Structured Light In Sunlight

Mohit Gupta
Columbia University
New York, NY 10027
mohitg@cs.columbia.edu

Qi Yin
Columbia University
New York, NY 10027
qiyin@cs.columbia.edu

Shree K. Nayar
Columbia University
New York, NY 10027
nayar@cs.columbia.edu

Abstract

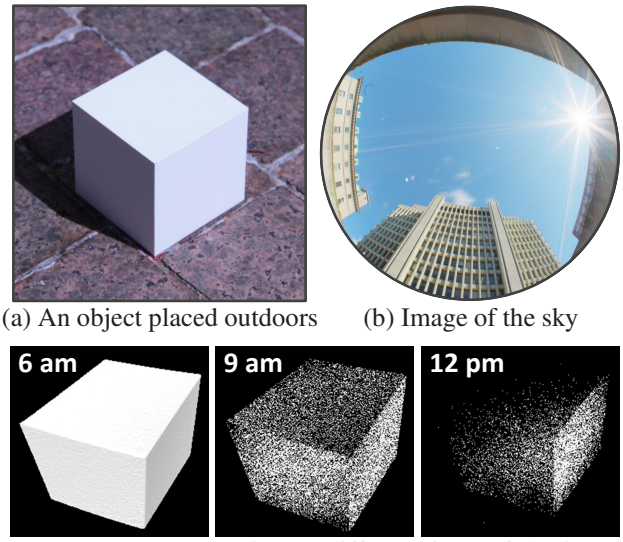
Strong ambient illumination severely degrades the performance of structured light based techniques. This is especially true in outdoor scenarios, where the structured light sources have to compete with sunlight, whose power is often 2-5 orders of magnitude larger than the projected light. In this paper, we propose the concept of light-concentration to overcome strong ambient illumination. Our key observation is that given a fixed light (power) budget, it is always *better to allocate it sequentially in several portions of the scene, as compared to spreading it over the entire scene at once. For a desired level of accuracy, we show that by distributing light appropriately, the proposed approach requires 1-2 orders lower acquisition time than existing approaches. Our approach is illumination-adaptive as the optimal light distribution is determined based on a measurement of the ambient illumination level. Since current light sources have a fixed light distribution, we have built a prototype light source that supports flexible light distribution by controlling the scanning speed of a laser scanner. We show several high quality 3D scanning results in a wide range of outdoor scenarios. The proposed approach will benefit 3D vision systems that need to operate outdoors under extreme ambient illumination levels on a limited time and power budget.*

1. Introduction

Structured light 3D scanning, because of its accuracy and simplicity, is the method of choice for 3D reconstruction in several applications, including factory automation, performance capture, digitization of cultural heritage and autonomous vehicles. In many real-world settings, structured light sources have to compete with strong ambient illumination. In these scenarios, because of the limited dynamic range of image sensors, the signal (intensity due to structured light) in the captured images can be extremely low, resulting in poor 3D reconstructions.

This is especially true outdoors, where sunlight is often 2-5 orders of magnitude brighter than the projected structured light. For instance, it is known that Kinect, a popular structured light device, cannot recover 3D shape in strong sunlight [1]. While several optical techniques for ambient light reduction have been proposed [10], they achieve only moderate success. An example using an off-the-shelf laser 3D scanner is shown in Figure 1. Under strong ambient illumination, the reconstruction quality of an object placed outdoors degrades, even when spectral filtering is used.

One obvious solution to the ambient light problem is to increase the power of the light source. Unfortunately, this is not always possible. Especially in outdoor scenarios, vision systems often operate on a limited power budget. Moreover, low-cost hand-held projectors (e.g., pico projectors) are increasingly becoming popular as structured light sources. For these low-power devices to be useful outdoors, it is important to be able to handle strong ambient illumination on a tight power budget.



(a) An object placed outdoors (b) Image of the sky

(c-e) 3D reconstructions at different times of the day

Figure 1. **Effect of ambient illumination on structured light 3D scanning.** (a) An object placed outdoors on a clear day receives strong ambient illuminance R_a from the sun and the sky. (b) Image of the sky at 9am. (c-e) 3D reconstructions using conventional methods at different times of the day. From left to right, as the day progresses, R_a increases (2000 lux, 24,000 lux and 90,000 lux, respectively) and the reconstruction quality degrades.

In this paper, we introduce the concept of *light concentration* in order to deal with strong ambient illumination. The key idea is that even with a small light budget, signal level can be increased by concentrating the available projector light on a small portion of the scene. This is illustrated in Figure 2 (center). At first glance, it may appear that concentrating the light will require more measurements, as only a fraction of the scene is illuminated and encoded at a time. However, we show that, it is possible to achieve significantly lower acquisition times by concentrating light as compared to existing approaches that spread the available light over the entire projector image plane, and then reduce image noise by frame-averaging. We call this the *light-concentration advantage*.

