Single Pixel Amplitude-Modulated Time-of-Flight Camera

Andrew Ponec
Stanford University
ponec@stanford.edu

Cedric Yue Sik Kin
Stanford University
cedyue@stanford.edu

Abstract

Depth sensing is an increasingly important area of imaging research due to the proliferation of applications which require knowledge of the three-dimensional space surrounding a user or vehicle. Applications such as autonomous vehicles, augmented reality, and robotics have led to the recent development or advancement of a large number of depth imaging mechanisms, including stereo triangulation, structured light, coded aperture, and time-of-flight cameras. Current depth cameras are limited by a fundamental tradeoff between resolution and cost. We propose and demonstrate a novel single pixel camera design that measures depth with a single inexpensive optical sensor by measuring the phase of backscattered LED light from an actively illuminated scene. High resolution is obtained from the single sensor through the use of a Digital Micromirror Device (DMD), which acts as a spatial light modulator (SLM). Our results show a reduction in noise when using orthogonal spatial modulation masks derived from Hadamard Matrices rather that a simple raster scanning approach due to increased signal strength resulting from the use of a larger portion of the light returning from the scene.

1. Introduction

1.1. Motivation

Recent years have seen a surge of research into novel depth sensing technologies in order to overcome the limitations of existing systems and improve the cost, lateral and depth resolution, capture speed, safety, and range of depth cameras. One of the most promising depth sensing technologies is time-of-flight (ToF) sensing. Time-of-flight sensors utilize the fixed, known speed of light to determine depth by measuring the round-trip time it takes projected light to reach an object in the scene and return to an onboard sensor. Time-of-flight cameras, especially direct ToF cameras which measure returns from a single laser pulse, provide relatively unambiguous depth data and have low processing requirements, good depth resolution, and fast capture speed. However, these cameras suffer from a difficult tradeoff between cost and resolution. Many direct ToF cameras (also referred to as 3D flash LIDARs) cost tens of thousands of dollars despite relatively low lateral resolutions [10]. This prevents their use on all but the most cost insensitive military and space applications.

Another approach to resolving the high cost of depth sensors in mechanical scanning. A number of systems, including the popular Velodyne LIDAR systems used on many autonomous vehicles, have used mechanical scanning of a small number of accurate LIDAR depth sensors[8]. This approach also allows greater resolution from a smaller number of depth sensors, but costs remain high (Velodynes LIDARs are estimated in tens of thousands of dollars [9]) and introduces other disadvantages: mechanical scanning increases scene acquisition time, and the addition of moving mechanical components increases the cost and weight of the system while reducing reliability.

1.2. Related Work

A promising approach to sidestep the resolution/cost tradeoff in depth sensing without resorting to mechanically sweeping the scene is to use a spatial light modulator (SLM) to encode scene depth information on a single sensing element. Such an approach for non-depth cameras was introduced by Duart et al. in their foundational paper describing compressive sensing with a single pixel camera using a DMD for spatial light modulation [4].

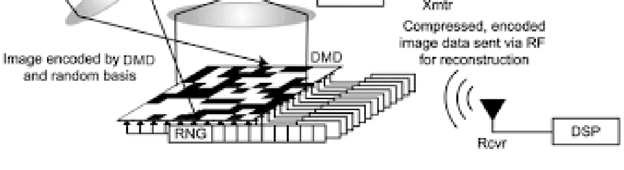

Figure 1. Single-Pixel camera concept.

The simplest implementation of such a system would send simple raster patterns to the spatial light modulator, which accomplishes the same goal as mechanically scanning the scene but offloads the scanning task from macro-mechanical systems to Micro-Electro-Mechanical Systems (MEMS) such as Digital Micromirror Devices (DMDs). An improved approach uses a set of masks with many pixels switched on during each capture. In this approach, the patterns are drawn from a set of orthogonal vectors, and the intensity from all of the on pixels is integrated on the single sensor. If the basis set of patterns is chosen correctly, the matrix modifying the scene is invertible, and the scene can be reconstructed from the captured data by applying the inverse matrix to the captured data. For this reconstruction technique, the total number of measurements must be equal to the desired resolution of the resulting image.

Hadamard matrices have orthogonal and symmetric properties that allow it to be inverted and the rows of Hadamard matrices can be used as a basis set of patterns for spatial light modulation.

$$\mathrm{b} = \mathrm{H}^{-1}\mathrm{a}$$

where a is the captured data, b is the reconstructed scene and H is the Hadamard matrix.

