MC3D: Motion Contrast 3D Scanning

Nathan Matsuda
Northwestern University
Evanston, IL

Oliver Cossairt
Northwestern University
Evanston, IL

Mohit Gupta
Columbia University
New York, NY

Abstract

Structured light 3D scanning systems are fundamentally constrained by limited sensor bandwidth and light source power, hindering their performance in real-world applications where depth information is essential, such as industrial automation, autonomous transportation, robotic surgery, and entertainment. We present a novel structured light technique called Motion Contrast 3D scanning (MC3D) that maximizes bandwidth and light source power to avoid performance trade-offs. The technique utilizes motion contrast cameras that sense temporal gradients asynchronously, i.e., independently for each pixel, a property that minimizes redundant sampling. This allows laser scanning resolution with single-shot speed, even in the presence of strong ambient illumination, significant inter-reflections, and highly reflective surfaces. The proposed approach will allow 3D vision systems to be deployed in challenging and hitherto inaccessible real-world scenarios requiring high performance using limited power and bandwidth.

1. Introduction

Many applications in science and industry, such as robotics, bioinformatics, augmented reality, and manufacturing automation rely on capturing the 3D shape of scenes. Structured light (SL) methods, where the scene is actively illuminated to reveal 3D structure, provide the most accurate shape recovery compared to passive or physical techniques [7, 33]. Here we focus on triangulation-based SL techniques, which have been shown to produce the most accurate depth information over short distances [34]. Most SL systems operate with practical constraints on sensor bandwidth and light source power. These resource limitations force concessions in acquisition speed, resolution, and performance in challenging 3D scanning conditions such as strong ambient light (e.g., outdoors) [25, 16], participating media (e.g. fog, dust or rain) [19, 20, 26, 14], specular materials [31, 27], and strong inter-reflections within the scene [15, 13, 11, 30, 4]. We propose a SL scanning architecture that overcomes these trade-offs by replacing the traditional camera with a differential motion contrast sensor to maximize light and bandwidth resource utilization.

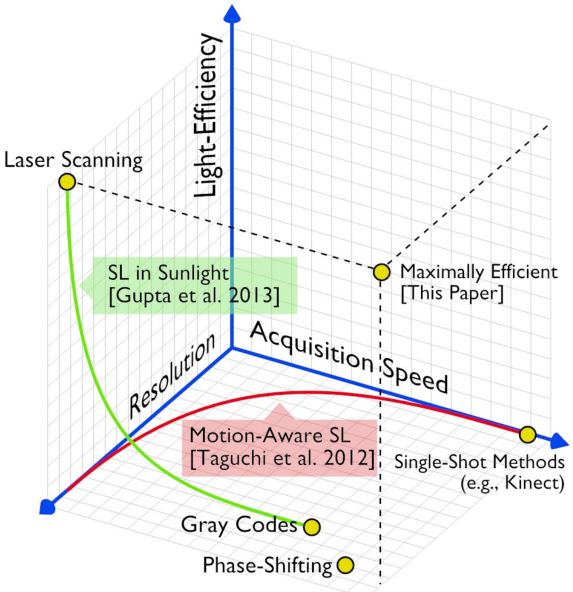


Figure 1: **Taxonomy of SL Systems:** SL systems face trade-offs in acquisition speed, resolution, and light efficiency. Laser scanning (upper left) achieves high resolution at slow speeds. Single-shot methods (mid-right) obtain lower resolution with a single exposure. Other methods such as Gray coding and phase shifting (mid-bottom) balance speed and resolution but have degraded performance in the presence of strong ambient light, scene inter-reflections, and dense participating media. Hybrid techniques from Gupta et al. [16] (curve shown in green) and Taguchi et al. [36] (curve shown in red) strike a balance between these extremes. This paper proposes a new SL method, motion contrast 3D scanning (denoted by the point in the center), that simultaneously achieves high resolution, low acquisition speed, and robust performance in exceptionally challenging 3D scanning environments.

Speed-resolution trade-off in SL methods: Most existing SL methods achieve either high resolution or high acquisition speed, but not both. This trade-off arises due to limited sensor bandwidth. On one extreme are the point/line scanning systems [5] (Figure 1, upper left), which achieve high quality results. However, each image captures only one point (or line) of depth information, thus requiring hundreds or thousands of images to capture the entire scene. Improvements can be made in processing, such as the space-time analysis proposed by Curless et al. [12] to improve accuracy and reflectance invariance, but ultimately traditional point scanning remains a highly inefficient use of camera bandwidth.

Methods such as Gray coding [32] and phase shifting [35, 15] improve bandwidth utilization but still require capturing multiple images (Figure 1, lower center). Single-shot methods [37, 38] enable depth acquisition (Figure 1, right) with a single image but achieve low resolution results. Content-aware techniques improve resolution in some cases [18, 23, 17], but at the cost of reduced capture speed [36]. This paper introduces a method achieving higher scan speeds while retaining the advantages of traditional laser scanning.

Speed-robustness trade-off: This trade-off arises due to limited light source power and is depicted by the green SL in sunlight curve in Figure 1. Laser scanning systems concentrate the available light source power in a smaller region, resulting in a large signal-to-noise ratio, but require long acquisition times. In comparison, the full-frame methods (phase-shifting, Gray codes, single-shot methods) achieve high speed by illuminating the entire scene at once but are prone to errors due to ambient illumination [16] and indirect illumination due to inter-reflections and scattering [13].

Limited dynamic range of the sensor: For scenes composed of highly specular materials such as metals, the dynamic range of the sensor is often not sufficient to capture the intensity variations of the scene. This often results in large errors in the recovered shape. Mitigating this challenge requires using special optical elements [27] or capturing a large number of images [31].

Motion contrast 3D scanning: In order to overcome these trade-offs and challenges, we make the following three observations:

> **Observation 1**: In order for the light source to be used with maximum efficiency, it should be concentrated on the smallest possible scene area. Point light scanning systems concentrate the available light into a single point, thus maximizing SNR.
>
> **Observation 2**: In conventional scanning based SL systems, most of the sensor bandwidth is not utilized. For example, in point light scanning systems, every captured image has only one sensor pixel [1] that witnesses an illuminated spot.
>
> **Observation 3**: If materials with highly specular BRDFs are present, the range of intensities in the scene often exceed the sensor's dynamic range. However, instead of capturing absolute intensities, a sensor that captures the temporal gradients of logarithmic intensity (as the projected pattern varies) can achieve invariance to the scene's BRDF.

