

Epipolar Time-of-Flight Imaging

SUPREETH ACHAR, Carnegie Mellon University
JOSEPH R. BARTELS, Carnegie Mellon University
WILLIAM L. 'RED' WHITTAKER, Carnegie Mellon University
KIRIAKOS N. KUTULAKOS, University of Toronto
SRINIVASA G. NARASIMHAN, Carnegie Mellon University

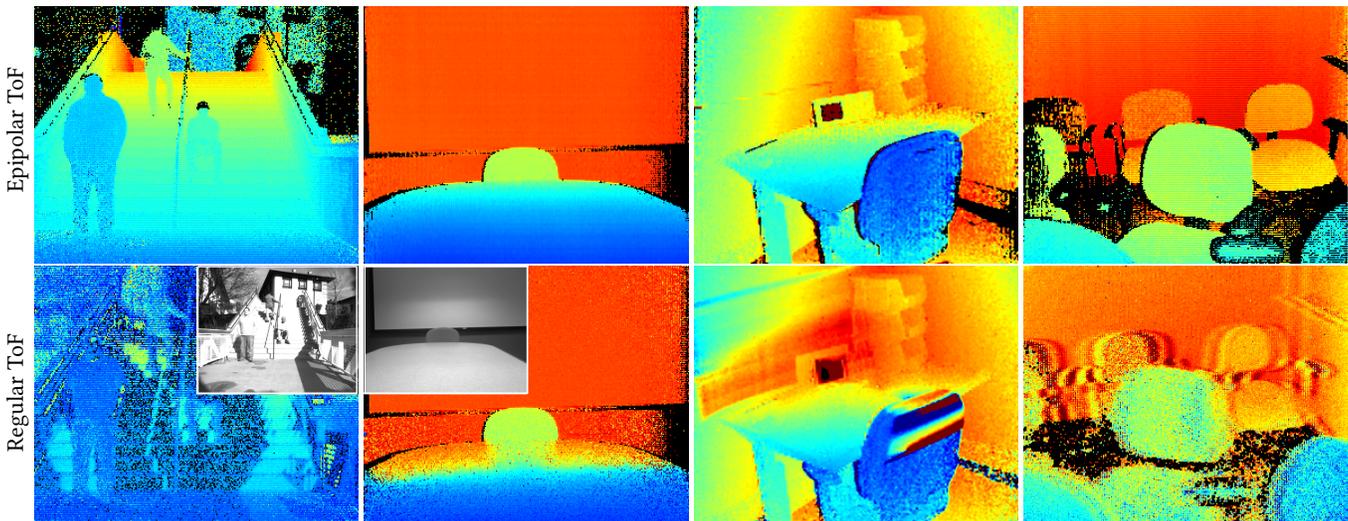


Fig. 1. We demonstrate, for the first time: (a) Live 3D CW-ToF imaging in sunlight (70 klx) with 15 m range (people walking on stairs, phase wraps around on distant building). (b) Elimination of most global light transport during image capture (glossy reflection from projection screen onto a conference table). (c) Capture virtually no interference from other ToF devices operating at the same frequency (observe errors on the wall and chair with regular ToF). (d) Live un-distortion of depth maps even when the camera is shaken violently during scene exposure (observe hard to remove ghosting errors in regular ToF)

Consumer time-of-flight depth cameras like Kinect and PMD are cheap, compact and produce video-rate depth maps in short-range applications. In this paper we apply energy-efficient epipolar imaging to the ToF domain to significantly expand the versatility of these sensors: we demonstrate live 3D imaging at over 15 m range outdoors in bright sunlight; robustness to global transport effects such as specular and diffuse inter-reflections—the first live demonstration for this ToF technology; interference-free 3D imaging in the presence of many ToF sensors, even when they are all operating at the same optical wavelength and modulation frequency; and blur-free, distortion-free 3D video in the presence of severe camera shake. We believe these achievements can make such cheap ToF devices broadly applicable in consumer and robotics domains.

CCS Concepts: •Computing methodologies → 3D imaging; Active vision; Epipolar geometry; Computational photography;

Additional Key Words and Phrases: Time-of-flight, depth cameras

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2017 ACM. 0730-0301/2017/7-ART37 \$15.00
DOI: <http://dx.doi.org/10.1145/3072959.3073686>

ACM Reference format:

Supreeth Achar, Joseph R. Bartels, William L. 'Red' Whittaker, Kiriakos N. Kutulakos, and Srinivasa G. Narasimhan. 2017. Epipolar Time-of-Flight Imaging. *ACM Trans. Graph.* 36, 4, Article 37 (July 2017), 8 pages.
DOI: <http://dx.doi.org/10.1145/3072959.3073686>

1 INTRODUCTION

Time-of-flight (ToF) depth sensors have become the technology of choice in diverse applications today, from automotive and aviation to robotics, gaming and consumer electronics. These sensors come in two general flavors: LIDAR-based systems that rely on extremely brief pulses of light to sense depth, and continuous-wave systems that emit a modulated light signal over much longer duration. The former can acquire centimeter-accurate depth maps up to a kilometer away in broad daylight but they have low measurement rates and their cost per pixel is orders of magnitude higher than latter—whose range, outdoor operation and robustness are extremely limited. Since low cost, large-scale production and high measurement rate often trump other considerations, continuous-wave ToF (CW-ToF) sensors continue to dominate the consumer electronics and low-end robotics space despite their shortcomings.

In this paper we present a first attempt to significantly reduce these shortcomings by *energy-efficient epipolar imaging*. The idea is to project a continuously-modulated sheet of laser light onto a sequence of epipolar planes that is chosen carefully and that spans the field of view. At the same time, only the strip of CW-ToF pixels that belong to each epipolar plane is exposed (Figure 2). Our prototype implementation couples a specially-built projection system to an off-the-shelf CW-ToF sensor that has a controllable region of interest. It outputs live 320×240 3D video at 7.5 frames per second, with the frame rate only limited by the sensor’s API.