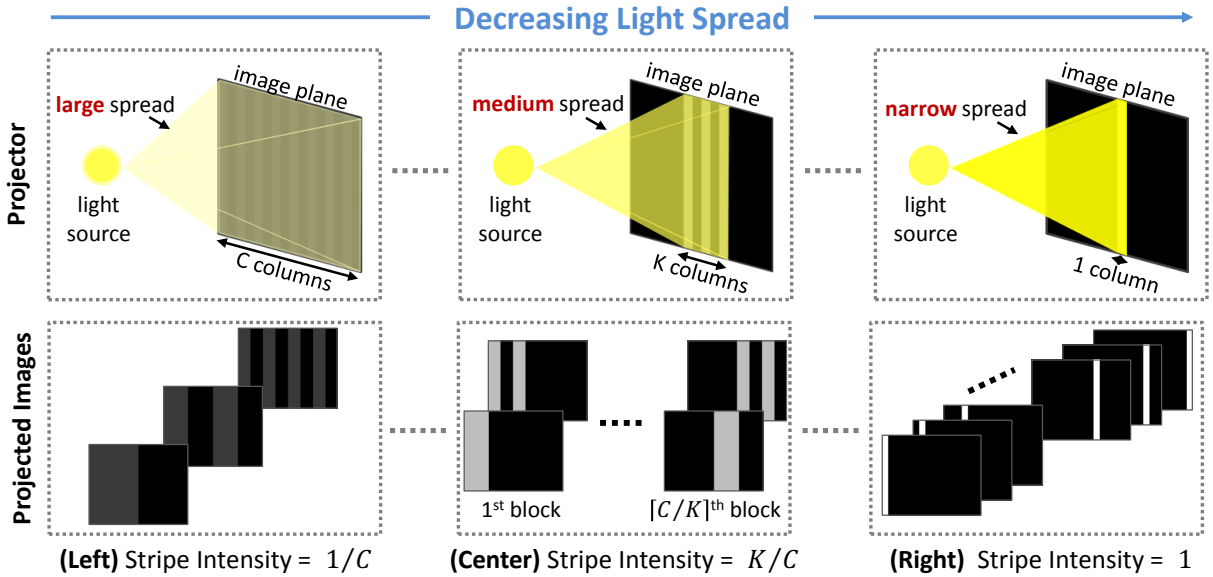


Figure 2. **Light redistribution for structured light.** We consider different light distributions for designing structured light systems that perform under strong ambient illumination. Given a fixed light budget, as the light spread decreases (from left to right), the intensity of each projected stripe increases. Existing structured light techniques lie at the two extremes of the power distribution scale. (Left) Systems where light is distributed over the entire projector image plane yield low signal strength and hence poor reconstruction quality. (Right) Systems where all the light is concentrated in to a single column require a large number of measurements. (Center) We show that by concentrating the light appropriately, it is possible to achieve fast and high-quality 3D scanning even in strong ambient illumination.

The light-concentration advantage arises from the fact that frame-averaging increases signal-to-noise-ratio (SNR) by a factor of *square-root* of the number of averaged frames. However, the same time and power budget, if allocated into smaller scene regions, increases SNR *linearly* with the number of measurements. We show that for the same accuracy (SNR) level, while the number of measurements required by existing approaches is linear in the ambient illuminance level R_a, i.e., $\mathcal{O}(R_a)$, the proposed approach requires only $\mathcal{O}(\sqrt{R_a})$ measurements. For outdoor ambient illuminance levels, this translates into 1-2 orders of magnitude (10-100 times) lower acquisition time.

Scope and contributions: This paper introduces light redistribution as a new dimension in the design of structured light systems. We do not introduce a new structured light coding scheme. Instead, we show that by managing the light budget appropriately, it is possible to perform fast and accurate 3D scanning outdoors on a limited power budget. After determining the optimal light distribution based on the ambient illuminance level, any one of the existing high-SNR structured light coding scheme [14, 4, 6] can be used. The proposed approach can *adapt* to the ambient light level. For instance, as ambient illuminance decreases, the acquisition time required by our approach decreases. The proposed techniques are not restricted only to ambient illumination due to sunlight. They are applicable in any scenario that has a wide range of ambient illumination.

Hardware prototype and practical implications: Existing projectors distribute light over the entire image plane; they do not have the ability to distribute light in a flexible manner. We have developed a prototype projector with flexible light distribution ability by using an off-the-shelf laser scanner. Different light distributions over the projector image plane are achieved by varying the scanning speed of the scanner. The proposed approach achieves fast and high-quality (pixel-level) 3D reconstruction for even the most extreme scenarios (direct sunlight, low-powered light source). These features make our approach especially suitable for moving platforms such as autonomous cars which need to operate outdoors under varying ambient illuminance levels on a limited power budget.

2. Related Work

Structured light 3D scanning: Structured light techniques are classified based on the coded patterns that they project on the scene. Some typical examples are single line stripes [3], sinusoidal patterns [13], binary patterns [11] and deBruijn codes [15]. For a comprehensive survey on existing coding schemes, the reader is referred to [12].

Significant work has been done towards designing high SNR structured light coding schemes [14, 4, 6]. It has been shown that in scenarios with extremely low SNR (such as strong ambient illumination), optimal SNR is achieved by using patterns with the fewest possible intensity levels (binary patterns with two intensity levels) [6]. In Figure 1, despite binary Gray code patterns being used, result quality degrades as ambient illumination increases. This is because using high SNR patterns without considering light redistribution is not sufficient to achieve high-quality results under strong ambient illumination.

Optical methods for suppressing ambient illumination: Examples of such methods include using a narrow spectral bandwidth laser (sunlight has broad bandwidth) with a narrow-band spectral filter [10] and a polarized light source (sunlight is unpolarized) with a polarization filter [10].

This paper proposes a different approach. Given a fixed level of ambient illuminance (after optical suppression), we

determine the optimal distribution of the light (of the structured light source) in order to maximize the SNR. The increase in SNR achieved by our method is in addition to, and much higher than, that achieved by the optical methods. In order to deal with extreme ambient illumination scenarios, optical suppression techniques can be used in a complementary manner to our method.

Recently, a pulsed light source with a fast shutter [8] was used to suppress ambient illumination. Our approach is inspired by this work, which corresponds to the right extreme of the light distribution scale in Figure 2. In this method, all the light is concentrated into a single column. While effective, it requires a large number of images. In contrast, we consider the entire range of light distributions. Given the same power budget, our method, by distributing the available light efficiently, requires 10-100 times fewer images in most outdoor scenarios.

3. Structured Light In Ambient Illumination

We model the structured light source L as a projector that has an image plane with C columns. The projector projects spatio-temporally coded patterns on the scene so that a unique intensity code is assigned to each column [1]. The power of the light source is fixed at P Watts. If the available power is spread equally among all C columns, each column generates $\frac{P}{C}$ Watts of light. This is illustrated in Figure 2 (left).

The intensity of a scene point S in a captured image is:

$$I = I_l + I_a + \eta, \tag{1}$$

where I_l and I_a are intensities corresponding to the light source L and ambient illumination A, respectively. η is the camera noise. The goal is to extract the signal component I_l reliably from the captured images. The accuracy of the estimated signal $\hat{I}_l$ (and hence the depth-accuracy) is proportional to the signal-to-noise-ratio: $SNR = \frac{I_l}{\eta}$.

3.1. Ambient illumination and depth accuracy

The components I_l and I_a are proportional to the illuminance values R_l and R_a at scene point S due to the light source L and ambient illumination A, respectively:

$$I_l = \alpha R_l, \quad I_a = \beta R_a, \tag{2}$$

where α and β encapsulate the scene point's BRDF, light fall-off, and camera's spectral gain [2]. R_l is proportional to the source power P. We assume the affine camera noise model, with both signal-dependent and signal-independent terms [5]:

$$\eta^2 = \sigma_r^2 + \frac{\alpha R_l + \beta R_a}{g}, \tag{3}$$

where σ_r is the standard deviation of the signal-independent sensor read noise, and g is camera gain. In scenarios with strong ambient illumination, $R_a >> R_l$, and the dominant source of noise is the signal-dependent photon noise, i.e., $\sigma_r^2 << \frac{\beta R_a}{g}$. Then, the SNR is approximated as:

$$SNR \approx \lambda \frac{R_l}{\sqrt{R_a}}, \tag{4}$$

where λ is a constant. In order to achieve a desired depth accuracy δ, the SNR should be higher than a threshold τ, i.e., $SNR > \tau$ [3]. Substituting in Eq. 4:

$$\frac{R_l}{\sqrt{R_a}} \geq \frac{\tau}{\lambda}. \tag{5}$$

We call this the *decodability condition*. In order to achieve the desired depth accuracy, all the captured images must satisfy the decodability condition.