Based on these observations, we present motion contrast 3D scanning (MC3D), a technique that simultaneously achieves the light concentration of light scanning methods, the speed of single-shot methods, and a large dynamic range. The key idea is to use biologically inspired *motion contrast sensors* in conjunction with point light scanning. The pixels on motion contrast sensors measure temporal gradients of logarithmic intensity independently and asynchronously. Due to these features, for the first time, MC3D achieves high quality results for scenes with strong specularities, significant ambient and indirect illumination, and near real-time capture rates.

Hardware prototype and practical implications: We have implemented a prototype MC3D system using off the shelf components. We show high quality 3D scanning results achieved using a single measurement per pixel, as well as robust 3D scanning results in the presence of strong ambient light, significant inter-reflections, and highly specular surfaces. We establish the merit of the proposed approach by comparing with existing systems such as Kinect [2], and binary SL. Due to its simplicity and low-cost, we believe that MC3D will allow 3D vision systems to be deployed in challenging and hitherto inaccessible real-world scenarios which require high performance with limited power and bandwidth.

2. Ambient and Global Illumination in SL

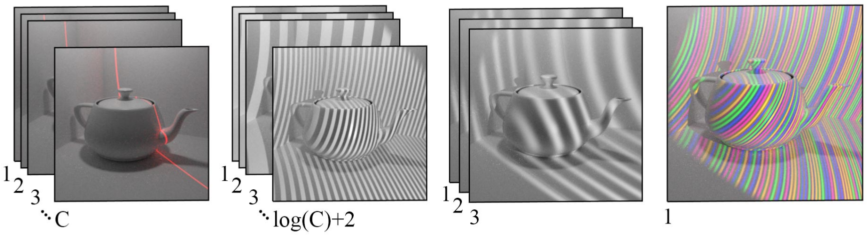
SL systems rely on the assumption that light travels directly from source to scene to camera. However, in real-world scenarios, scenes invariably receive light indirectly due to inter-reflections and scattering, as well as from ambient light sources (e.g., sun in outdoor settings). In the following, we discuss how point scanning systems are the most robust in the presence of these undesired sources of illumination.

Point scanning and ambient illumination. Let the scene be illuminated by the structured light source and an ambient light source. Full-frame SL methods (e.g., phase-shifting, Gray coding) spread the power of the structured light source over the entire scene. Suppose the brightness of the scene point due to the structured light source and ambient illumination are P and A, respectively. Since ambient illumination contributes to photon noise, the SNR of the intensity measurement can be approximated as $\frac{P}{\sqrt{A}}$ [16]. However, if the power of the structured light source is concentrated into only a fraction of the scene at a time, the effective source power increases and higher SNR is achieved. We refer to

[1] Assuming the sensor and source spatial resolutions are matched.

[2] We compare with the first-generation Kinect, which uses active triangulation depth recovery, instead of the new Kinect, which is based on Time-of-Flight.

	SDE	LCR
Point Scan	$R\times C$	1
Line Scan	C	$1/R$
Binary	$log(C)+2$	$1/(R\times C)$
Phase Shifting	3	$1/(R\times C)$
Single-Shot	1	$1/(R\times C)$



(a) Line Scan (b) Binary SL (c) Phase Shift (d) Single-Shot

Figure 2: **SL methods characterized by SPD and LER:** (a) Line scanning captures all disparity measurements in C images. (b) Binary patterns reduce the images to $log_2(C)+2$. (c) Phase shifting needs a minimum of three sinusoidal patterns. (d) Single-shot methods require only a single exposure but make smoothness assumptions that reduces resolution.

this fraction as the *Light Concentration Ratio* (LCR). The resulting SNR is given as $\frac{P}{LCR\sqrt{A}}$. Since point scanning systems maximally concentrate the light (into a single scene point), they achieve the minimum LCR and produce the most robust performance in the presence of ambient illumination for any SL system.

Point scanning and global illumination. The contributions of both direct and indirect illumination may be modeled by the light transport matrix T that maps a set of $R\times C$ projected intensities $\mathbf{p}$ from a projector onto the $M \times N$ measured intensities $\mathbf{c}$ from the camera.

$$\mathbf{c} = T\mathbf{p}. \tag{1}$$

The component of light that is directly reflected to the i^{th} camera pixel is given by $T_{i,\alpha}p_\alpha$ where the index α depends on the depth/disparity of the scene point. All other entries of T correspond to contributions from indirect reflections, which may be caused by scene inter-reflections, sub-surface scattering, or scattering from participating media. SL systems project a set of K patterns which are used to infer the index α that establishes projector-camera correspondence. For SL techniques that illuminate the entire scene at once, such as phase-shifting SL and binary SL, the sufficient condition for estimating α is that direct reflection must be greater than the sum of all indirect contributions:

$$T_{i,\alpha} > \sum_{k\neq\alpha} T_{i,k}. \tag{2}$$

For scenes with significant global illumination, this condition is often violated, resulting in depth errors [13]. For point scanning, a set of $K = R \times C$ images are captured, each corresponding to a different column $\mathbf{t_i}$ of the matrix T. In this case, a sufficient condition to estimate α is simply that direct reflection must be greater than each of the individual indirect sources of light, i.e:

$$T_{i,\alpha} > T_{i,k},\ \forall k \in \{1, \cdots R \times C\},\ k \neq \alpha. \tag{3}$$

If this condition is met, α can be found by simply thresholding each column $\mathbf{t_i}$ such that only one component remains. Since Equation 3 is a significantly less restrictive requirement than Equation 2, point scanning systems are much more robust in the presence of significant global illumination (e.g. a denser T matrix).

Sampling efficiency: While point scanning produces optimal performance in the presence of ambient and global illumination, it is an extremely inefficient sampling strategy. We define the sampling efficiency in terms of the number of pixel samples required per depth estimate (SDE). Ideally, we want $SDE = 1$, but conventional point scanning (as well as several other SL methods) captures many images for estimating depth, thus resulting in $SDE > 1$.