Epipolar imaging was first proposed by O’Toole *et al.* [2014b] for acquiring live direct-only or global-only video with a conventional (non-ToF) video sensor. The approach was extended to the ToF domain by O’Toole *et al.* [O’Toole et al. 2014a] but its energy efficiency was very low and it involved capturing more than 500 images to calculate a single “direct-only” ToF image. In the context of triangulation-based 3D imaging, O’Toole *et al.* [2015] showed that significant improvements in energy efficiency and robustness can be achieved with a 2D scanning-laser projector and a rolling shutter camera. Our approach can be thought of as extending this idea to the ToF domain; as such, it inherits all the advantages of non-ToF energy-efficient epipolar imaging while also tackling major challenges that are specific to CW-ToF.

First and foremost, the range of CW-ToF sensors is severely limited by power consumption and eye safety considerations. Although most sensors electronically subtract the DC component of incident light [Oggier et al. 2005], photon noise from strong ambient sources such as sunlight easily overwhelms the CW-ToF signal at distances more than a few meters outdoors and typical frame rates [Cossairt et al. 2015; Gupta et al. 2013; Mertz et al. 2012]. By concentrating the light source’s energy into a single sheet, epipolar ToF boosts this range to ten meters and acquires useful—albeit noisier—depth signal at over 15 m outdoors.

Second, the depth accuracy of CW-ToF sensors is strongly affected by global illumination effects such as inter-reflections. These effects produce longer light paths and thus show up as a source of structured additive noise. They are extremely common indoors (*e.g.*, corners between walls, shiny surfaces of tables and floors, mirrors, *etc.*). Methods that aim to cancel the effects of global illumination *a posteriori* [Dorrington et al. 2011; Gupta et al. 2015; Heide et al. 2014; Kadambi et al. 2013] require extra image measurements and make strong assumptions in how they model transient responses. In contrast, epipolar CW-ToF optically blocks almost all global light paths *prior* to acquisition. This gives us significant robustness to all forms of global light transport without having to capture additional images.

Third, as devices equipped with CW-ToF depth sensors become increasingly common indoors and outdoors, they must be able to operate without interfering with each other. While non-interference between devices of a given make and model can be achieved by varying modulation frequency across them, robustness against the broader ecosystem of CW-ToF sensors is desirable. We show that epipolar ToF enables interference-free live 3D imaging even for devices that have the exact same modulation frequency and light source wavelength.

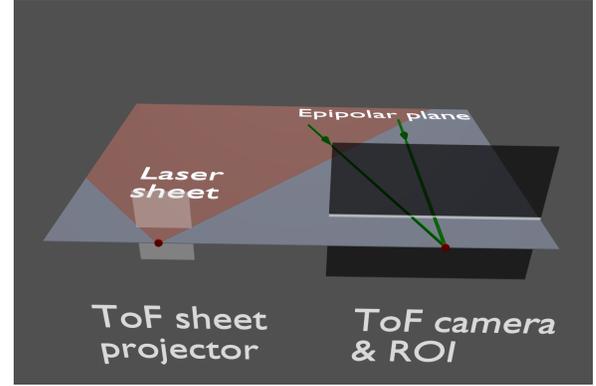


Fig. 2. Epipolar time of flight. A projector that generates a steerable sheet of modulated laser light is combined with a ToF sensor whose rows can be exposed one at a time. The projector and camera are placed in a rectified stereo configuration so that the light sheet always lies on an epipolar plane between the projector and the camera. At any given instant, only the row of camera pixels on the epipolar plane is exposed to light..

Last but not least, CW-ToF sensors must acquire two or more frames with a different phase of emitted light in order to compute a single depth map. This makes them highly sensitive to camera shake: unlike conventional cameras where this shake merely blurs the image [Fergus et al. 2006], camera shake in CW-ToF causes the static-scene assumption to be violated [Shrestha et al. 2016]. This leads to depth maps that are both blurry and corrupted by motion artifacts. Epipolar ToF makes it possible to address both of these problems: motion blur is minimized because only a very short exposure is used for each epipolar plane; motion artifacts and depth errors are minimized by acquiring multiple phase measurements *per epipolar plane* rather than per frame. The rolling-shutter-like distortions [Baker et al. 2010; Kerl et al. 2015] due to the sequential nature of epipolar-plane ToF can be reduced by scheduling the sequence of epipolar planes so that post-acquisition distortion correction becomes easier.

2 CONTINUOUS WAVE TIME OF FLIGHT

The operating principles of CW-ToF cameras are discussed in [Lange and Seitz 2001]. To summarize, these cameras use a temporally-modulated light source and a sensor where the exposure is also modulated during integration. If the illumination modulation function is $f_{\omega}(t) = \cos(\omega t)$ and the sensor modulation function is $g_{\omega, \phi}(t) = \cos(\omega t + \phi)$ where ω is the modulation frequency in rad/s and ϕ is the phase offset between the source and sensor modulation functions, then the measurement integrated over an exposure time of t_{exp} at a pixel x is

$$I_{\omega, \phi}(x) = \int_0^{t_{\text{exp}}} f_{\omega}(t) * [h_x(t) + A_x] g_{\omega, \phi}(t) dt \quad (1)$$

$$= \frac{t_{\text{exp}}}{2} \int_0^{\infty} \cos(\omega \tau - \phi) h_x(\tau) d\tau, \quad (2)$$

where $*$ denotes convolution, $h_x(t)$ represents a pixel’s transient response to the active light source and A_x is the response due to

the DC component of the active light source as well as other ambient sources. In practice, $I_{\omega, \phi}(x)$ is measured by integrating incoming light to two different storage sites (called taps) depending on whether $g_{\omega, \phi}(t)$ is positive or negative and then taking the difference between the stored values. Thus even though A_x drops out of the integral, ambient light still adds to the measurement shot noise.