If R_a is significantly larger than R_l, the decodability condition is not satisfied. This results in large errors in the recovered shape, as illustrated in Figure 1. As R_a increases, the quality of the reconstructed shape deteriorates.

3.2. Increasing SNR by multi-frame capture

A common technique for increasing SNR is by capturing multiple frames per image[4] and combining them into a single image. For instance, by capturing f frames $[F_1, \ldots, F_f]$ for each image I, and computing the average image $\hat{I} = \frac{\sum_i F_i}{f}$, noise can be reduced by a factor of $\sqrt{f}$.

How many frames should be combined so that the decodability condition is satisfied? Using Eq. 4, the SNR for an image computed by averaging f frames is $SNR_{av} = \sqrt{f}\lambda\frac{R_l}{\sqrt{R_a}}$. Since SNR_{av} should be greater than τ, we get:

$$f \geq \left(\frac{\tau}{\lambda R_l}\right)^2 R_a \ . \tag{6}$$

Let N_C be the number of images required by the particular structured light coding scheme used to encode all the projector columns uniquely, and f, as defined above, is the number of frames to be averaged per image. Then, the total number of measurements M is given as:

$$M = N_C \times f \ . \tag{7}$$

From Eqs. 6 and 7, we arrive at the following result:

Result 1 (Acquisition time for frame averaging) *Given a fixed power budget P, the number of measurements M (and hence the acquisition time) using frame-averaging* is linear *in the ambient illuminance level R_a, i.e., $\mathcal{O}(R_a)$.*

Thus, while frame-averaging can be an effective method for increasing SNR in weak ambient illumination (e.g., indoors), the acquisition time is prohibitively large for outdoor ambient illumination levels that are $10^2 - 10^3$ times the typical indoor illumination.

In view of this tradeoff between desired accuracy and acquisition time, we ask the following question: Is it possible

[1] Because of epipolar geometry between the projector and camera, only 1D coding (e.g., along the columns) on the projector plane is sufficient to perform depth recovery using triangulation. In the rest of the paper, all the pixels within a column are grouped together as a single entity - a column.

[2] β also includes the effect of any optical (e.g., spectral or polarization) filtering used for reducing ambient illumination. In all our experiments, we used a narrow-band laser light source and spectral filter in front of our camera. This suppresses ambient illumination by a factor of about 20.

[3] The threshold τ depends on the structured light coding and decoding algorithms. It increases monotonically with δ. The analytical expressions for λ and τ are derived in a technical report available at [2].

[4] In this paper, we distinguish images from frames. Images correspond to measurements captured under different illumination patterns. Frames are measurements captured under the same illumination pattern. Multiple frames may be combined to compute a single image.

to achieve high depth accuracy while also requiring a small number of measurements, even in extremely strong ambient light conditions and with a limited power budget?

4. The Light-Concentration Advantage

Suppose the SNR needs to be increased by a factor of s in order to satisfy the decodability condition. Our key observation is that SNR can be improved much more efficiently as compared to frame-averaging by *concentrating* light into a smaller region of the scene. This is different from blocking the light, which results in light-loss. The total light budget remains the same - it is just concentrated into a smaller region. This is illustrated in Figure 2 (center).

In particular, let the projector image plane be divided into s blocks of size $K = \frac{C}{s}$ columns each. Suppose all the available light is concentrated into a single block at a time, and each block is encoded independently. We call this the *concentrate-and-scan* strategy, as light is concentrated in a selected region of the scene, and then the illuminated region is scanned over the entire scene. The averaging strategy defined in the previous section is called *spread-and-average*, as it includes spreading all the light over the entire projector image plane, and then averaging frames.

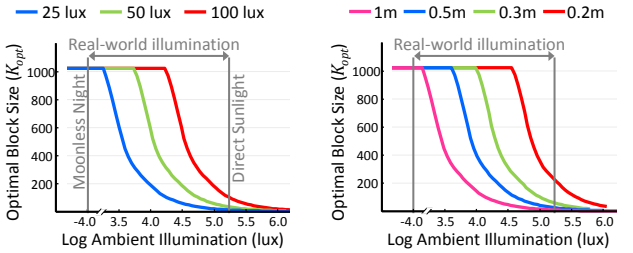
While the concentrate-and-scan approach requires s times more images (as each of the s block is encoded independently), since each column receives s times more light, SNR is increased by a factor of s, without requiring any frame-averaging. Thus, the decodability condition is satisfied with only s times more measurements. In contrast, as discussed in the previous section, the spread-and-average approach would require s^2 times more measurements to increase the SNR by a factor of s. Thus, we get:

Result 2 (Light-concentration advantage) *Given a fixed power budget, in order to achieve a desired accuracy level (SNR), it is* always *better to increase the signal directly by using the concentrate-and-scan approach, instead of reducing noise by the spread-and-average approach.*

The above result, called the *light-concentration advantage (LCA)*, forms the basis of the proposed techniques. As we show later, as a consequence of the LCA, concentrate-and-scan requires a much lower acquisition time (1-2 orders of magnitude smaller), as compared to spread-and-average in extreme ambient illumination conditions. In the following, we formalize the concepts introduced above.

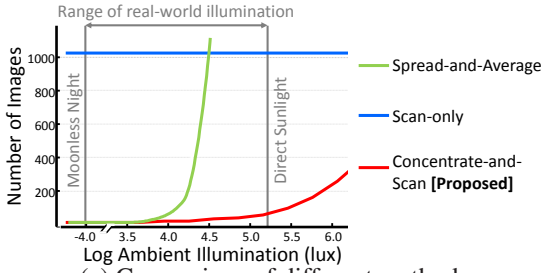
4.1. Concentrate-and-scan structured light

Suppose we could concentrate all the light into any block of size K columns, where K $(1 \leq K \leq C)$ could be chosen arbitrarily. Then, given a fixed block size K, concentrate-and-scan structured light consists of dividing the projector image plane into $\lceil \frac{C}{K} \rceil$ non-overlapping blocks of K columns each. Let the blocks be $B_1, B_2, \ldots, B_{\lceil \frac{C}{K} \rceil}$. Then, for each block B_i, only the columns within B_i are encoded (using any existing coding scheme) while spreading light only within that block. This step is repeated sequentially for all the blocks by concentrating light in a single block at a time. This is illustrated in Figure 2 (center).