2.1. SDE and LCR of Existing Methods

Figure 2 compares SDE and LCR values for existing SL methods. We consider Point Scan, Line Scan (Figure 2a), Binary SL/ Gray coding (Figure 2b), Phase Shifted SL (Figure 2c), and Single-shot SL (Figure 2d). Scanning methods have small LCR but require numerous image captures, resulting in a larger SDE. Binary SL, Phase Shifted SL, and Single-shot methods require fewer images, but this is achieved by increasing LCR for each frame.

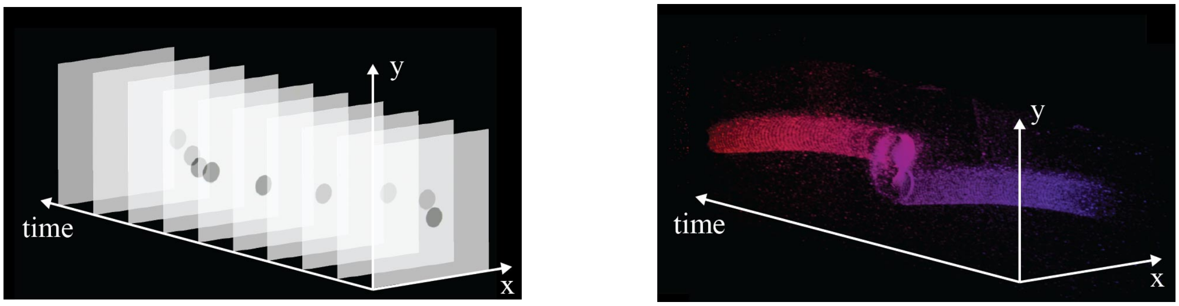
Hybrid methods: Hybrid techniques can achieve higher performance by adapting to scene content. Motion-aware SL, for example, uses motion analysis to reallocate bandwidth for either increased resolution or lower acquisition time given a fixed SDE [36]. A recent approach [16] proposes to increase LCR in high ambient lighting by increasing SDE. Hybrid methods aim to prioritize the allocation of LCR and SDE depending on scene content and imaging conditions, but are still subject to the same trade-offs as the basic SL methods.

2.2. The Ideal SL System

An ideal SL system maximizes both bandwidth and light source usage as follows:

Definition 1 *A* ***Maximally Efficient SL System*** *satisfies the constraint:*

$$SDE = 1, LCR = 1/(R \times C)$$



(a) Conventional Camera

(b) Motion Contrast Camera

Figure 3: **Conventional vs. Motion Contrast Output:** (a) The space-time volume output of a conventional camera consists of a series of discrete full frame images (here a black circle on a pendulum). (b) The output of a motion contrast camera for the same scene consists of a small number of pixel change events scattered in time and space. The sampling rate along the time axis in both cameras is limited by the camera bandwidth. The sampling rate for motion contrast is far higher because of the naturally sparse distribution of pixel change events.

Intuitively, $LCR = 1/(R \times C)$ implies the use of point scanned illumination, i.e., the structured illumination is concentrated into one scene point at a time. On the other hand, $SDE = 1$ means that each scene point is sampled only once, suggesting a single-shot method. Unfortunately, scanned illumination methods have low SDE and single-shot methods have low LCR. How can a system be both single-shot and scanning?

We reconcile this conflict by revisiting our observation that illumination scanning systems severely under-utilize camera bandwidth. Ideally, we need a sensor that measures only the scene points that are illuminated by the scanning light source. Although conventional sensors do not have such a capability, we draw motivation from biological vision where sensors that only report salient information are commonplace. Organic photoreceptors respond to changes in instantaneous contrast, implicitly culling static information. If such a sensor observes a scene lit with scanning illumination, measurement events will only occur at scene points containing the moving spot. Digital sensors mimicking the differential nature of biological photoreceptors are now available as commercially packaged camera modules. Thus, we can use these off-the-shelf components to build a scanning system that utilizes both light power and measurement bandwidth in the maximally efficient manner.

3. Motion Contrast Cameras

Lichtsteiner et al. [24] recently introduced the biologically inspired *Motion Contrast Camera*, in which pixels on the sensor independently and asynchronously generate output when they observe a temporal intensity gradient. When plotted in x, y, and time, the motion contrast output stream appears as a sparse distribution of discrete events corresponding to individual pixel changes. Figure 3b depicts the output of a motion contrast camera when viewing a black circle attached to a pendulum swinging over a white background. Note that the conventional camera view of this action, shown in Figure 3a, samples slowly along the time axis to account for bandwidth consumed by the non-moving parts of the image. For a scanning SL system, this wasted bandwidth contains measurements that provide no depth estimates, raising the SDE of the system. The motion contrast camera only makes measurements at points that are illuminated by the scanned light, enabling a SDE of 1.

For our prototype, we use the iniLabs DVS128 [24]. The camera module contains a 1st generation 128x128 CMOS motion contrast sensor, which has been used in research applications such as high frequency tracking [28], unsupervised feature extraction [8], and neurologically-inspired robotic control systems [21]. This camera has also been used to recover depth by imaging the profile of a fixed-position, pulsed laser in the context of terrain mapping [9].

The DVS128 uses event time-stamps assigned using a 100kHz counter [24]. For our 128 pixel line scanning setup this translates to a maximum resolvable scan rate of nearly 800Hz. The dynamic range of the DVS is more than 120dB due to the static background rejection discussed earlier [24].

4. Motion Contrast 3D Scanning

We now present Motion Contrast 3D scanning (MC3D). The key principle behind MC3D is the conversion of spatial projector-camera disparity to temporal events recorded by the motion contrast sensor. Interestingly, the idea of mapping disparity to time has been explored previously in the VLSI community, where several researchers have developed highly customized CMOS sensors with on-pixel circuits that record the time of maximum intensity [6, 22, 29]. The use of a motion contrast sensor in a 3D scanning system is similar to these previous approaches with two important differences: 1) The differential logarithmic nature of motion contrast cameras improves performance in the presence of ambient illumination and arbitrary scene reflectance, and 2) motion contrast cameras are currently commercially available while previous techniques required custom VLSI fabrication, limiting access to only the small number of research labs with the requisite expertise.