If there are no indirect light paths between the light source and sensor pixel x , then $h_x(t) \propto \delta(t - l(x)/c)$ where c is the speed of light and $l(x)$ is the length of the path from the light source to the scene point corresponding to x and back to the sensor. Assuming the scene is static, we can recover the path length $l(x)$ by capturing a pair of images at the same frequency but two different modulation phases $\phi = 0$ and $\phi = \pi/2$:

$$l(x) = \frac{c}{2\omega} \operatorname{atan2} \left(I_{\omega, \frac{\pi}{2}}(x), I_{\omega, 0}(x) \right). \quad (3)$$

The pixel depth $z(x)$ can be computed from $l(x)$ using the geometric calibration parameters of the light source and sensor.

3 EPIPOLAR TIME OF FLIGHT

To realize the geometry of Figure 2, we use a line laser source with a 1D-scanning mirror that projects a steerable light sheet onto the scene. No currently-available CW-ToF sensor provides controllable exposure coding across the 2D pixel array. Taking into account available off-the-shelf hardware, there are three possible ways to restrict exposure to pixels on a single epipolar plane: (1) use a Digital Micromirror Device (DMD) to mask all other pixels; (2) use a 1D sensor and a controllable mirror to select the epipolar plane it should image; or (3) use a 2D sensor with a controllable region of interest (ROI). We chose the third option because it is much more light-efficient than a DMD mask and leads to a much simpler design. We make the ROI one row tall to match the requirements of epipolar ToF.

Epipolar plane sampling. As explained in Section 2, CW-ToF requires at least two images to recover depth. To cover an entire scene using epipolar ToF, the active epipolar plane must be swept across the field of view. This offers flexibility to choose the order in which epipolar planes are sampled.

Figure 3 illustrates several such ordering schemes. For instance, the ordering scheme of Figure 3c illustrates the operation of a hypothetical rolling-shutter ToF camera, where one complete image is acquired for each modulation phase. This scheme is undesirable because if the scene or camera move while acquiring these images, the recovered depth map will contain hard-to-correct errors.

A better ordering strategy is to loop through the set of modulation phases at one epipolar plane before imaging the next row (Figure 3d). Since each row's exposure time is very short, all phases required for a single row can be acquired quickly enough to minimize depth and motion blur artifacts due to camera/scene motion.

Under this strategy, each row is captured at a slightly different time. Although this induces a rolling-shutter-like effect in the acquired depth map, the individual depth values will be blur- and artifact-free and can be combined into a consistent model by post-processing [Alismail et al. 2014; Kerl et al. 2015].

To make such post-processing even easier while obeying the kinematic constraints of the mirror's actuator, we order epipolar planes in a sawtooth pattern (Figure 3e). This essentially provides full-field-of-view depth maps at twice the frame rate but half the vertical resolution, making depth correction easier for fast camera shake and/or scene motions. More generally, Figure 3f shows an example of a non-uniform sampling scheme in which epipolar planes corresponding to lower image rows are sampled more frequently. This type of sampling could be useful on a vehicle where lower portions of the field of view are usually closer and move faster, requiring acquisition at a faster sampling rate.

4 EPIPOLAR TOF PROTOTYPE

Our prototype device for epipolar time-of-flight imaging uses a galvomirror-based light sheet projector for illumination and a ToF sensor with an adjustable region of interest for imaging.

The time-of-flight sensor we use is the EPC660 (from Espros Photonics) which has a resolution of 320x240 and pixels that implement ambient saturation prevention. The sensor is fitted with a 8 mm F1.6 low distortion lens and an optical bandpass filter (650 nm center frequency, 20 nm bandwidth). The sensor allows the ROI to be changed with every sensor readout and we use this feature to select which row to image. We read data out of the sensor using the sensor development kit (DME660) from the manufacturer.

Our line projector uses a 658 nm laser diode with a peak power of 700 mW. Light from the diode is collimated and passed through a Powell lens that stretches the beam cross-section into a diverging, almost uniformly illuminated straight line with a 45° fanout angle. The laser light is directed at a 1D scanning galvomirror that can be rotated to deflect the sheet. The rotational range of the mirror gives the projector a 40° vertical field of view. The projector's effective center of projection moves as the mirror rotates but this effect can be ignored because the distance between the fanout point and the galvomirror is very small compared to depths in the scene.

A microcontroller is used to synchronize the sensor and light source. The microcontroller communicates with the sensor over an I2C bus to set the exposure time, modulation frequency/phase and region-of-interest row and also to trigger each capture. The microcontroller also actuates the projector's galvomirror. In addition, the microcontroller can read the camera's rotational velocity from a MEMS IMU (inertial magnetic unit) that we have attached to the sensor. A frequency generator circuit allows us to select a modulation frequency between 1 MHz and 24 MHz in steps of 1 MHz.

We align the projector and camera side-by-side in a rectified stereo configuration as required for epipolar imaging. When correctly aligned, the projected light sheet illuminates a single row of pixels in the camera and this row is independent of depth. A mirror calibration is performed to determine the mapping between galvomirror angle and illuminated camera row.

Sensor Calibration. In practice, we observe that the measurements read out from the sensor do not match their expected values. There are a number of reasons for this discrepancy, including fixed-pattern noise, non-uniform pixel sensitivity, crosstalk between taps and small variations in the phase of the exposure modulation function at each pixel. We model the relation between the expected sensor