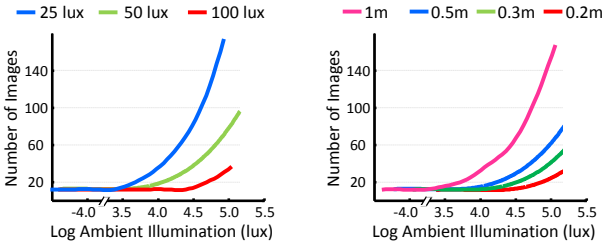


(a) Different source powers (b) Different scene distances

Figure 3. **Optimal block size K_{opt} for the proposed concentrate-and-scan method.** (a) Variation of K_{opt} with ambient illuminance level, for different light source powers P (resulting in illuminance of 25, 50 and 100 lux, respectively at a normally facing scene point 1 meter away). (b) Variation of K_{opt} for different scene-source distances, for the 50 lux light source.



(a) Comparison of different methods



(b) Different source powers (c) Different scene distances

Figure 4. **Number of measurements.** (a) Comparison of the number of measurements required by different methods. Scene-source distance is assumed to be 1 meter, and source illuminance is 50 lux. For most scenarios, the concentrate-and-scan method requires 1-2 orders of magnitude fewer images than existing methods. Number of images required by our method, for (b) different source power ratings, and (c) for different scene–source distances.

4.2. What is the optimal block size K?

The block size K determines the total number of measurements, and also the SNR of each measurement. A large block size K requires fewer measurements, but also results in low SNR per measurement. On the other hand, small K requires more measurements, but higher SNR per measurement. Given this trade-off, what K should be used?

In order to fully exploit the light-concentration advantage, the block size K should be chosen so that the decodability condition is satisfied without requiring frame-averaging. Let R_l be the source illuminance when light is spread over the entire image plane. Then, the illuminance when light is concentrated into K columns is $R_l \frac{C}{K}$. Substituting in the decodability condition (Eq. 5), we get:

$$K \leq \frac{\lambda C}{\tau} \frac{R_l}{\sqrt{R_a}}. \tag{8}$$

On the other hand, K should be as large as possible (up

to a maximum of C) in order to minimize the number of required images. Thus, the optimal block size K_{opt} is:

$$K_{opt} = \frac{\lambda C}{\tau} \frac{R_l}{\sqrt{R_a}}. \quad (9)$$

As expected, K_{opt} is inversely correlated with R_a. As R_a increases, K_{opt} becomes smaller. This ensures that the available light is concentrated into a smaller region so that the decodability condition is satisfied.

Figure 3 (a) shows the variation of K_{opt} with R_a, for different source powers. The three sources correspond approximately to a small laser pointer, a desktop laser scanner and a pico projector (resulting in illuminance of 25 lux, 50 lux and 100 lux respectively at a normally facing scene point 1 meter away). The number of projector columns is $C = 1024$. For these settings, $\lambda = 4.47$ (see the supplementary technical report [2] for details of computation of λ). The constant $\tau = 3.0$ was calculated assuming binary structured light coding [5], and the accuracy level is 0.5 pixels - accuracy is defined in terms of the difference between the estimated projector column correspondence and the ground-truth. As the source power P increases, the curve shifts to the right. Similarly, increasing the source-scene distance effectively reduces the source power, and shifts the plot to the left, as shown in Figure 3 (b).

4.3. Acquisition time

Let N_K be the number of images required to encode a block of size K columns. N_K depends on the coding scheme used within each block. The number of measurements M_{cs} required for concentrate-and-scan is simply the product of N_K and the number of blocks $\frac{C}{K}$: $M_{cs} = N_K \times \frac{C}{K}$. Note that no frame-averaging is required as the SNR of each measurement is sufficiently high to satisfy the decodability condition. Substituting the value of K_{opt} from Eq. 9, we get:

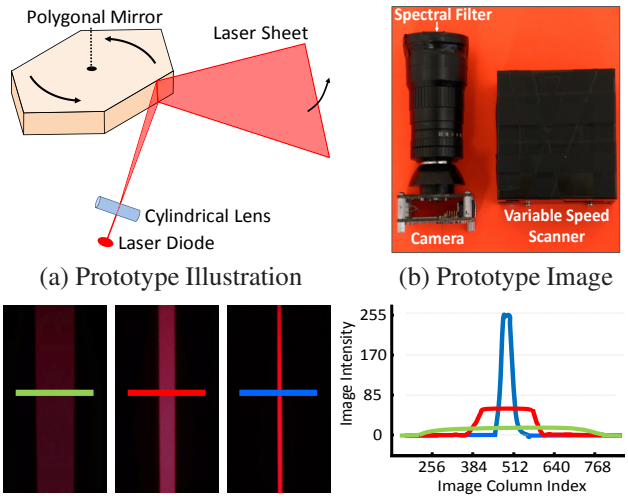
$$M_{cs} = N_K \frac{\tau}{\lambda R_l} \sqrt{R_a}. \quad (10)$$

Thus, we get the following result:

Result 3 (Acquisition time for concentrate-and-scan) *Given a fixed power budget P, the number of measurements M_{cs} (and hence the acquisition time) using the concentrate-and-scan approach is proportional to $\sqrt{R_a}$, i.e., $\mathcal{O}\left(\sqrt{R_a}\right)$.*

In contrast, recall from Result 1, that the number of measurements M_{sa} required for the spread-and-average approach is $\mathcal{O}(R_a)$. Thus, as R_a increases, M_{sa} increases much more rapidly as compared to M_{cs}. Figure 4 (a) shows the number of measurements required by the concentrate-and-scan and spread-and-average techniques for a wide range of ambient illumination levels. The camera, scene settings and the accuracy level are the same as in Figure 3 (a). It was assumed that binary Gray codes are used, so that $N_K = \log_2 K$. The source illuminance is 50 lux.

[5] Similar analysis can be performed for other structured light schemes such as phase-shifting [13] and N-ary coding [6]. See the supplementary technical report [2] for analysis and results for N-ary coding.



(a) Prototype Illustration (b) Prototype Image

(c-e) Different Light Spreads (f) Comparison

Figure 5. **Hardware Prototype.** (a-b) Our hardware system is based on an off-the-shelf laser scanner. The scanner has a rotating polygonal mirror that sweeps a laser sheet. Flexible light distribution capability is implemented by varying the mirror's rotation speeds. (c-e) A scene illuminated at different rotation speeds. As the speed decreases (from left to right), the illuminated area decreases, but the illumination strength increases. (f) Comparison of the intensities along the marked scanlines. Because the total energy is the same, the area under the three plots is the same.