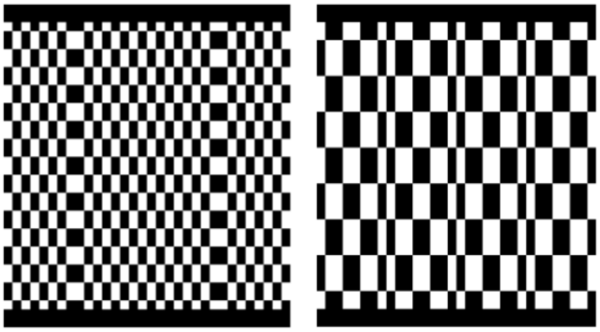

Figure 2. Example of Hadamard patterns.

Duart et al. demonstrated improvements to this fully-sampled approach by exploiting the inherent sparseness in most natural images. By using appropriate reconstruction priors, it is possible to accurately capture and reconstruct a scene with many fewer samples than the Nyquist sampling theorem requires. This approach, called compressive sensing, has been applied to a number of imaging applications [1][11][7].

A good example of a system that can take advantage of the unique characteristics of a single pixel camera setup is hyperspectral imaging. Hyperspectral image sensors (spectrometers) are extremely expensive compared to standard image sensors, making the production of a spectrometer array to capture hyperspectral images prohibitively expensive. A single pixel camera setup can allow a single spectrometer to capture an entire scene without mechanical motion, and possibly using compressive sensing, greatly improving the usefulness of such as system.

Together, the advantages of achieving the full resolution of the spatial light modulator with only a single imaging sensor, as well as the possibility to capture compressed images directly, has made single pixel cameras an intriguing solution to a number of optical sensing problems.

Single pixel sensing has also been applied to the field of depth sensing. Edgar et. al. [2] demonstrated a system that utilized a pulsed laser and accurate timing electronics to measure scene depth, with the spatial light modulation applied to the outgoing signal. The following diagram illustrates the experimental setup from that work.

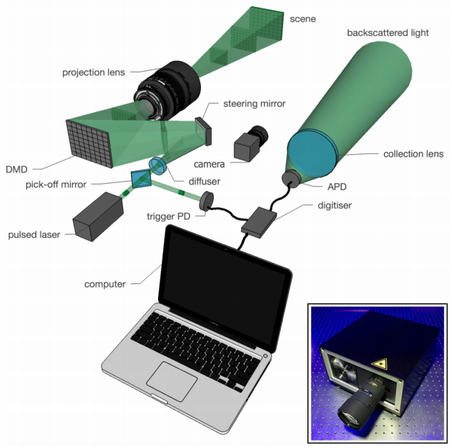

Figure 3. Single pixel depth sensor using pulsed laser.

Kirmani et. al.[6] demonstrated a similar setup using compressive depth sensing in which a SLM is applied to outgoing laser pulses, with timing-critical electronics directly capturing the elapsed time between the outgoing pulses and the return pulses from scene reflections.

While effective in reducing the number of active elements while preserving high resolution with a spatial light modulator, the reliance on fast pulsed lasers and high resolution time-to-digital converters increases complexity and cost, limiting the practical use of these depth cameras.

2. New Single Pixel Depth Camera Using Phase Detection

We propose a single pixel depth camera utilizing phase-based rather than time-based depth measurement. Previous implementations of single-pixel depth cameras have relied on fast repetitive laser pulses and extremely fast detection electronics in order to directly measure the time between the outgoing and return pulses. Our method draws upon the use of a relative phase shift between continuously amplitude-modulated light in order to determine depth. This approach enables for the use of LEDs as light sources rather than lasers, which allows the scene to be illuminated with higher optical power without safety concerns. Additionally, the measurement of phase using RF measurement ICs is standard and inexpensive, and easier to work with than the time-to-digital converters needed for pulsed laser setups that must measure transmission to reception times with picosecond accuracy. The primary disadvantage of phase-based measurement approaches is phase wrapping, which limits the unambiguous measurement range to 1/2 the modulation wavelength. This limitation can be overcome by using multiple modulation frequencies, which extends the unambiguous range to the least common multiple of the ranges used, but this variation increases system complexity and is not used in all phase-based systems.

