves, $slope_{\text{with}}/slope_{\text{without}}$.

Figure 9: Integration Time vs. Power reading.

Figure 10: Left: Intensity map. Right: Intensity noise.
Figure 11: Depth map with systematic error potentially due to the data being uncalibrated.