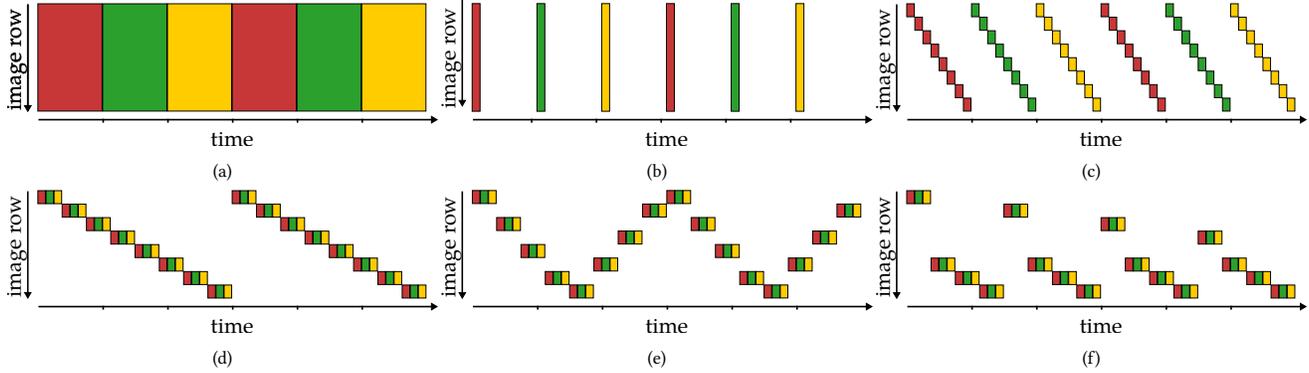


Fig. 3. Epipolar plane sampling schemes and row exposures in ToF imaging. (a) In conventional CW-ToF all epipolar planes are illuminated simultaneously and all camera rows are exposed at the same time. This requires long exposures and leads to severe artifacts due to motion, ambient light, global light transport and interference between devices. (b) Sending a very brief, high-intensity pulse of light for CW-ToF confers resistance to ambient light but is still prone to artifacts due to global light transport and motion. (c) Ordering the epipolar ToF planes similarly to a rolling-shutter camera confers robustness to ambient light, global illumination and motion blur. Sensitivity to motion remains, however, because of the significant delay between the phase measurements acquired for each row. (d) Interleaving those measurements plane by plane minimizes such artifacts. (e) Scanning the entire field of view twice within the same total exposure time yields higher temporal sampling of the scene and makes consistent merging of individual depth map rows easier. (f) For certain applications, scanning different portions of the field of view with different temporal sampling rates can be beneficial.

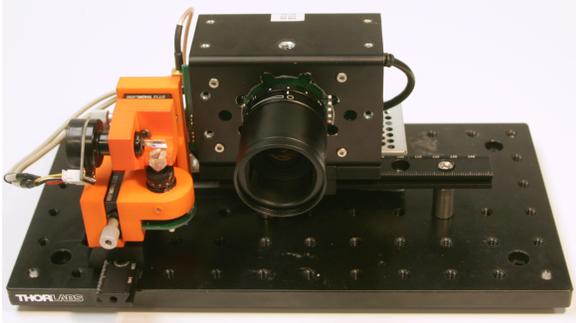


Fig. 4. Our Epipolar ToF prototype uses a custom built steerable light sheet projector (orange assembly) and a DME660 camera with fast ROI control to capture arbitrary rows of pixels.

measurements $I_\omega(x)$ and the actual measurements $\hat{I}_\omega(x)$ using a 3×3 correction matrix $H_\omega(x)$ at each pixel

$$\begin{bmatrix} I_{\omega,0} \\ I_{\omega,\frac{\pi}{2}} \\ 1 \end{bmatrix} = H_\omega(x) \begin{bmatrix} \hat{I}_{\omega,0} \\ \hat{I}_{\omega,\frac{\pi}{2}} \\ 1 \end{bmatrix} \quad (4)$$

To find $H_\omega(x)$, we place a fronto-parallel surface at a set of known distances z_k , $k = 1, \dots, K$. For each position of the plane, we collect sensor measurements at different aperture settings ($s = 1, \dots, S$) to simulate the effect of varying scene albedos. For each plane position k , we can compute the path length at a pixel $l_k(x)$ and thus the expected phase $\frac{2\omega l_k(x)}{c}$. We then compute the correction matrix that best explains the sensor measurements $I_{\omega,k,s}(x)$ by finding the matrix that minimizes the least-squares error between the corrected measurements and the expected phase.

These calibration parameters are dependent on both modulation frequency and exposure time so we repeat the process for all the frequencies and exposure times used in the experiments. Although the modulation signals we input to the sensor and light source driver are square waves, at modulation frequencies of 20 MHz and above, we noticed that the harmonics were largely suppressed and so the modulation functions were well approximated by sinusoids.

Timing. The time needed to image a row (and by extension the frame rate) with our prototype is a function of several quantities: the number of readouts per row n , the exposure time t_{exp} , the readout time for a row t_{read} and the time t_{mirror} taken by the galvo mirror to move to the next row position in the sampling sequence:

$$t_{\text{row}} = nt_{\text{exp}} + (n-1)t_{\text{read}} + \max(t_{\text{read}}, t_{\text{mirror}}) \quad (5)$$

With a two-tap sensor like the one in our prototype, at least $n = 2$ readouts are needed to measure depth using a single modulation frequency. Figure 5 shows a timing example. t_{read} is 175 μs and for most of our experiments we set t_{exp} to 100 μs . In our row sampling sequence, the mirror rotates through two rows (approximately 0.33°) per step and t_{mirror} for this step size is roughly 100 μs . In total, t_{row} is 550 μs when $n = 2$ yielding a frame rate of 7.5fps (or 3.8fps when $n = 4$).

Limitations. Currently, the main bottleneck for frame rate is readout time. Our approach needs data from only one row of the sensor per readout, but the smallest region of interest the EPC660 sensor supports is 4 rows tall. We are forced to read out 4 rows at a time when we actually use just one. Also, the development kit we have used limits the sensor data readout bus to 20 MHz but the sensor itself supports bus rates upto 80 MHz. The minimum value of t_{exp} depends on the light source's peak power and desired range. Our prototype has a source with a peak power of 700 mW while most other experimental time-of-flight systems have a peak light source