We also plot the number of images required for single line-striping, where all the light is concentrated into a single column (as illustrated in Figure 2 (right)). This scan-only technique [8] is a special case of concentrate-and-scan approach with $K = 1$. The scan-only technique requires $M_s = C$ images, irrespective of the ambient illumination levels.

Implications (from Figure 4 (a)): *For typical low power projectors, the concentrate-and-scan approach requires 1-2 orders of magnitude (10-100 times) lower acquisition time than the existing schemes, for all outdoor ambient illuminance levels ($R_a > 10^4$).*

Conversely, given the same time budget, concentrate-and-scan approach achieves a significantly higher SNR and result quality. Figures 4 (b) and (c) show the variation of M_{cs}, for different source powers P and different scene-source distances d_{ss}. Again, the number of required images is relatively small even for the most extreme cases (direct sunlight, low-powered light source and large d_{ss}).

5. Hardware Prototype

In order to implement the concentrate-and-scan approach, we need a projector whose light could be distributed programmatically into any contiguous subset of K columns on the image plane. It should be possible to vary K. This functionality is not available in existing off-the-shelf projectors, which distribute light over the entire image plane. How can we achieve flexible light-distribution capability?

Scanning based projectors: While several existing projectors spread light spatially (e.g., using a lens), there is a class of projectors that raster-scan a narrow beam of light rapidly across the image plane. These are called scanning-projectors. For example, all laser-video projectors (e.g., Mi-

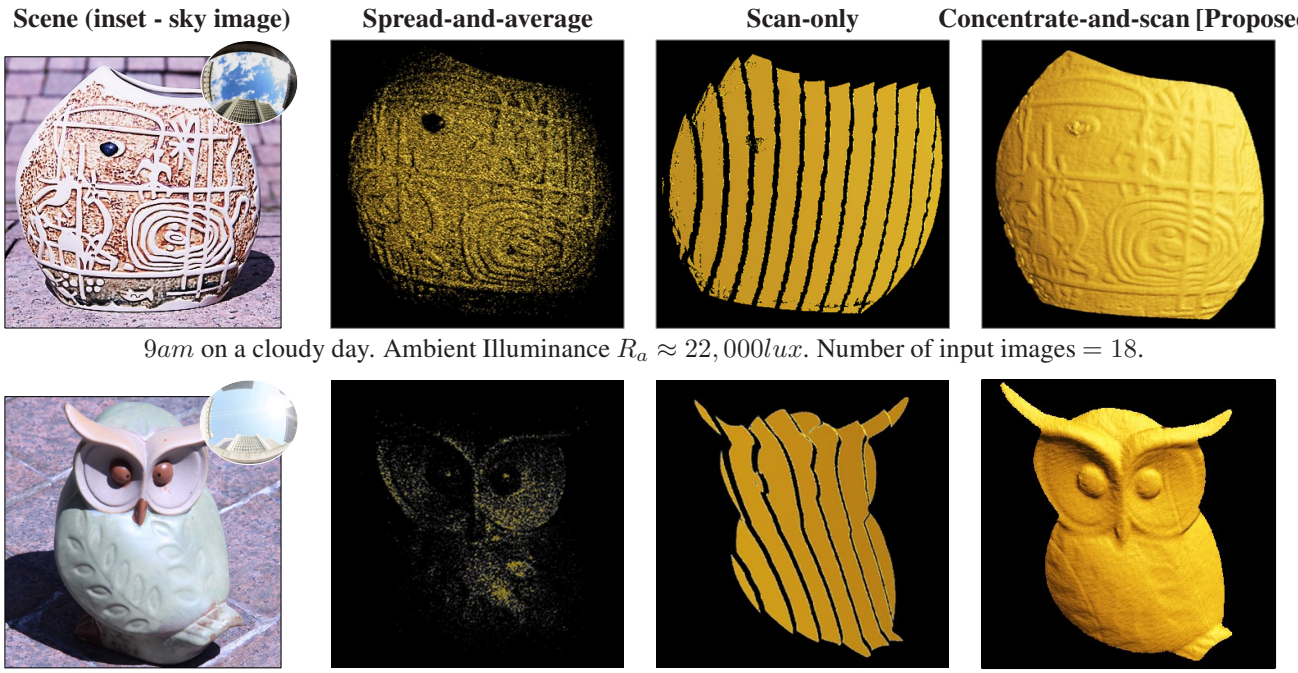


Figure 6. **Results of 3D scanning in sunlight.** (a) Objects placed outdoors in two different ambient illumination conditions - $9am$ on a cloudy day (top row) and $1pm$ on a bright sunny day (bottom row). 3D scanning results using (b) spread-and-average, (c) scan-only, and (d) the proposed concentrate-and-scan approaches, respectively. **For each row, the same capture time and power budget were used for all three techniques.** The spread-and-average method achieves a low SNR, resulting in large holes in the recovered shapes. The scan-only methods results in low resolution, thus losing all the surface details. Moreover, there are holes due to discontinuities at the boundaries. In contrast, the proposed method achieves high-quality results. The optimal block sizes for the concentrate-and-scan approach are $K_{opt} = 512$ and 256 columns for the top and bottom rows, respectively. The total number of projector columns $C = 1024$.

croVision SHOWWX+™ Laser Projector) and laser scanners belong to this category. The scanning mechanism is rapid enough that the beam traverses the entire image plane within the duration of one projected image. There are several realizations of the scanning mechanism, e.g. a galvanometer or a rapidly rotating polygonal mirror. Our hardware system is based on an off-the-shelf laser scanner from Spacevision Ltd. (www.space-vision.jp). The scanner uses a rotating polygonal mirror, and is illustrated in Figure 5.

The key observation is that it is possible to implement different light distributions by changing the speed of the scanning mechanism (in our case, the rotation velocity of the polygonal mirror) [6] Let the total power of the source be P. Suppose the scanning frequency is S scans-per-second (sps). The camera frame rate is also S frames-per-second (fps) so that one image is captured for every scan. If the total number of projector columns is C, the energy radiated by a single column during a single image capture is $P_c = \frac{P}{S \times C}$. If the scanning speed is reduced by a factor ω, only $\frac{C}{\omega}$ columns are illuminated in every captured image. The energy radiated by a column increases to $\omega \times P_c$. Figure 5 (c-e) shows a scene illuminated at three different rotation speeds. As the speed decreases, the illuminated area decreases, but the illumination strength increases.

