3dim: Compact and Low Power Time-of-Flight Sensor for 3D Capture Using Parametric Signal Processing

Andrea Colaço[1,2], Ahmed Kirmani[1], Nan-Wei Gong[1,2],
Tim McGarry[3], Laurence Watkins[3], and Vivek K Goyal[1]
[1]Massachusetts Institute of Technology [2]3dim Tech, Inc. [3]Princeton Optronics

Abstract

We introduce an active optical sensor, 3dim, for capturing three-dimensional (3D) scene structure. 3dim is an active time-of-flight (TOF) technique, which is based on parametric modeling and processing of scene impulse responses. As opposed to estimating one depth value per pixel in a conventional TOF sensor, 3dim operates by estimating multiple object depths at each of the baseline-separated photodetectors. 3dim offers simple hardware construction, high frame rates, small size, insensitivity to ambient light, and low power consumption. Unlike existing depth sensors, it does not provide a depth map with full spatial resolution for general scenes, but rather high 3D object localization accuracy in application-specific scenarios. Our 3dim prototype is built using standard, commercially-available components. We demonstrate centimeter range resolution for gesture tracking applications on mobile platforms, a resolution that is very fine relative to the system bandwidth and optical power.

Overview

In the post-PC era, 3D sensing is being increasingly integrated in consumer electronic devices to enable gesture-based user interfaces. The demand for accurate, low power and compact depth cameras is thus evident. The past year alone has seen the launch of several sensors for consumer electronics applications such as the Leap Motion Controller [1], ultrasonic touchless technology by Elliptic Labs [2], and compact depth cameras by PrimeSense [3] and PMD [4]. The key performance features that are desirable in 3D sensors are high depth accuracy and resolution, high frame rate, insensitivity to ambient light, and large working volume. It remains difficult to satisfy these performance requirements given the constraints on size and power for mobile devices (see Fig. 1).

In this paper, we introduce 3dim, an optical sensor combining a TOF-based module with an RGB camera module to achieve 3D localization and tracking of objects such as users' hands while gesturing to a mobile device, or a real-world object augmented with virtual information. Unlike TOF cameras and other general-purpose 3D acquisition technologies, 3dim is an application-specific sensor. Its TOF module uses neither focusing optics nor a full-resolution 2D sensor array, and hence it does not produce scene depth maps. Instead, it provides 3D acquisition capabilities specific to tracking gestures and planar objects.

The current image processing and computer vision pipeline operates by first capturing a full 2D or 3D image of the scene, then processing to detect objects, and finally performing object parameter estimation. This pipeline works for general scenes but requires significant computation and acquisition resources. It obtains a full-resolution image even though the objects of interest are simple and few. Our acquisition architecture can be used to disrupt the standard image processing and computer vision pipeline. With few sensors and low computational complexity, 3dim can track hands or infer planar object pose and orientation with high accuracy. This obviates high-complexity processing on high-resolution image data.

	Frame rate	Daylight sensitivity	Depth resolution	Total power	Working range
3dim	100 fps	low	< 1 cm	< 150 mW	0-2 m
Leap	90-100 fps	high	1 mm	3-5 W	< 0.6 m
Ultrasound	50 fps	none	coarse ~ 2-5 cm	~ 300 mW	< 1 m
2D cameras	25-30 fps	none	coarse > 5 cm	200 mW	0-2 m
Structured light 3D	30 fps	high	coarse > 2 cm	3-5 W	0.8-3 m
Time-of-flight 3D	30 fps	high	< 1 cm	3-5 W	0-2 m

Figure 1: Comparison of 3D sensing technologies for gesture-sensing applications.

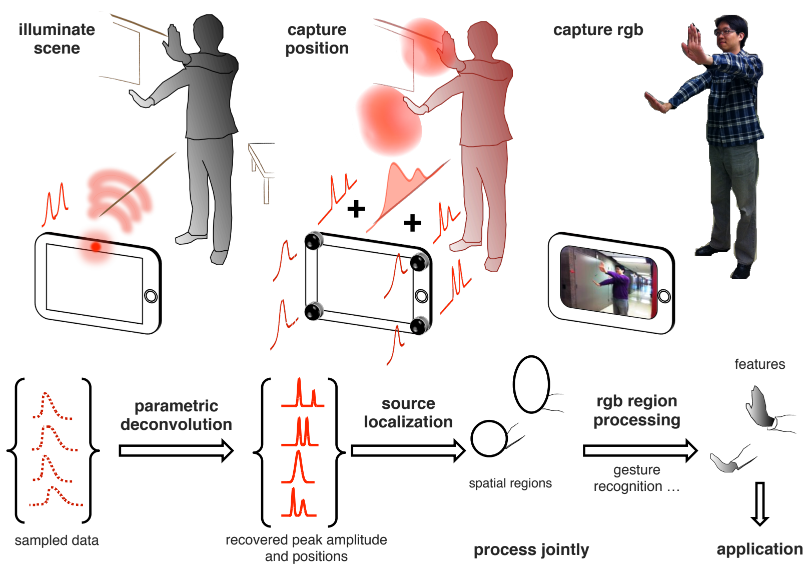


Figure 2: The signal processing pipeline for 3dim.