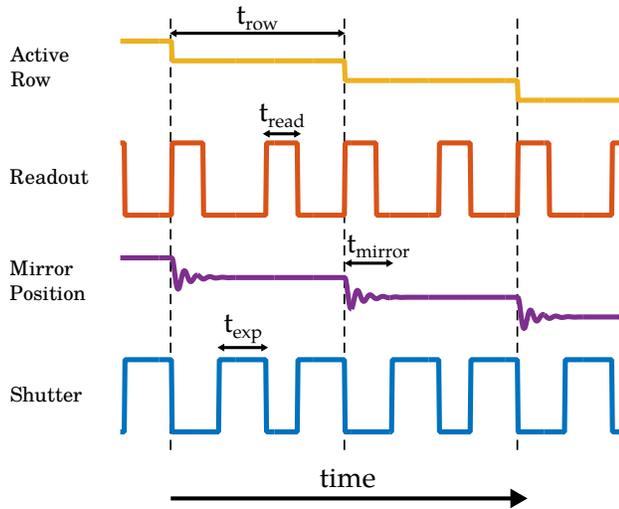


Fig. 5. Timing diagrams for camera exposure, readout and mirror position for a particular sequencing of the rows. First, the scanning mirror is moved to the new active row and takes t_{mirror} to settle in the position. When the previous row readout is complete (which takes t_{read}) and the mirror is in position, the camera is triggered. Each exposure lasts for t_{exp} and at the end of each exposure the row is read out.

power in the 3 W to 10 W range. With a brighter light source we could use a far shorter exposure time without loss of range. Lastly, the low-cost galvomirror we used could be replaced with a faster 1D MEMS mirror. With these improvements a system based on our prototype would operate at video frame rates.

The sensor used in our prototype supports a maximum modulation frequency of only 24 MHz whereas most other time-of-flight sensors can run in the 50 MHz to 100 MHz range. This limits our prototype’s ability to accurately scan smaller objects or be used for transient imaging. The EPC660 datasheet specifies that the sensor ADC returns 12-bit values but the version of the sensor we are using only returns 10 bits. This affects range and makes the output depth maps noisier.

Eye Safety. Eye safety requirements place a limit on the power that can be emitted by a CW-ToF system’s light source. This has implications for accuracy, range and frame rate. The quantity of interest in determining eye safety for a laser source is the Maximal Permissible Exposure or MPE. MPE is expressed in terms of energy or power per unit area [American National Standards Institute 2014] and is function of light source wavelength and exposure time among other factors. In our laser sheet projector, light spreads out from a spot so the power density drops as the distance from the source increases. For our current system, the energy density is safe at a distance of at least 66 cm from the source. By switching to a near-infrared (850 nm) laser, the eye safe distance of our system can be reduced to 40 cm. Details of the calculation are listed in Table 1. The laser diode source we currently use is effectively a point source. The permissible energy limits for extended light sources are considerably higher. Switching to a small extended area source such as a Vertical-Cavity Surface Emitting (VCSEL) array, would allow us to make our

prototype eye safe at shorter distances and/or extend the maximum working range.

5 RESULTS

We demonstrate the benefits of epipolar ToF imaging by comparing to regular ToF imaging in different scenes and conditions. There are two ways we could implement regular ToF imaging with our prototype sensor. The first is to remove the galvomirror and the line generator lens from the laser sheet projector and replace them with a diffuser. The second is to keep the entire sensor exposed until the sheet projector has swept across the full field of view. In the multi-device interference and camera motion experiments, we use a diffuser. For the ambient light comparisons, we use the full frame ROI approach. This prevents light loss at the diffuser from affecting our comparisons.

Ambient Light. Figure 6 shows a simulation that illustrates the benefits of applying epipolar imaging to ToF in brightly lit environments. For a given light source power, depth accuracy degrades rapidly with regular imaging as ambient light levels increase from 0 lx (complete darkness) to 100 klx (direct sunlight). With epipolar imaging, the degradation is much more gradual.

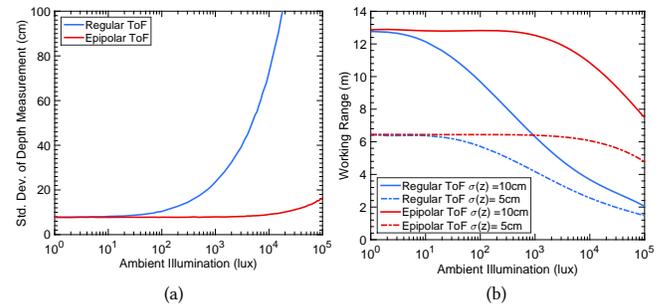


Fig. 6. A simulation of the standard deviation in depth measurements obtained using regular and epipolar ToF imaging (15 MHz modulation frequency) for a target 10 m from the camera as a function of ambient light level is shown in (a). For both cases, the peak light source power is 2 W and the total exposure time is the same (7.2 ms per image) but epipolar ToF is more robust to ambient light because it concentrates light source power and uses a short exposure for each row (30 μ s). (b) shows the working range of the same simulated camera at different levels of acceptable range accuracy. Note that simulated camera’s parameters differ from prototype, see supplementary material.

Figure 7 quantitatively compares our sensor prototype operating outdoors in regular ToF and epipolar ToF imaging modes under cloudy and sunny conditions. Regular ToF mode performs poorly in bright sunlight, while epipolar ToF is considerably more robust. Figure 8 shows an example scene with both strong ambient light and global illumination effects.

Global Illumination. Figure 9 demonstrates the ability of epipolar imaging to suppress the effects of global illumination in a few common indoor environments. These results are generated using a single modulation frequency (24 MHz). In regular ToF mode, diffuse interreflections between the walls and ceiling cause depths to be