MC3D consists of a laser line scanner that is swept relative to a DVS sensor. The event timing from the DVS is used

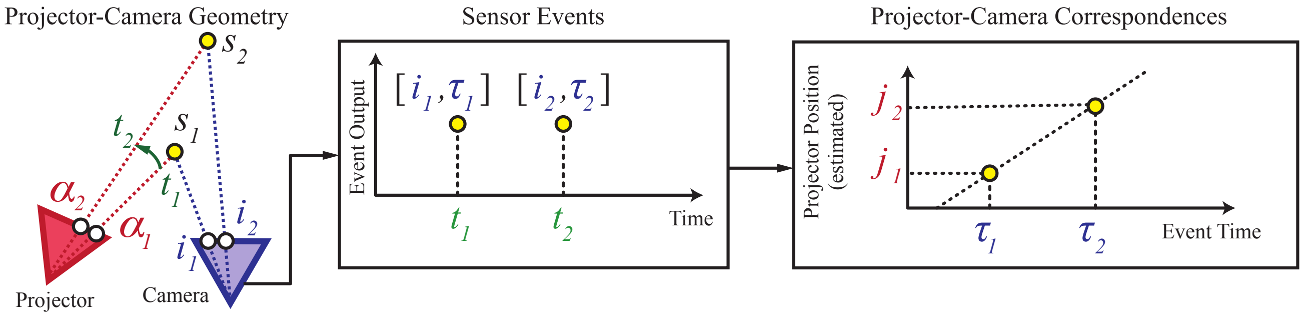


Figure 4: **System Model:** A scanning source illuminates projector positions α_1 and α_2 at times t_1 and t_2, striking scene points s_1 and s_2. Correspondence between projector and camera coordinates is not known at runtime. The DVS sensor registers changing pixels at columns i_1 and i_2 at times t_1 and t_2, which are output as events containing the location/event time pairs [i_1, τ_1] and [i_2, τ_2]. We recover the estimated projector positions j_1 and j_2 from the event times. Depth can then be calculated using the correspondence between event location and estimated projector location.

to determine scan angle, establishing projector-camera correspondence for each pixel. The DVS was used previously for SL scanning by Brandli et al. [9] in a pushbroom setup that sweeps an affixed camera-projector module across the scene. This technique is useful for large area terrain mapping but ineffective for 3D scanning of dynamic scenes. Our focus is to design a SL system capable of 3D capture for exceptionally challenging scenes, including those containing fast dynamics, significant specularities, and strong ambient and global illumination.

For ease of explanation, we assume that the MC3D system is free of distortion, blurring, and aberration; that the projector and camera are rectified and have equal focal lengths f; and are separated by a baseline b [3]. We use a 1D analysis that applies equally to all camera-projector rows. A scene point $s = (x, z)$ maps to column i in the camera image and the corresponding column α in the projector image (see Figure 4). Referring to the right side of Equation 1, after discretizing time by the index t the set of $K = R \times C$ projected patterns from a point scanner becomes:

$$P = [\mathbf{p}_1, \cdots \mathbf{p}_K] = I_0 \delta_{i,t} + I_b, \tag{4}$$

where δ is the Kronecker delta function, I_0 is the power of the focused laser beam, and I_b represents the small amount of background illumination introduced by the projector (e.g. due to scattering in the scanning optics). From Equation 1, the light intensity directly reflected to the camera is:

$$c_{i,t} = T_{i,\alpha} P_{\alpha,t} = (I_0 \delta_{\alpha,t} + I_b) T_{i,\alpha}, \tag{5}$$

where $T_{i,\alpha}$ denotes the fraction of light reflected in direction i that was incident in direction α (i.e. the BRDF) and the pair $[i, \alpha]$ represent a projector-camera correspondence. Motion contrast cameras sense the time derivative of the logarithm of incident intensity [24]:

$$\begin{aligned} c_{i,t}^{MC} &= \log(c_{i,t}) - \log(c_{i,t+1}), \qquad (6)\\ &= \log\left(\frac{I_0 + I_b}{I_b}\right) \delta_{\alpha,t}. \qquad (7) \end{aligned}$$

Next, the motion contrast intensity is thresholded and the set of space and time indices are transmitted asynchronously as tuples:

$$[i, \tau],\ s.t.\ c_{i,t}^{MC} > \epsilon, \tau = t + \sigma, \tag{8}$$

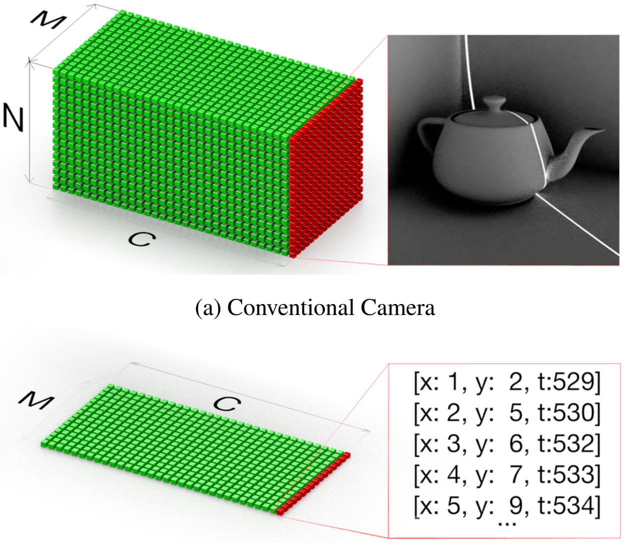
where σ is the timing noise that may be present due to pixel latency, multiple event firings, and projector timing drift. The tuples are transmitted as an asynchronous stream of events (Figure 4 middle) which establish correspondences between camera columns i and projector columns $j = \tau \cdot S$ (Figure 4 right), where S is the projector scan speed in columns/sec. The depth is then calculated as:

$$z(i) = \frac{bf}{(i - \tau \cdot S)}. \tag{9}$$

Fundamentally, MC3D is a scanning system, but it differs from conventional implementations because the motion contrast sensor implicitly culls unnecessary measurements. A conventional camera must sample the entire image for each scanned point (see Figure 5a), while the motion contrast camera samples only one pixel, drastically reducing the number of measurements required (see Figure 5b).