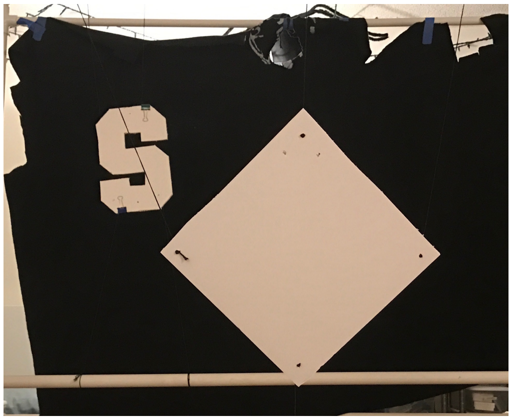

Figure 5. Test scene with objects at 1.3m, 1.7m and background at 2.1m

2.1. Image Processing

We use both raster masks and Hadamard matrix masks to encode scene information and compare the results. Currently we use the full Hadamard basis set; future work will include compressive sensing using a subset of the basis.

To denoise the phase data, our general formulation for the image reconstruction is

$$\text{minimize } \frac{1}{2}\|\text{Cx - b}\|_2^2 + \Gamma(\text{x})$$

where $\Gamma(\text{x})$ is the regularizer modeling prior knowledge of phase data in natural images. Given the sparcity of the data, we use the total variation (TV) prior [5]. However using the TV prior makes the above equation convex, with the l1-norm not differentiable, making it hard to solve the regularized linear system. We use the alternating direction method of multipliers (ADMM) to solve it [3].

3. Experimental Results

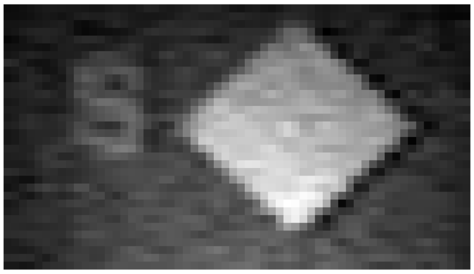

Figure 6. Intensity map from Raster.

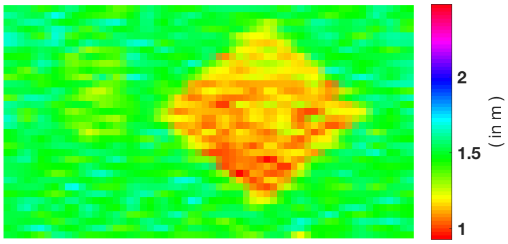

Figure 7. Depth as hue from Raster.

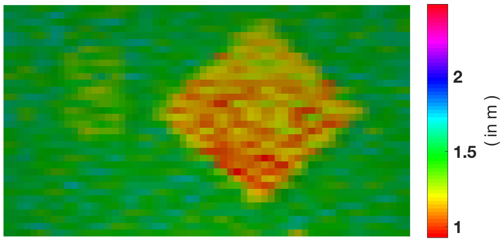

Figure 8. Depth and intensity as hue and value from Raster.

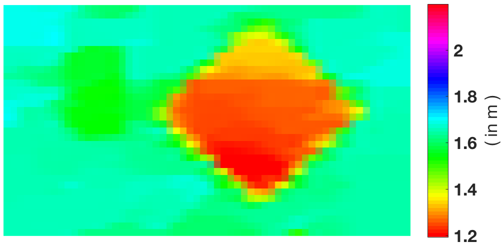

Figure 9. Depth as hue from Raster - denoised.