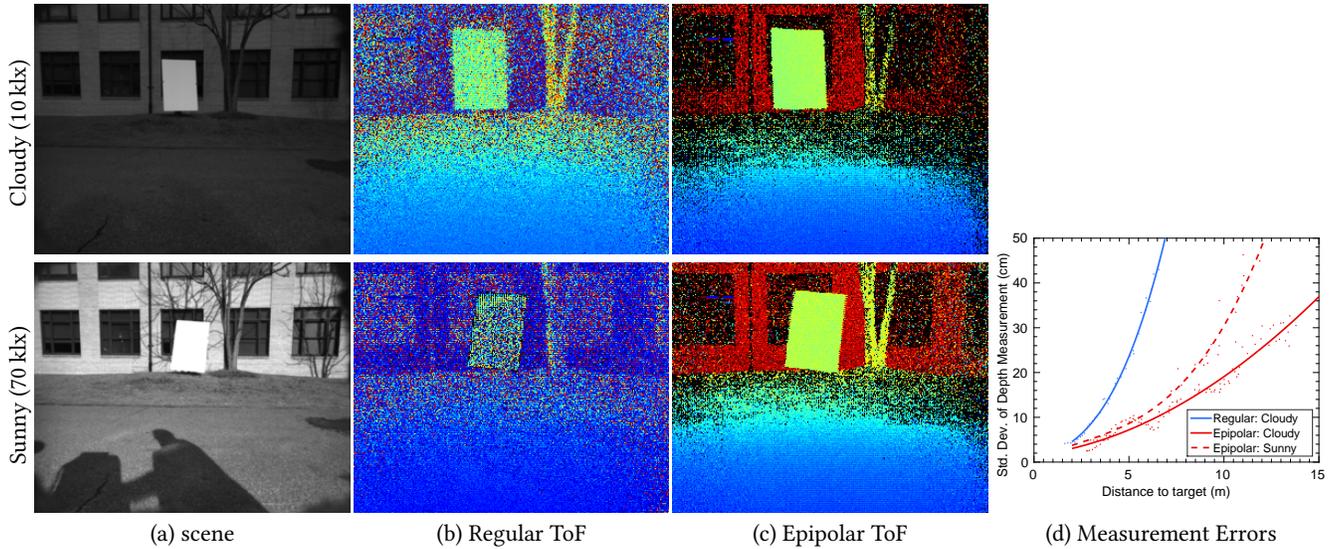


Fig. 7. We placed a white planar target at a range of distances from the sensor in both cloudy weather and bright sunshine. Even under cloudy conditions, epipolar ToF imaging produced far less noisy depth measurements than regular ToF. Under bright sunlight, regular ToF failed completely whereas epipolar ToF still provided useful depth returns. Depth maps in column (b) and (c) range from 0 m to 15 m. The camera modulation frequency was set to 10 MHz. (d) shows standard deviation in depth estimates versus distance to target (slower rising curves are better). Our prototype has depth error of around 3% at 10 m in bright sunlight.

overestimated and the corner to be rounded out. With epipolar imaging, the walls appear straight and meet at a sharp right angle. The conference table in the second row appears specular at grazing angles. In the bathroom scene, the ghosting on the wall due to reflections from the mirror is suppressed by epipolar imaging. The water fountain is particularly challenging because the direct return from its metallic surface is very weak, but the surface reflects a lot of indirect light back to the sensor. For epipolar imaging, we combine 3 exposures to try recover a useable direct signal. Longer exposures do not help regular imaging because the interreflections cause the sensor to saturate.

Multi-Camera Interference. With epipolar CW-ToF imaging, two systems running at the same modulation frequency can usually only interfere with each other at a sparse set of pixels in each image. Each system illuminates and images a single line in the scene at a time, so at any instant the second system can only interfere with the first at the points where its illuminated line intersects with the first system's exposed row of pixels. A degenerate case occurs when two systems happen to be aligned in such a way that they have identical epipolar planes and their cameras are synchronized by chance. This, however, can be considered a very rare occurrence.

If more than two epipolar ToF systems are present, each pair of cameras has only a sparse set of pixels that may be affected by interference. When a set of epipolar ToF systems are running at different modulation frequencies, the contribution of each system to shot noise in the others is greatly reduced. Figure 1 shows the result of operating two CW-ToF cameras simultaneously at the same frequency in either regular or epipolar imaging modes. In epipolar mode, the interference between the cameras is minimal. It should be

noted that the two cameras are operating completely independently of each other without any form of synchronization between them.

Camera Motion. Consider the case of a rotating camera with known rotational trajectory obtained from a MEMS gyroscope. With regular imaging, each captured ToF measurement has motion blur and strong artefacts at depth discontinuities because the measurements are not aligned to each other. This could be partially corrected using a spatially varying deconvolution but high frequencies in the image would be recovered poorly. With epipolar ToF imaging, motion blur has basically no effect and a depth map with a rolling-shutter-like effect is acquired. This can be corrected with a simple image warp computed from the rotation. Figure 1 shows an example from a rapidly panning camera. The video accompanying this paper shows an extended result.

6 DISCUSSION

Epipolar imaging for continuous-wave time-of-flight depth cameras mitigates many of the problems commonly encountered with these sensors. These problems include highly degraded performance in brightly lit conditions, systematic errors due to global illumination, errors due to inter-device interference and artifacts induced by sensor motion.

Compared to depth cameras, systems like scanning LIDAR that illuminate and image a single point at a time are very robust to all these effects but have a low measurement rate. Epipolar imaging can be thought of as a compromise between these two extremes of full-field capture and point-by-point capture. Because epipolar imaging illuminates and captures a single line at a time, it allows