Independence to scene reflectivity. A closer look at Equations 5 and 7 reveal that while the intensity recorded by a conventional laser scanning system depends on scene reflectivity, MC3D does not. Strictly speaking, the equation only takes direct reflection into account, but BRDF invariance still holds approximately when ambient and global illumination are present. This feature, in combination with the logarithmic response, establishes MC3D as a much more robust technique for estimating depth of highly reflective objects, as demonstrated by the experiments shown in Figure 9.

[3] Lack of distortion, equal focal lengths, etc., are not a requirement for the system and can be accounted for by calibration.



(a) Conventional Camera

```
[x: 1, y:  2, t:529]
[x: 2, y:  5, t:530]
[x: 3, y:  6, t:532]
[x: 4, y:  7, t:533]
[x: 5, y:  9, t:534]
...
```

(b) Motion Contrast Camera

Figure 5: **Traditional vs. MC3D Line Scanning:** (a) A traditional camera capturing C scanned lines will require $M \times N \times C$ samples for a single scan. The camera data is reported in the form of 2D intensity images. (b) A motion contrast camera only reports information for each projected line and uses a bandwidth of just $M \times C$ per scan. The motion contrast output consists of an x, y, time triplet for each sample.

5. Experimental Methods and Results

DVS operation: In our system, DVS sensor parameters are set via a USB interface. In all our experiments, we maximized the built-in event rate cap to use all available bandwidth and maximized the event threshold ϵ to reject extraneous events.

Light source: We used two different sources in our prototype implementation: a portable, fixed-frequency point scanner and a variable-frequency line scanner. The portable scanner was a SHOWWX laser pico-projector from Microvision, which displays VGA input at 848x480 60Hz by scanning red, green, and blue laser diodes with a MEMS micromirror [2]. The micromirror follows a traditional raster pattern, thus functioning as a self-contained 60Hz laser spot scanner. For the variable-frequency line scanner, we used a Thorlabs GVSM002 galvanometer coupled with a Thorlabs HNL210-L 21mW HeNe Laser and a cylindrical lens. The galvanometer is able to operate at scan speeds from 0-250Hz.

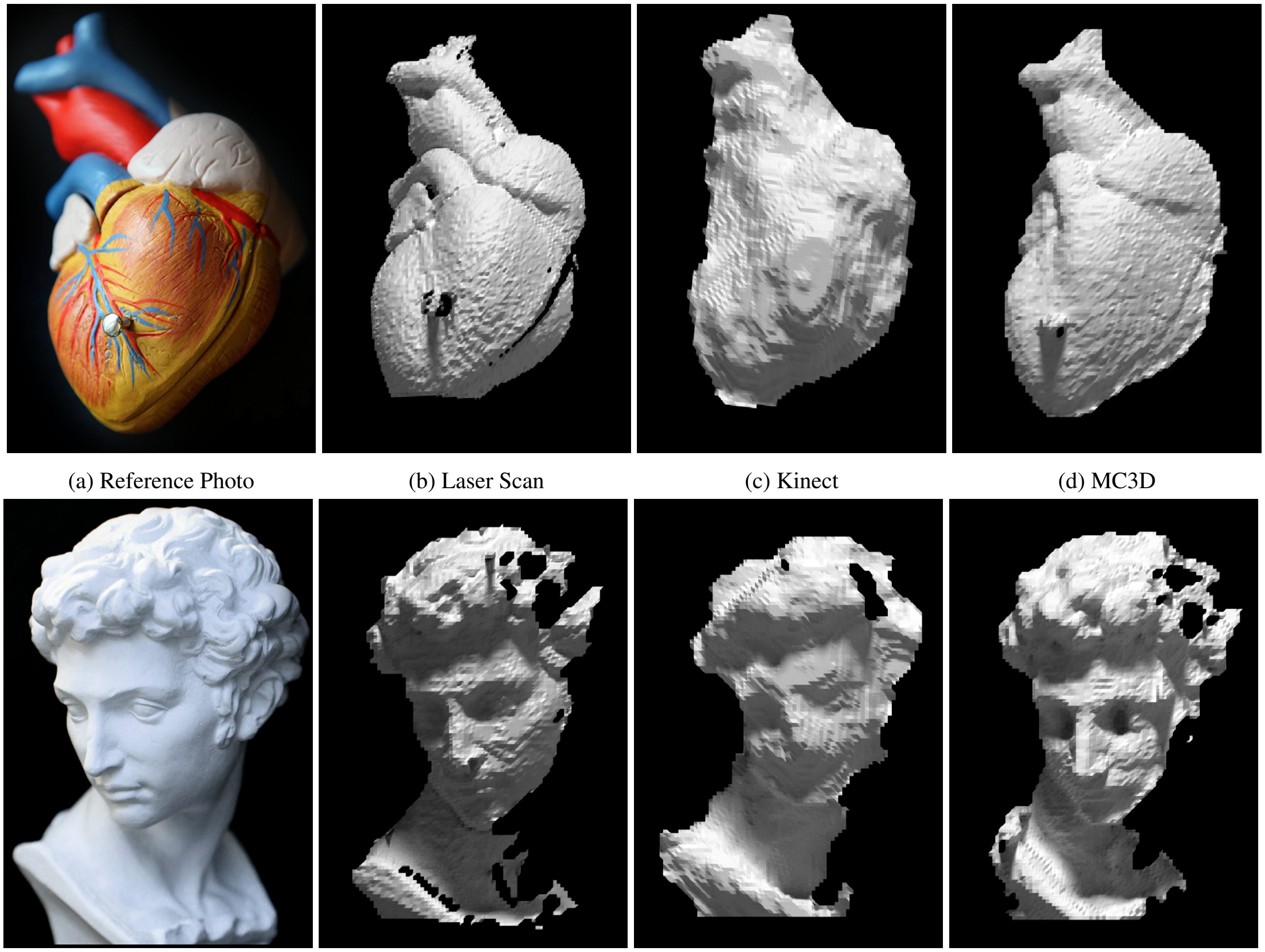
Evaluation of simple shapes: To quantitatively evaluate the performance of our system, we scanned a plane and a sphere. We placed the plane parallel to the sensor at a distance of 500 mm and captured a single scan (one measurement per pixel). Fitting an analytic plane to the result using least squares, we calculated a depth error of 7.849 mm RMSE. Similarly, for a 100 mm diameter sphere centered at 500 mm from the sensor, depth error was 12.680 mm RMSE. In both cases, $SDE = 1$ and $LCR = 1/(R \times C)$ and the SHOWWX projector was used as the source.