Concentrate-and-scan structured light implementation: Let the optimal block size be K_{opt}; the image plane is divided into $\frac{C}{K_{opt}}$ blocks. Let the number of images required for encoding each block be N_K. Let the projected images be $\{T_i^j \mid 1 \leq i \leq N_K, 1 \leq j \leq \frac{C}{K_{opt}}\}$, where the subscript and the superscript are the image-index within a block, and the block index, respectively. Note that each T_i^j has K_{opt} columns. Concentrate-and-scan structured light consists of the following steps (for a pictorial explanation of the algorithm, see the project video available at [2]):

1. Reduce the scanner speed from S to $\frac{S \times K_{opt}}{C}$ sps. The frame rate of the camera remains the same at S fps.
2. For every i, concatenate all $\{T_i^j \mid 1 \leq j \leq \frac{C}{K_{opt}}\}$ images into a single image T_i^{cat} (having C columns). T_i^{cat} is projected during a single projector scan. In this duration, the camera captures $\frac{C}{K_{opt}}$ images I_i^j, one corresponding to each block.
3. For each camera pixel $\mathbf{x}$, identify the block j so that $I_i^j(\mathbf{x}) > 0$ for some i. This is the *corresponding block*, that contains the corresponding column (no column is encoded with an all zeros code). The corresponding column is estimated using the decoding algorithm for the coding scheme used within each block.

6. Results

Figure 6 shows 3D scanning results for objects placed outdoors under different ambient illuminance levels. The optimal block size was determined using Eq. 9. The constants $\lambda = 4.47$ and $\tau = 3.0$ were estimated using the expressions given in the technical report [2]. I_l and I_a were

[6] Different laser scanning speeds have been used for generating different camera exposures in a structured light setup [7].

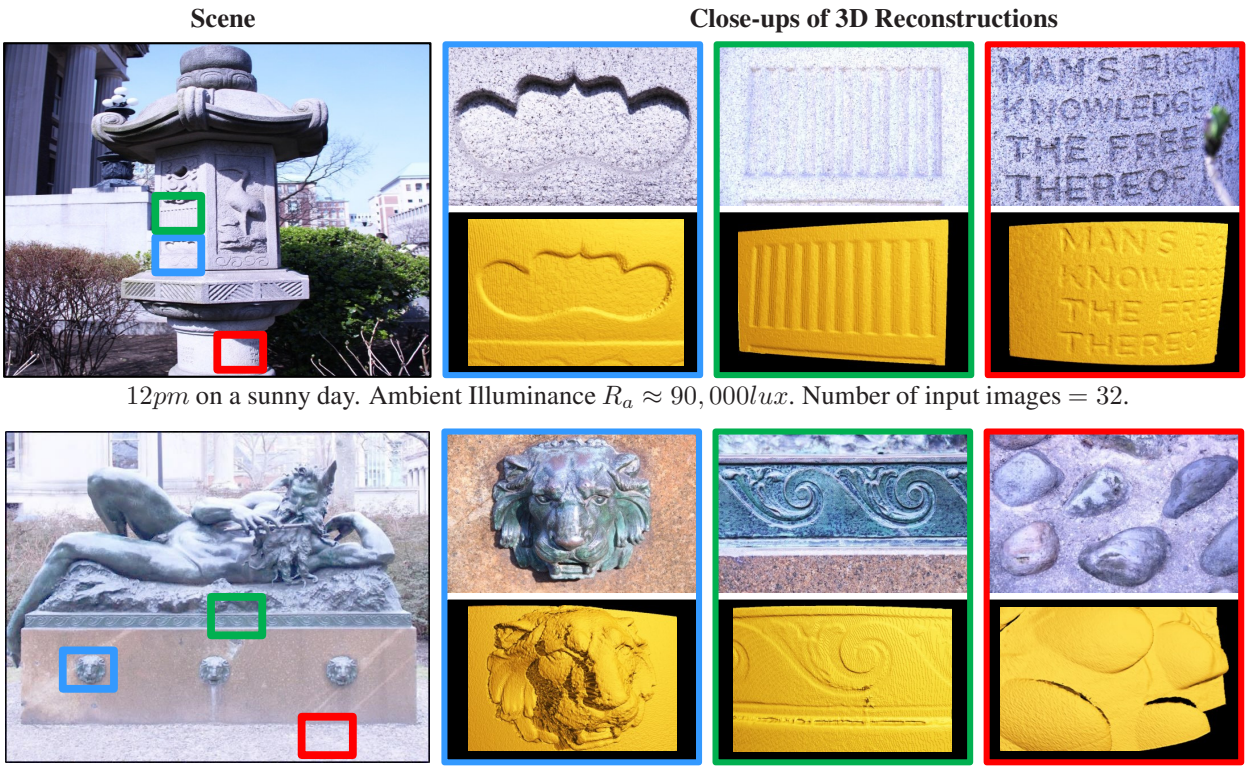


Figure 7. **Structured light in the wild.** 3D scanning results for two outdoor scenes with strong ambient light. In both cases, our method achieves highly detailed 3D structure with a limited power budget (illuminance from source ≈ 50 lux) and few (< 50) images.

measured by capturing two HDR images of the scene - one with the projector on, and one with the projector off [7]. Binary Gray code patterns were used as the structured light encoding scheme. Camera exposure times were chosen to be just below the saturation level.

We compare with the results of the spread-and-average and the scan-only methods. **The same capture time and power budget were used for all the methods.** Despite averaging, the spread-and-average method achieves low SNR, resulting in large holes in the recovered shapes. For the scan-only method, the width of the stripe was increased to $\frac{C}{M}$, where M is the number of measurements that can be captured within the time budget, so that the whole image plane is covered. Because the depth resolution is inversely correlated to the stripe-width, this method achieves a very low depth resolution. Notice that all the surface details are lost. Moreover, there are discontinuities at the stripe boundaries. In contrast, the proposed method achieves results with both high-resolution and high SNR.

Structured Light in the Wild: Figure 7 shows 3D scanning results for two outdoor scenes with strong ambient light (90,000 and 22,000 lux). In both cases, our method achieves highly detailed 3D structure with less than 50 images with a very limited power budget (source illuminance ≈ 50 lux).

Illumination-adaptive structured light: Since the optimal block size K_{opt} can be determined automatically using image-based measurements, we have implemented an illumination-adaptive structured light system. Figure 8 shows a scene scanned at various times of the day. As the day progresses, ambient illuminance increases, and the number of measurements increases accordingly (10, 18, 18, 32 and 56). For each illumination level, we show comparison with the spread-and-average method. **For each instant, the capture time and power budget are the same for both the methods.** For low ambient illumination, $K_{opt} = C$. In this case, our method behaves like the spread-only method. As ambient illuminance increases, the result quality of the spread-and-average scheme deteriorates. **For a time-lapse video of results, see the project video [2].**

7. Discussion

Contributions: This paper proposes light distribution as a new dimension in the design of structured light systems. We show that by controlling the distribution of the light, it is possible to develop fast and accurate 3D scanning systems that work in a wide range of outdoor scenarios with a limited time and power budget.