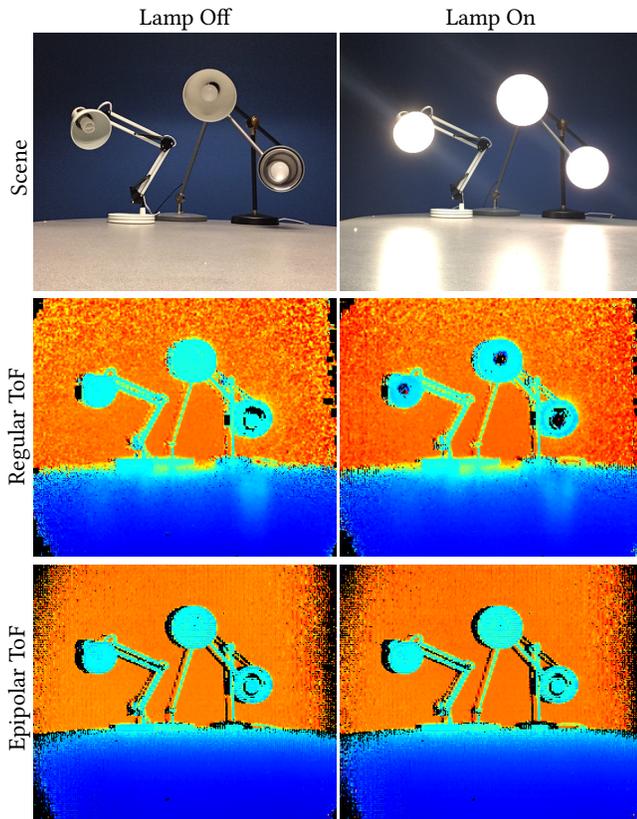


Fig. 8. Epipolar ToF imaging provides accurate depth measurements from the surface of the light bulbs even when they are turned on. Also note how reflections from the table’s surface cause errors with regular ToF, but these are suppressed with epipolar imaging.

a depth camera to have most of the robustness of point scanning while still having a high measurement rate.

Cycling through multiple phases or patterns at one row before proceeding to the next row is directly applicable to structured light as well. Such a scheme would make it possible to apply multi-image structured light methods to dynamic scenes for generating high-quality depth maps where currently only single-shot methods can be used.

In our prototype, the scanning mirror follows a sawtooth pattern and captures rows in an ordered sequence. However, with a faster scanning mirror, pseudo-random row sampling strategies could be implemented that might allow epipolar imaging to be used in conjunction with compressed sensing or similar techniques. This would allow recovery of temporally super-resolved depth maps of fast-moving scenes.

ACKNOWLEDGMENTS

This work is sponsored by the Office of Naval Research (Grant N000141512358, DURIP N000141612906), the National Aeronautics and Space Administration (Grant NNX16AD98G), the Defense Advanced Research Projects Agency (REVEAL Grant HR00111620021)

Quantity	Symbol	Formula	Value
laser peak power	P		700 mW
laser wavelength	λ		658 nm
laser spot diameter	d		0.4 cm
laser horizontal fanout	θ_x		0.785 rad
laser vertical fanout	θ_y		0.002 rad
eye-laser distance	z		
illuminated width	$w(z)$	$d + z\theta_x$	
illuminated height	$h(z)$	$d + z\theta_y$	
illuminated area	$A(z)$	$w(z)h(z)$	
exposure duration	t_{exp}		100 μs
energy per exposure	E	$\sqrt{z}Pt_{\text{exp}}$	49.5 μW
energy density limit	MPE		1.80 $\mu\text{J}/\text{cm}^2$
eye safe distance	z_s	$A(z_s) = \frac{E}{MPE}$	66 cm

Table 1. The maximal permissible exposure (MPE) is the highest safe energy density for a light source. The rules for computing MPE depend on exposure time and wavelength amongst other factors. Our current prototype is eye safe at distances greater than 66 cm. A near-infrared (850 nm) version of our current prototype would be eye safe at distances greater than 40 cm. We designed the laser sheet projector so that its horizontal fanout angle matches the camera’s horizontal field of view and the vertical fanout angle is slightly less than the angle covered by a camera pixel.

and the Natural Sciences and Engineering Research Council of Canada (NSERC) under the RPDG and SPG programs. J. Bartels was supported by NASA fellowship NNX14AM53H.

The authors would like to thank Dr. Sandu Sonoc, Laser Safety Officer at the University of Toronto for his help with the laser safety calculations.

REFERENCES

- Hatem Alismail, L Douglas Baker, and Brett Browning. 2014. Continuous trajectory estimation for 3D SLAM from actuated lidar. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 6096–6101.
- American National Standards Institute. 2014. *American National Standard for Safe Use of Lasers Z136.1*. Laser Institute of America.
- Simon Baker, Eric Bennett, Sing Bing Kang, and Richard Szeliski. 2010. Removing rolling shutter wobble. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2392–2399.
- Oliver S Cossairt, Nathan Matsuda, and Mohit Gupta. 2015. Motion Contrast 3D Scanning. In *Computational Optical Sensing and Imaging*. Optical Society of America, CT2E–1.
- A. A. Dorrington, J. P. Godbaz, M. J. Cree, A. D. Payne, and L. V. Streeter. 2011. Separating true range measurements from multi-path and scattering interference in commercial range cameras. (2011). DOI: <http://dx.doi.org/10.1117/12.876586>
- Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T Roweis, and William T Freeman. 2006. Removing camera shake from a single photograph. In *ACM Transactions on Graphics (TOG)*, Vol. 25. ACM, 787–794.
- Mohit Gupta, Shree K. Nayar, Matthias B. Hullin, and Jaime Martín. 2015. Phasor Imaging: A Generalization Of Correlation-Based Time-of-Flight Imaging. *ACM Transactions on Graphics* 34 (2015).
- Mohit Gupta, Qi Yin, and Shree K Nayar. 2013. Structured Light In Sunlight. *International Conference on Computer Vision* (2013).
- Felix Heide, Lei Xiao, Andreas Kolb, Matthias B. Hullin, and Wolfgang Heidrich. 2014. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Opt. Express* 22, 21 (Oct 2014), 26338–26350. DOI: <http://dx.doi.org/10.1364/OE.22.026338>
- Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh Raskar. 2013. Coded Time of Flight Cameras: Sparse Deconvolution to Address Multipath Interference and Recover Time Profiles. *ACM Trans. Graph.* 32, 6, Article 167 (Nov. 2013), 10 pages. DOI: <http://dx.doi.org/10.1145/2508363.2508428>