Evaluation of complex scenes: To demonstrate the advantages of our system in more realistic situations, we used two test objects: a medical model of a heart and a miniature plaster bust. These objects both contain smooth surfaces, fine details, and strong silhouette edges.

We captured these objects with our system and the Microsoft Kinect depth camera [1]. The Kinect is based on a single-shot scanning method and has a similar form factor and equivalent field of view when cropped to the same resolution as our prototype system. For our experimental results, we captured test objects with both systems at identical distances and lighting conditions. We fixed the exposure time for both systems at 1 second, averaging all input data during that time to produce a single disparity map. We applied a 3x3 median filter to the output of both systems. The resulting scans, shown in Figure 6, clearly show increased fidelity in our system as compared to the Kinect. The SHOWWX projector was used as the source in these experiments.

We also captured the same scenes with traditional laser scanning using the same galvanometer setup and an IDS UI348xCP-M Monochrome CMOS camera. The image was cropped using the camera's hardware region of interest to 128x128. The camera was then set to the highest possible frame rate at that resolution, or 573fps. This corresponds to a total exposure time of 28.5s, though the real world capture time was 22 minutes. Note that MC3D, while requiring several orders of magnitude less capture time than traditional laser scanning, achieves similar quality results.

Ambient lighting comparison: Figure 7 shows the performance of our system under bright ambient lighting conditions as compared to Kinect. We floodlit the scene with a broadband halogen lamp whose emission extends well into the infrared region used by the Kinect sensor. The ambient intensity was controlled by adjusting the lamp distance from the scene. Errors in the Kinect disparity map become significant even for small amounts of ambient illumination as has been shown previously [10]. In contrast, MC3D achieves high quality results for a significantly wider range of ambient illumination. The illuminance of the laser pico-projector used in this experiment is around 150 lux, measured at the object. MC3D performs well under ambient flux an order of magnitude above that of the projector. The SHOWWX projector was used as the source in these experiments, which has a listed laser power of 1mW. The Kinect, according to the hardware teardown at [3], has a 60mW laser source. The Kinect is targeted at indoor, eye-safe usage, but our experimental setup nonetheless outperforms the Kinect ambient light rejection at even lower power levels due to the light concentration advantage of laser scanning.



(a) Reference Photo (b) Laser Scan (c) Kinect (d) MC3D

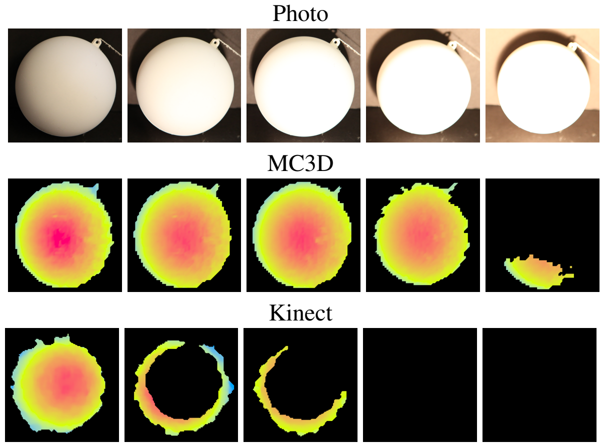
(e) Reference Photo (f) Laser Scan (g) Kinect (h) MC3D

Figure 6: **Comparison with Laser Scanning and Microsoft Kinect:** Laser scanning performed with laser galvanometer and traditional sensor cropped to 128x128 with total exposure time of 28.5s. Kinect and MC3D methods captured with 1 second exposure at 128x128 resolution (Kinect output cropped to match) and median filtered. Object placed 1m from sensor under ∼150 lux ambient illuminance measured at object. Note that while the image-space resolution for all 3 methods are matched, MC3D produces depth resolution equivalent to laser scanning, whereas the Kinect depth is more coarsely quantized.

Strong scene inter-reflections: Figure 8 shows the performance of MC3D for a scene with significant inter-reflections. The test scene consists of two pieces of white foam board meeting at a 30 degree angle. The scene produces significant inter-reflections when illuminated by a SL source. As shown in the cross-section plot on the right, MC3D faithfully recovers the V-groove of the two boards while Gray coding SL produces significant errors that grossly misrepresent the shape. The galvanometer line scanner was used as the source in these experiments.

Specular materials: Figure 9 shows the performance of MC3D for a highly specular steel sphere using the galvanometer line scanner. The reflective appearance produces a wide dynamic range that is particularly challenging for conventional SL techniques. Because MC3D senses differential motion contrast, it is more robust for scenes with a wide dynamic range. As shown in the cross-section plot on the right, MC3D faithfully recovers the spherical surface while Gray coding SL produces significant errors at the boundary and center of the sphere.

Motion comparison: We captured a spinning paper pinwheel using the SHOWWX projector to show the system's high rate of capture. Four frames from this motion sequence are shown at the top of Figure 10. Each image corresponds to consecutive 16ms exposures captured sequentially at 60fps. A Kinect capture at the bottom of the figure shows the pinwheel captured at the maximum 30fps frame rate of that sensor.



(a) 150lux (b) 500lux (c) 1000lux (d) 2000lux (e) 5000lux

Figure 7: **Output Under Ambient Illumination:** Disparity output for both methods captured with 1 second exposure at 128x128 resolution (Kinect output cropped to match) under increasing illumination from 150 lux to 5000 lux measured at middle of the sphere surface. The illuminance from our projector pattern was measured at 150lux. Note that in addition to outperforming the Kinect, MC3D returns usable data at ambient illuminance levels an order of magnitude higher than the projector power.