Limitations: Our approach assumes that the power of the light source, when completely concentrated into a single line, is sufficient for the decodability condition to be satisfied. While this is true in most settings even for a low-power light source, for parts of a highly specular objects, the image component due to ambient illumination may be too strong. In this case, even concentrating all the light into a single column fails to overcome ambient illumination. An example is illustrated in Figure 9. It is possible to overcome this limitation partly by diffusing the projected patterns [9].

[7] It is assumed that R_a is constant across the scene. If there is large variation in R_a (e.g., due to a shadow edge), different block sizes can be used for different parts of the scene.

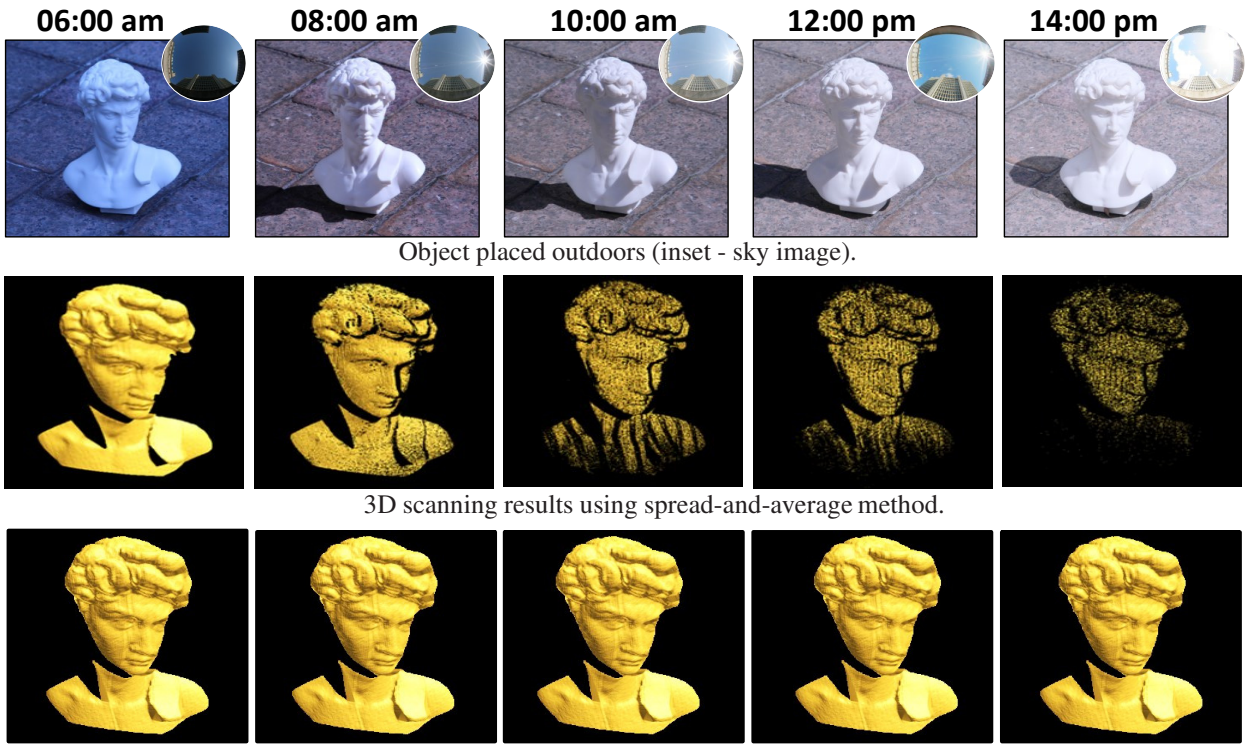


Figure 8. **Illumination-adaptive structured light.** 3D scanning results at different times of the day. **For each instant (each column), the capture time and power budget are the same for both methods.** For low ambient illuminance (left), both concentrate-and-scan and spread-and-average methods produce good results. As the day progresses, concentrate-and-scan method adapts to the ambient illuminance level (increasing from left to right) by choosing the appropriate block size, and achieves results of much higher quality.

Acknowledgments: This research was supported by NSF (grant IIS 09-64429) and ONR (grant N00014-11-1-0285). The authors are grateful to Yukio Sato of Space-Vision Inc. for making the laser scanner and associated software available for the experiments reported in this paper.

References

[1] Kinect outdoors. *www.youtube.com/watch?v=rI6CU9aRDIo*. 1

[2] Project webpage. *http://www.cs.columbia.edu/CAVE/projects/StructuredLightInSunlight/*. 3, 5, 6, 7

[3] G. J. Agin and T. O. Binford. Computer description of curved objects. *IEEE Transactions on Computers*, 25(4), 1976. 2

[4] D. Caspi, N. Kiryati, and J. Shamir. Range imaging with adaptive color structured light. *IEEE PAMI*, 20(5), 1998. 2

[5] S. W. Hasinoff, F. Durand, and W. T. Freeman. Noise-optimal capture for high dynamic range photography. In *CVPR*, 2010. 3

[6] E. Horn and N. Kiryati. Toward optimal structured light patterns. *Proc. 3DIM*, 1997. 2, 5

[7] I. Ihrke, I. Reshetouski, A. Manakov, A. Tevs, M. Wand, and H.-P. Seidel. A Kaleidoscopic Approach to Surround Geometry and Reflectance Acquisition. *Proc. IEEE CCD Workshop*, 2012. 6

[8] C. Mertz, S. Koppal, S. Sia, and S. Narasimhan. A low-power structured light sensor for outdoor scene reconstruction and dominant material identification. *Proc. IEEE PROCAMS*, 2012. 3, 5

[9] S. K. Nayar and M. Gupta. Diffuse structured light. In *ICCP*, 2012. 7

[10] D. Padilla, P. Davidson, J. Carlson, and D. Novick. Advancements in sensing and perception using structured lighting techniques: An ldrd final report. *Sandia National Lab Report*, 2005. 1, 2

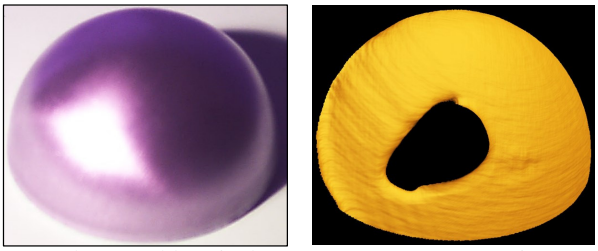
[11] J. Posdamer and M. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1), 1982. 2

[12] 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. 2

[13] V. Srinivasan, H. Liu, and M. Halious. Automated phase-measuring profilometry: A phase mapping approach. *Applied Optics*, 24, 1985. 2, 5

[14] Y. Wang, K. Liu, D. Lau, Q. Hao, and L. Hassebrook. Maximum snr pattern strategy for phase shifting methods in structured light illumination. *JOSA A*, 27(9), 2010. 2

[15] L. Zhang, B. Curless, and S. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. *Proc. IEEE 3DPVT*, 2002. 2



Specular metal hemisphere | 3D reconstruction

Figure 9. **Failure case.** (a) A metal hemisphere reflects sunlight specularly. (b) Inside the highlight, even concentrating all the projector light into a single column fails to overcome ambient illumination, resulting in a large hole in the reconstructed shape.