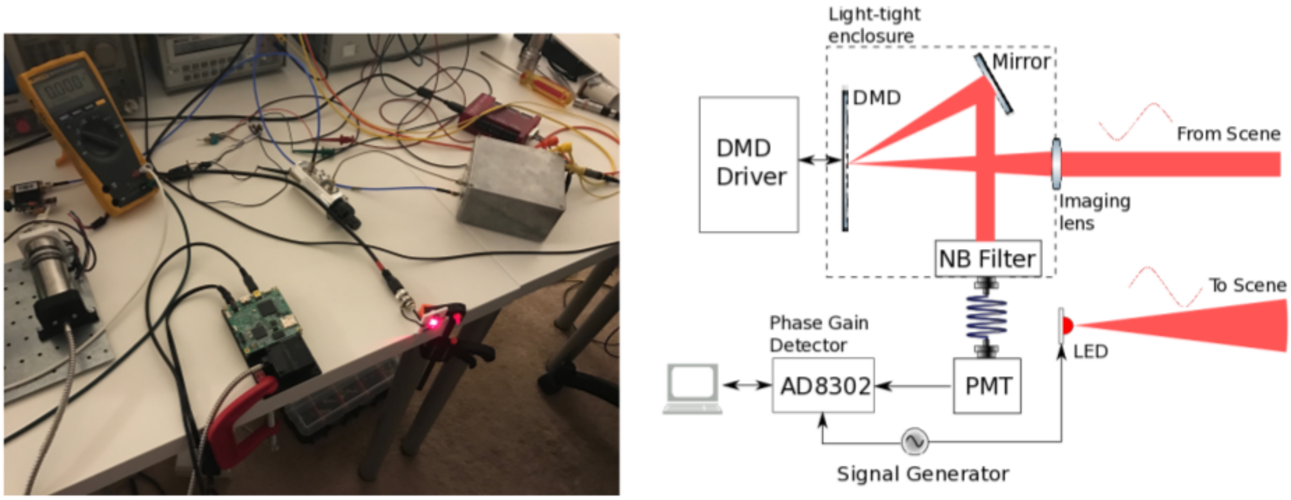

Figure 4. Picture of our camera (left). Our schematic (right).

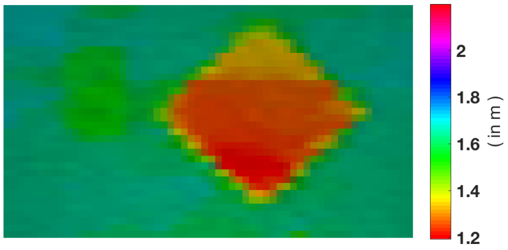

Figure 10. Depth and intensity as hue and value from Raster - denoised.

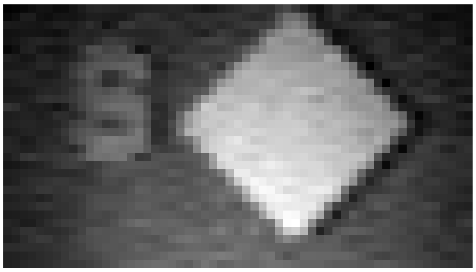

Figure 11. Intensity map from Hadamard reconstruction.

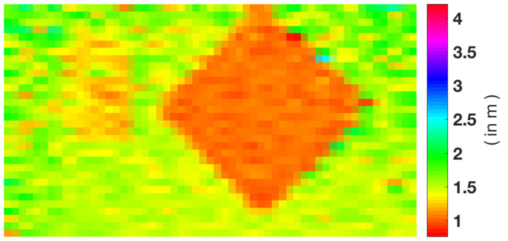

Figure 12. Depth as hue from Hadamard reconstruction.

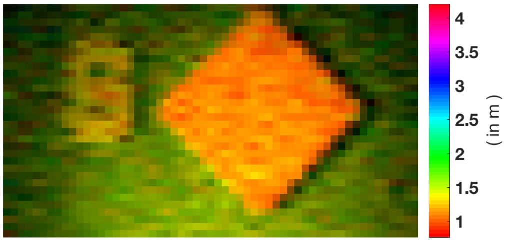

Figure 13. Depth and intensity as hue and value from Hadamard reconstruction.

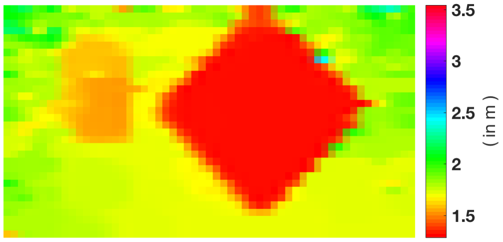

Figure 14. Depth as hue from Hadamard reconstruction - denoised.

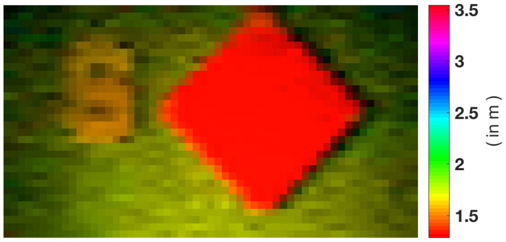

Figure 15. Depth and intensity as hue and value from Hadamard reconstruction - denoised.