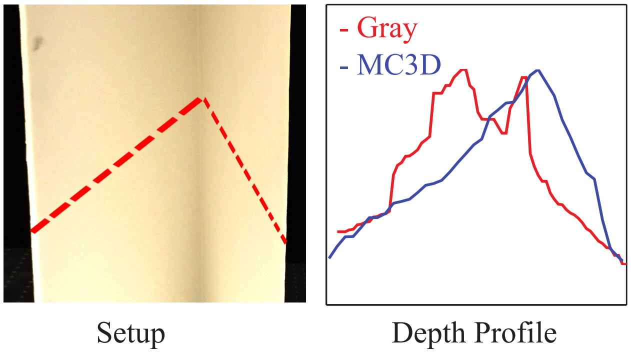


Figure 8: **Performance with Interreflections:** The image on the left depicts a test scene consisting of two pieces of white foam board meeting at a 30 degree angle. The middle row of the depth output from Gray coding and MC3D are shown in the plot on the right. Both scans were captured with an exposure time of 1/30th second. Gray coding used 22 consecutive coded frames, while MC3D results were averaged over 22 frames. MC3D faithfully recovers the V-groove shape while the Gray code output contains gross errors.

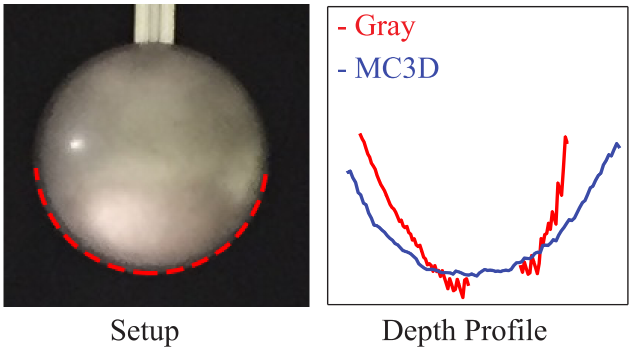
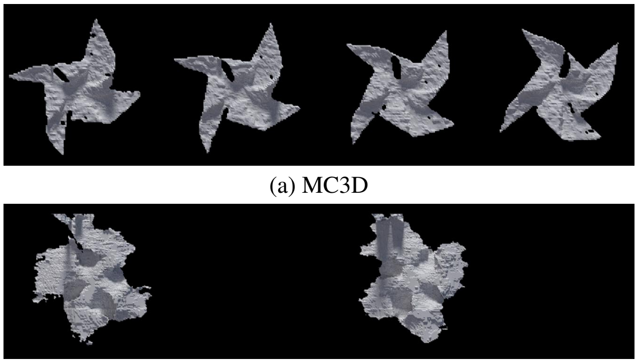


Figure 9: **Performance with Reflective Surfaces:** The image on the left depicts a reflective test scene consisting of a shiny steel sphere. The plot on the right shows the depth output from Gray coding and MC3D. Both scans were captured with an exposure time of 1/30th second. The Gray coding method used 22 consecutive coded frames, while MC3D results were averaged over 22 frames. The Gray code output produces significant artifacts not present in MC3D output.



(a) MC3D

(b) Kinect

Figure 10: **Motion Comparison:** The top row depicts 4 frames of a pinwheel spinning at roughly 120rpm, captured at 60fps using MC3D. The bottom row depicts the same pinwheel spinning at the same rate, over the same time interval captured with the Kinect. Only 2 frames are shown due to the 30fps native frame rate of the Kinect. **Please see movies of our real-time 3D scans in Supplementary Materials.**

6. Discussion and Limitations

We have introduced MC3D, a new approach to SL that eliminates redundant sampling of irrelevant pixels and maximizes laser scanning speed. This arrangement retains the light efficiency and resolution advantages of laser scanning while attaining the real-time performance of single-shot methods.

While our prototype system compares favorably against Kinect and Gray coding, it falls short of achieving laser scan quality. This is mostly due to the relatively small resolution (128×128) of the DVS and is not a fundamental limitation. The DVS used in our experiments is the first commercially available motion contrast sensor. Subsequent versions are expected to achieve higher resolution, which will enhance the quality of the results achieved by our technique. Furthermore, we intend to investigate superresolution techniques to improve spatial resolution.

There are several noise sources in our prototype system such as uncertainty in event timing due to internal electrical characteristics of the sensor, multiple event firings during one brightness change event, or downsampling in the

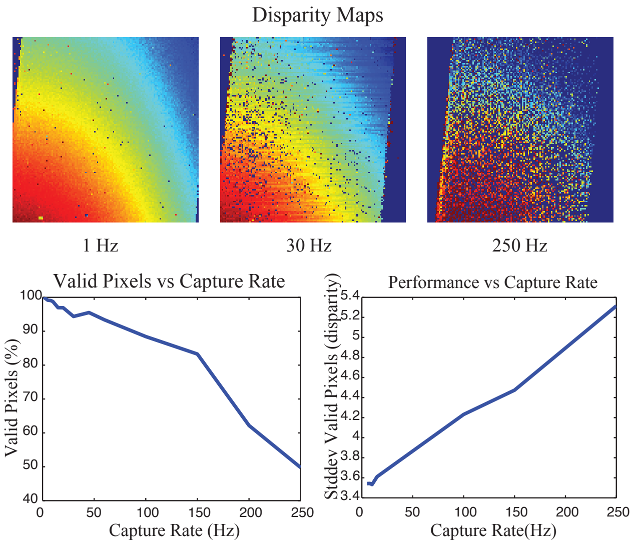


Figure 11: **MC3D Performance vs Scan Rates:** The row of images depict the disparity output from a single sweep of the laser at 1hz, 30hz, and 250hz. Bottom left, the number of valid pixels recovered on average for one scan at different scan rates decreases with increasing scan frequency. Bottom right, the standard deviation of the depth map increases with increasing scan frequency.

sensors digital interface. The trade-off between noise and scan speed is investigated in Figure 11. As scan speed increases, timing errors are amplified, resulting in an increased amount of dropped events (bottom-left), which degrades the quality of recovered depth maps (bottom-right). These can be mitigated through updated sensor designs, further system engineering, and more sophisticated point cloud processing. We plan to provide a thorough noise analysis in a future publication.

Despite limitations, our hardware prototype shows that this method can be implemented using off-the-shelf components with minimal system integration. The results from this prototype show promise in outperforming existing commercial single-shot SL systems, especially in terms of both speed and performance. Improvements are necessary to develop single-shot laser scanning into a commercially viable product, but nonetheless our simple prototype demonstrates that the MC3D concept has clear benefits over existing methods for dynamic scenes, highly specular materials, and strong ambient or global illumination.