Theory of Operation

3dim uses parametric modeling and processing of scene impulse response [5,6] to reduce the cost, power, and complexity of 3D acquisition. The intuition behind 3dim is that when objects—like a user's hands—that occupy a small portion of the sensor's field-of-view (FOV) are illuminated with a pulse of light, the reflected response is a short-duration signal that contributes high-frequency components to the scene impulse response (see Fig. 2). In contrast, large smooth objects—such as a user's body—contribute a smooth, low-frequency signal. 3dim uses parametric signal estimation [7] to recover the high-frequency components of the scene impulse response by first lowpass filtering followed by sampling and then nonlinear spectral estimation to estimate the hand locations in 3D.

The 2D camera module in the 3dim sensor captures an RGB image. This image could be used alone for hand gesture recognition using well-known computer vision techniques [8], but this has two weaknesses: it would not have the range or depth information required for 3D gestures; and

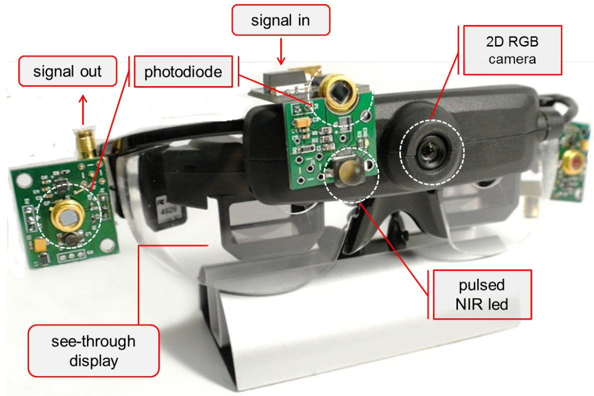


Figure 3: 3dim sensor hardware prototype mounted on smart glasses.

it would not be robust because the complexity of cluttered, real-world scenes cause conventional methods to fail. The 3dim sensor uses the locations provided by the TOF module to identify regions of interest for the RGB processing. This increases the robustness of the RGB processing while lowering its computational complexity.

Hardware Prototype and Applications

One 3dim prototype was built using four baseline-separated, unfocused CMOS photodetectors coupled with low-noise amplifiers, and one omnidirectional VCSEL laser diode [9], which modulates the emitted light intensity at 2 MHz with high opto-electronic efficiency. This prototype achieves a high frame rate of 60–90 frames per second (fps), maintains a compact form factor, remains insensitive to ambient light, uses only 10 mW average optical power, and works up to 2 m range with a 90° FOV. Another prototype with three photodetectors is shown mounted on Vuzix STAR 1200XL smart glasses in Fig. 3.

Fig. 4 shows signals acquired by the TOF module and a demonstration of 3D gesture capture built upon the TOF module alone. Fig. 5 shows the demo system operating in bright daylight; competing solutions cannot be used in this environment.

References

[1] Leap Motion. Leap Motion Controller. https://www.leapmotion.com/.

[2] Elliptic Labs. Windows 8 Gesture Suite. http://www.ellipticlabs.com/.

[3] PrimeSense. Capri 3d Sensor. http://www.primesense.com/.

[4] PMD. Camboard nano depth camera. http://www.pmdtec.com/.

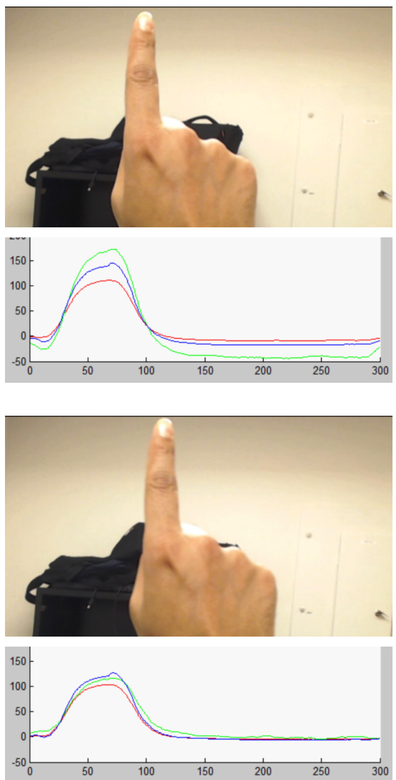
[5] A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal. Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor. *Optics Express*, 19(22):21485–21507, October 2011.

[6] A. Kirmani, A. Colaço, and V. K. Goyal. SFTI: Space-from-time imaging. In F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, editors, *Emerging Technologies for 3D Video*. Wiley, 2013. To appear.