4. Discussion

Our results demonstrate the viability of a phase-based single pixel depth camera as a depth sensor. While the experimental setup has significant room for improvement, the current images showed representative and fairly accurate depth images can be captured without the use of pulsed lasers and fast time to digital converters. Additionally, comparison of the result for the raster scanned image and the image captured with the Hadamard-derived masks show that the Hadamard images had lower noise, and the Stanford logo was easily differentiable from the background in the Hadamard but not the raster images. This effect demonstrates both the limitations of our experimental setup and the usefulness of the Hadamard approach. While the RF signal carrying scene information generated by the photomultiplier tube was filtered by a narrow-band quartz filter (10MHz) before being measured by the AD8302 phase/gain measurement chip, external RF interference from nearby sources added considerable noise to the signal. As this noise was signal-independent, the SNR of the experimental setup varied based on signal level.

When imaging in raster mode, the total captured light was only 1/1024 of the total light hitting the DMD, since only a single pixel in the 32x32 array sent light to the PMT. In contrast, when capturing with Hadamard masks, one half of the available light was sent from the DMD to the PMT, leading to much higher signal levels. The stronger signal in the presence of signal-independent noise led to the better captured image quality of the Hadamard images. Further improvements in image quality, especially in the phase/depth measurements, would combine more careful RF design to eliminate outside noise sources and would increase signal strength by using a more powerful light source. Our experimental setup used a single LED driven by a function generator; future implementations using a dedicated LED driver could easily produce an order of magnitude stronger illumination signal, which would likely significantly reduce phase noise. A final potential improvement could come in image capture rate. The experimental prototype relied on sending mask patterns sequentially to the lightcrafter over USB, which added 200 ms of latency per pattern. As the TI Lightcrafter is capable of displaying 4k binary frames per second, and the bandwidth of the phase-gain detector is 30 MHz, the image capture rate can be significantly improved.

While the captured images suffered from noise in the phase measurements, the de-noising algorithm using a Total Variation (TV) prior was effective in reducing the noise present in the captured phase signal. The scene was very favorable to the TV prior due to the sparse gradients present in the scene; future work involving more natural three dimensional scenes might require alternative de-noising techniques for good performance.

5. Conclusion

We have demonstrated a single pixel depth camera based on phase detection of LED light, which to the authors' knowledge is the first such system to be described in the academic literature. The experimental system, while not fully optimized, was capable of capturing depth images of a static scene, and post-processing on the captured data significantly reduced noise in the phase measurements. Future work will focus on SNR improvements through the reduction of externally coupled noise and by increasing the power output of the LED illuminator. Masks derived from Hadamard matrices showed promise in improving SNR, and future work will explore the use of compressive sensing techniques to undersample the image and decrease image capture time. As depth sensing technologies become more important for a wide range of applications, we believe the single pixel camera approach using phase-based depth detection could complement existing depth sensing technologies as a low-cost yet high-resolution depth imaging system.

6. Acknowledgments

We would like to acknowledge the support and advice of our project mentor Matthew OToole and Professor Gordon Wetzstein. We also would like to acknowledge Ben Johnson (Stanford M.S. EE 2013) for his invaluable support and advice, including providing the PMT, Phase/Gain measurement board, fiber and RF accessories, workspace, and lab equipment used in this project.

References

[1] W. L. e. a. Chan. A single-pixel terahertz imaging system based on compressed sensing. *Applied Physics Letters*, 93, 2008.

[2] e. a. Edgar, Matthew P. Real-time 3d video utilizing a compressed sensing time-of-flight single-pixel camera. *SPIE Nanoscience + Engineering*, 2016.

[3] B. et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 2010.

[4] M. F. D. et al. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25, 2008.

[5] R. L. et al. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 1992.

[6] e. a. Kirmani, Ahmed. Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor. *Optics Express*, 2011.

[7] N. e. a. Radwell. Single-pixel infrared and visible microscope. *Optica*, 2014.

[8] B. Schwarz. Lidar: Mapping the world in 3d. *Nature Photonics*, 2010.

[9] N. Shchetko. Laser eyes pose price hurdle for driverless cars, 2014.

[10] H. Shim and S. Lee. Performance evaluation of time-of-flight and structured light depth sensors in radiometric/geometric variations, 2012.

[11] . K. K. Sun, T. Compressive sensing hyperspectral imager. in computational optical sensing and imaging. *Optical Society of America.*, 2009.