7. Acknowledgments

This work was supported by funding through the Biological Systems Science Division, Office of Biological and Environmental Research, Office of Science, U.S. Dept. of Energy, under Contract DE-AC02-06CH11357. Additionally, this work was supported by ONR award number 1(GG010550)//N00014-14-1-0741.

References

[1] Microsoft Kinect. http://www.xbox.com/kinect. 6

[2] Microvision SHOWWX. https://web.archive.org/web/20110614205539/http://www.microvision.com/showwx/pdfs/showwx_userguide.pdf. 6

[3] Openkinect hardware info. http://openkinect.org/wiki/Hardware_info. 6

[4] S. Achar and S. G. Narasimhan. Multi Focus Structured Light for Recovering Scene Shape and Global Illumination. In *ECCV*, 2014. 1

[5] G. J. Agin and T. O. Binford. Computer description of curved objects. *IEEE Transactions on Computers*, 25(4), 1976. 1

[6] K. Araki, Y. Sato, and S. Parthasarathy. High speed rangefinder. In *Robotics and IECON'87 Conferences*, pages 184–188. International Society for Optics and Photonics, 1988. 4

[7] P. Besl. Active, optical range imaging sensors. *Machine vision and applications*, 1(2), 1988. 1

[8] O. Bichler, D. Querlioz, S. J. Thorpe, J.-P. Bourgoin, and C. Gamrat. Unsupervised features extraction from asynchronous silicon retina through spike-timing-dependent plasticity. *Neural Networks (IJCNN), International Joint Conference on*, 2011. 4

[9] C. Brandli, T. A. Mantel, M. Hutter, M. A. Höpflinger, R. Berner, R. Siegwart, and T. Delbruck. Adaptive pulsed laser line extraction for terrain reconstruction using a dynamic vision sensor. *Frontiers in neuroscience*, 7, 2013. 4, 5

[10] D. Castro and Mathur. Kinect outdoors. www.youtube.com/watch?v=rI6CU9aRDIo. 6

[11] V. Couture, N. Martin, and S. Roy. Unstructured light scanning robust to indirect illumination and depth discontinuities. *IJCV*, 108(3), 2014. 1

[12] B. Curless and M. Levoy. Better optical triangulation through spacetime analysis. In *IEEE ICCV*, 1995. 2

[13] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan. A practical approach to 3D scanning in the presence of interre.ections, subsurface scattering and defocus. *IJCV*, 102(1-3), 2012. 1, 2, 3

[14] M. Gupta, S. G. Narasimhan, and Y. Y. Schechner. On controlling light transport in poor visibility environments. In *IEEE CVPR*, pages 1–8, June 2008. 1

[15] M. Gupta and S. K. Nayar. Micro phase shifting. *IEEE CVPR*, 2012. 1, 2

[16] M. Gupta, Q. Yin, and S. K. Nayar. Structured light in sunlight. *IEEE ICCV*, 2013. 1, 2, 3

[17] K. Hattori and Y. Sato. Pattern shift rangefinding for accurate shape information. *MVA*, 1996. 2

[18] E. Horn and N. Kiryati. Toward optimal structured light patterns. *Image and Vision Computing*, 17(2), 1999. 2

[19] J. S. Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2), 1990. 1

[20] J. S. Jaffe. Enhanced extended range underwater imaging via structured illumination. *Optics Express*, (12), 2010. 1

[21] A. Jimenez-Fernandez, J. L. Fuentes-del Bosh, R. Paz-Vicente, A. Linares-Barranco, and G. Jimenez. Neuro-inspired system for real-time vision sensor tilt correction. In *IEEE ISCAS*, 2010. 4

[22] T. Kanade, A. Gruss, and L. R. Carley. A very fast VLSI rangefinder. In *IEEE ICRA*, pages 1322–1329, 1991. 4

[23] T. P. Koninckx and L. Van Gool. Real-time range acquisition by adaptive structured light. *IEEE PAMI*, 28(3), 2006. 2

[24] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128× 128 120 db 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 43(2), 2008. 4, 5

[25] C. Mertz, S. J. Koppal, S. Sia, and S. Narasimhan. A low-power structured light sensor for outdoor scene reconstruction and dominant material identification. *IEEE International Workshop on Projector-Camera Systems*, 2012. 1

[26] S. G. Narasimhan, S. K. Nayar, B. Sun, and S. J. Koppal. Structured light in scattering media. In *IEEE ICCV*, 2005. 1

[27] S. K. Nayar and M. Gupta. Diffuse structured light. In *IEEE ICCP*, 2012. 1, 2

[28] Z. Ni, A. Bolopion, J. Agnus, R. Benosman, and S. Régnier. Asynchronous event-based visual shape tracking for stable haptic feedback in microrobotics. *Robotics, IEEE Transactions on*, 28(5), 2012. 4

[29] Y. Oike, M. Ikeda, and K. Asada. A CMOS image sensor for high-speed active range finding using column-parallel time-domain ADC and position encoder. *IEEE Transactions on Electron Devices*, 50(1):152–158, 2003. 4

[30] M. O'Toole, J. Mather, and K. N. Kutulakos. 3D Shape and Indirect Appearance By Structured Light Transport. In *IEEE CVPR*, 2014. 1

[31] J. Park and A. Kak. 3d modeling of optically challenging objects. *IEEE TVCG*, 14(2), 2008. 1, 2

[32] J. Posdamer and M. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, 18(1), 1982. 2

[33] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8), 2010. 1

[34] R. Schwarte. *Handbook of Computer Vision and Applications, chapter Principles of 3-D Imaging Techniques*. Academic Press, 1999. 1

[35] V. Srinivasan, H.-C. Liu, and M. Halioua. Automated phase-measuring profilometry: a phase mapping approach. *Applied Optics*, 24(2), 1985. 2

[36] Y. Taguchi, A. Agrawal, and O. Tuzel. Motion-aware structured light using spatio-temporal decodable patterns. *ECCV*, 2012. 1, 2, 3

[37] L. Zhang, B. Curless, and S. M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. *International Symposium on 3D Data Processing Visualization and Transmission*, 2002. 2

[38] S. Zhang, D. V. D. Weide, and J. Oliver. Superfast phase-shifting method for 3-D shape measurement. *Optics Express*, 18(9), 2010. 2