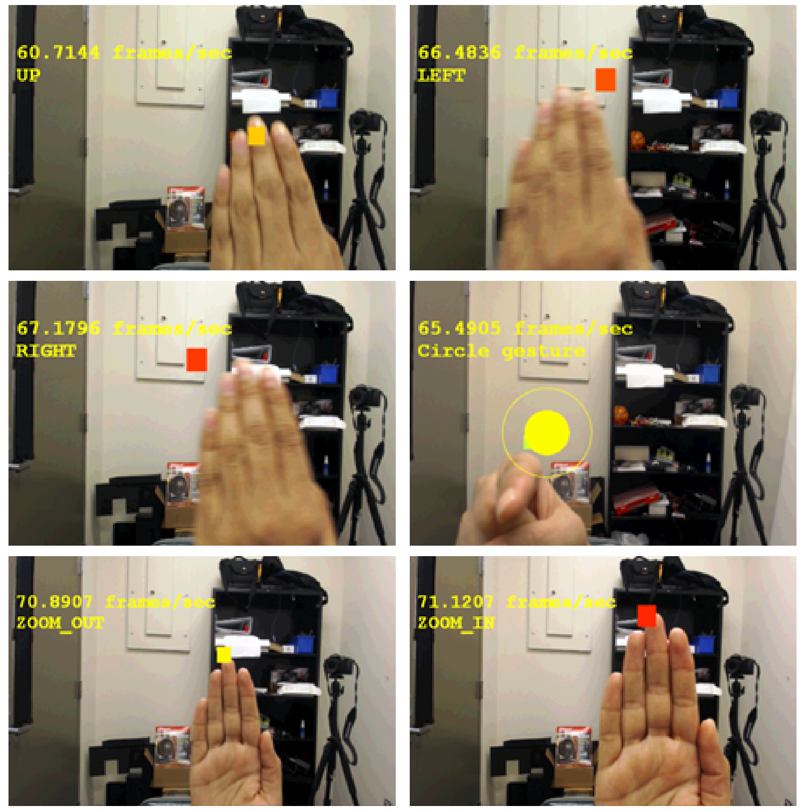
[7] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. Signal Process.*, 50(6):1417–1428, June 2002.

[8] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2010.

[9] Princeton Optronics. CW and QCW VCSEL array chip-on-submount. http://www.princetonoptronics.com/products/ProductMatrix.php.



(a) Tracking resolution.



(b) 3D motion-activated gestures.

Figure 4: (a) For a small change in finger position, there is a noticeable change in measured signal amplitudes. (b) Gestures recognized from TOF module alone (only 3D hand coordinates).

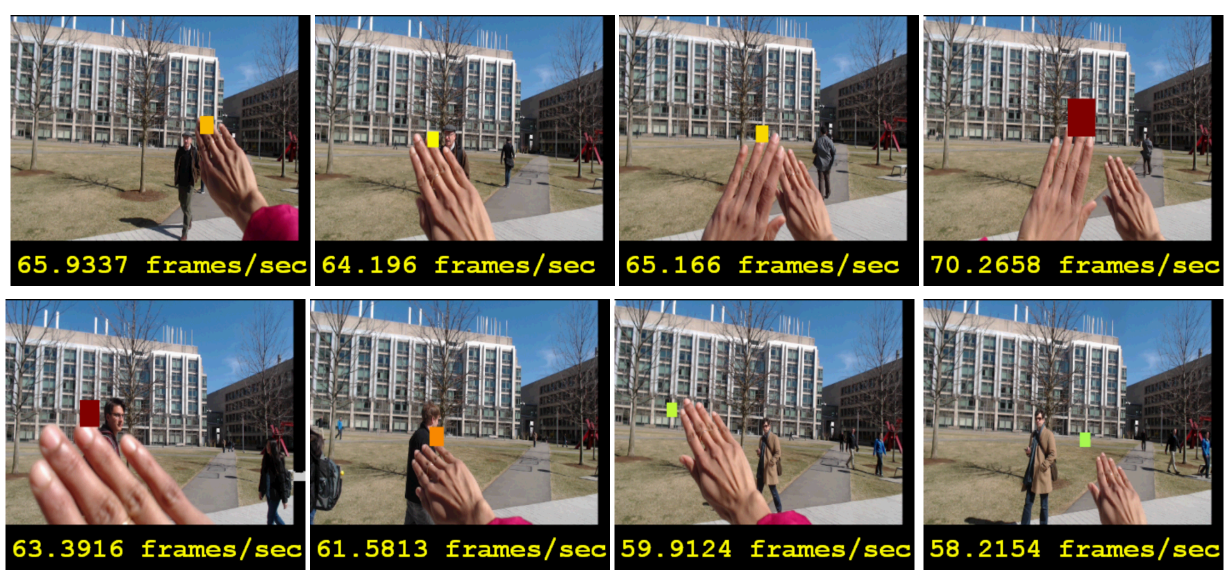


Figure 5: 3dim sensor operation in bright daylight.