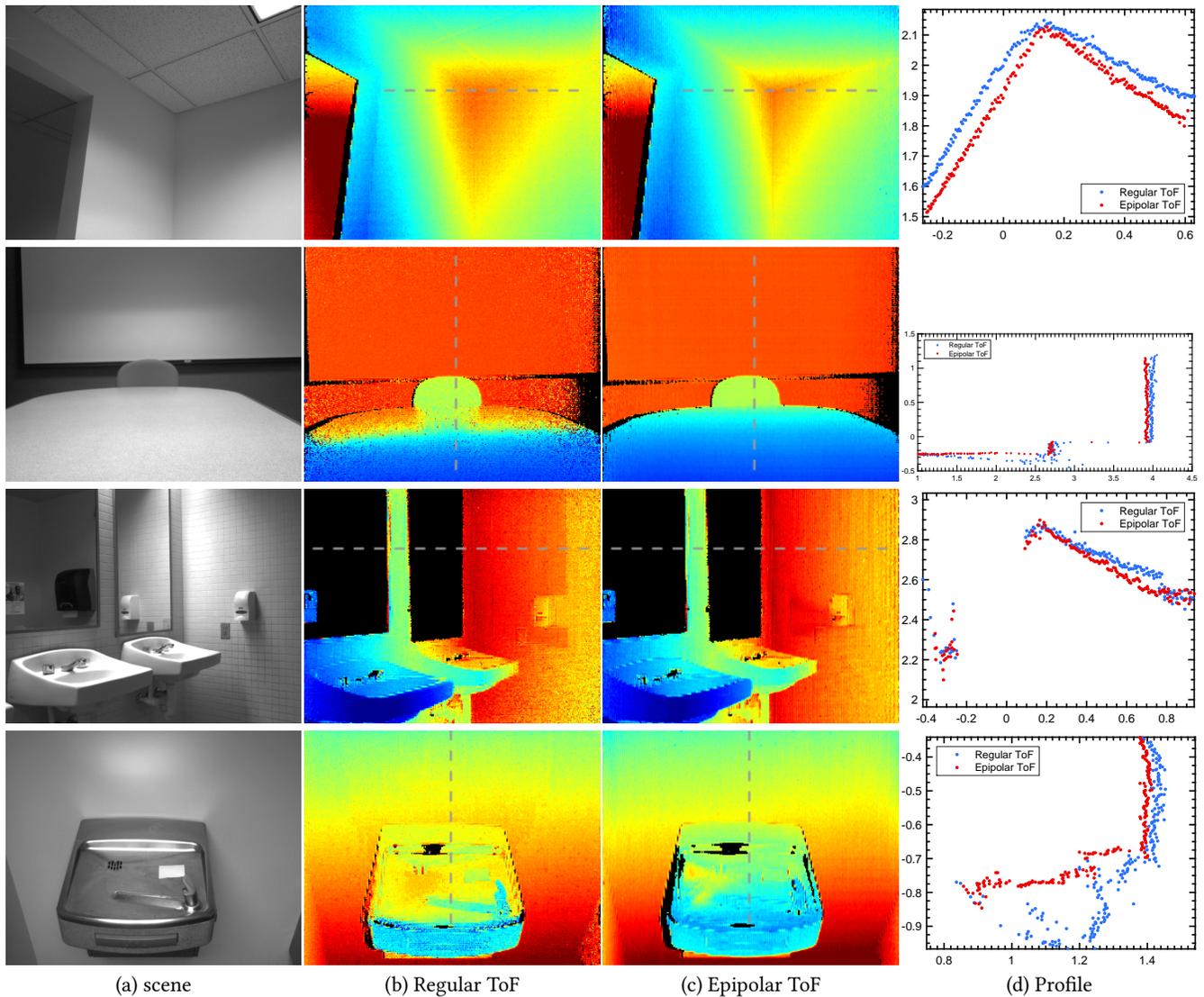


Fig. 9. Comparing depth maps with epipolar and regular ToF imaging in the presence of global light transport: diffuse interreflections at the corner, glossy interreflection from projection screen onto a shiny conference table, reflections from the mirrors in the restroom and in between the wall and the shiny water fountain. Epipolar ToF eliminates most of the global light transport resulting in depth maps that are significantly more accurate than regular ToF. All profile curves are in meters.

C. Kerl, J. Stueckler, and D. Cremers. 2015. Dense Continuous-Time Tracking and Mapping with Rolling Shutter RGB-D Cameras. In *ICCV*. Santiago, Chile.

Robert Lange and Peter Seitz. 2001. Solid-state time-of-flight range camera. *IEEE Journal of quantum electronics* 37, 3 (2001), 390–397.

Christoph Mertz, Sanjeev J Koppal, Solomon Sia, and Srinivasa Narasimhan. 2012. A low-power structured light sensor for outdoor scene reconstruction and dominant material identification. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 15–22.

Thierry Oggier, Rolf Kaufmann, Michael Lehmann, Bernhard Buttgen, Simon Neukom, Michael Richter, Matthias Schweizer, Peter Metzler, Felix Lustenberger, and Nicolas Blanc. 2005. Novel pixel architecture with inherent background suppression for 3D time-of-flight imaging. In *Electronic Imaging 2005*. International Society for Optics and Photonics, 1–8.

Matthew O’Toole, Supreeth Achar, Srinivasa G. Narasimhan, and Kiriakos N. Kutulakos. 2015. Homogeneous Codes for Energy-efficient Illumination and Imaging. *ACM*

Trans. Graph. 34, 4, Article 35 (July 2015), 13 pages. DOI: <http://dx.doi.org/10.1145/2766897>

Matthew O’Toole, Felix Heide, Lei Xiao, Matthias B. Hullin, Wolfgang Heidrich, and Kiriakos N. Kutulakos. 2014a. Temporal Frequency Probing for 5D Transient Analysis of Global Light Transport. *ACM Trans. Graph.* 33, 4, Article 87 (July 2014), 11 pages. DOI: <http://dx.doi.org/10.1145/2601097.2601103>

Matthew O’Toole, John Mather, and Kiriakos N Kutulakos. 2014b. 3d shape and indirect appearance by structured light transport. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3246–3253.

S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein. 2016. Computational Imaging with Multi-Camera Time-of-Flight Systems. *ACM Trans. Graph. (SIGGRAPH)* (2016).