An SNR analysis of structured light techniques in the presence of strong ambient illumination

Supplementary Technical Report
for Paper ID 0100

In this technical report, we provide an analysis on the image formulation of structured light based techniques in the presence of strong ambient illumination. In Section 1, we derive an analytical expression for the signal-noise-ratio (SNR) of structured light coding schemes. We show that the SNR is proportional to the power ratio between structured light signal and ambient illumination. In Section 2, we formulate the theoretical relation between depth accuracy and SNR. Based on this relation, the concept of decodability condition is present for general coding schemes. In Section 3, we derive a theoretical lower bound on acquisition time for different structured light coding schemes.

1 SNR formulation of structured light coding scheme in strong ambient illumination

Structured light methods belong to the category of active triangulation techniques. Its general setup consists of a strucutred light source (projector) and a camera. We model the structured light source L as a projector that has an image plane with C columns. The projector projects spatio-temporally coded patterns on the scene so that a unique intensity code is assigned to each column [1].

The intensity of a scene point S in a captured image is:

$$I = I_l + I_a + \eta, \tag{1}$$

where I_l and I_a are intensities corresponding to the light source L and ambient illumination A, respectively. η is the camera noise. The goal is to extract the signal component I_l reliably from the captured images.

The accuracy of the estimated signal $\hat{I}_l$ is proportional to the signal-to-noise-ratio:

[1] Because of epipolar geometry between the projector and camera, only 1D coding (e.g., along the columns) on the projector plane is sufficient to perform depth recovery using triangulation.

$$SNR = \frac{\Delta(I_l)}{\eta} = \frac{1}{L-1} \cdot \frac{I_l}{\eta}, \tag{2}$$

where $\Delta(I_l)$ is the minimal distance required to distinguish two different signals. L is number of different itensities used for coding. For example, since binary methods [3] only use two different intenties, $L = 2$ in this case. For N-ary coding [1] or color coding [4], L equals to N or the color number.

The components I_l and I_a are proportional to the irradiance values R_l and R_a at scene point S due to the light source L and ambient illumination A, respectively:

$$I_l = \alpha R_l, \quad I_a = \beta R_a, \tag{3}$$

where α and β encapsulate the scene point's BRDF, light fall-off, and camera gain. β also includes the effect of any optical (e.g., spectral or polarization) filtering used for reducing ambient illumination.

We assume the affine camera noise model, with both signal-dependent and signal-independent terms [2]:

$$\eta^2 = \sigma_r^2 + \frac{\alpha R_l + \beta R_a}{g}, \tag{4}$$

where σ_r is the standard deviation of the signal-independent sensor read noise, and g is camera gain. In scenarios with strong ambient illumination, $R_a >> R_l$, and the dominant source of noise is the signal-dependent photon noise, i.e., $\sigma_r^2 << \frac{\beta R_a}{g}$. Then, the SNR is approximated as:

$$SNR \approx \lambda \cdot \frac{1}{L-1} \cdot \frac{R_l}{\sqrt{R_a}}, \tag{5}$$

where $\lambda = \sqrt{\frac{\alpha^2 g}{\beta}}$. The reflection properties of the scene are assumed to be the same for structured light and ambient illumination. Also, we used a narrow-band laser light source and spectral filter in front of our camera. This suppresses ambient illumination by a factor of 20. In specific, $\alpha = 0.250$, $\beta = 0.0125$, $g = 4.0$ for our setup and λ is caculated as 4.47 accordingly.

2 Theoretical relation between SNR and depth accuracy

In this section, we derive the expression of depth accuracy given a SNR level. We have known that this relation depends on the structured light coding and decoding algorithms. For analysis simplicity, we assume that minimum-distance decoding is used for all coding schemes. Also, we assume the smallest hamming

distance of a coding scheme is always 1, which means the signal can be decoded correctly only when all the code digits are completely correct.

Suppose I_l^i is an intensity code, the captured intensity is $\hat{I_l^i} = I_l^i + \eta$ due to sensor noise. The condition that I_l^i can be decoded correctly with minimum-distance decoder is:

$$|\hat{I_l^i} - I_l^i| < |\hat{I_l^i} - I_l^j|, \text{ for all } j \neq i \tag{6}$$

For Eq. 6 to hold for even the nearest two codes, we get the condition:

$$\min(\frac{I_l^i - I_l^j}{2}) > \frac{\tau}{2} \cdot \eta, \text{ for all } j \neq i \tag{7}$$

where τ is a parameter to control the probability of decoding correctness. Since photon noise is the dominating noise source in strong ambient illumination, we use a gaussian distribution $N(\cdot)$ to approximate the noise distribution of each intensity code. Therefore, we can give the probability $p_r(i)$ that a single intensity code is decoded correctly:

$$p_r(i) = N(\min(\frac{I_l^i - I_l^j}{2}) > \frac{\tau}{2} \cdot \eta) = N(\frac{1}{L-1} \cdot I_l > \tau \cdot \eta) = N(\text{SNR} > \tau) \tag{8}$$

where L is the number of intensity levels as discussed in Section 1.

We assume the code length of a specific coding scheme is N_c and all code bits are independent to each other. The probability that the code sequence is correctly decoded is:

$$P_r = \prod p_r(i),\ 1 < i < N_c \tag{9}$$

The error tolerance of correspondence decoding is used as the measurement of depth accuracy.

$$\delta = \mu \cdot (1 - \prod p_r(i)),\ 1 < i < N_c \tag{10}$$

where μ is a weight constant set to be 10 in our setting.

In order to achieve a desired depth accuracy δ, the SNR should be higher than a threshold τ, i.e., $SNR > \tau$. The threshold τ increases monotonically with δ. Substituting in Eq. 5:

$$\frac{1}{L-1} \cdot \frac{R_l}{\sqrt{R_a}} \geq \frac{\tau}{\lambda}. \tag{11}$$

We call this the *decodability condition*. In order to achieve the desired depth accuracy, all the captured images must satisfy the decodability condition. In our experiment, we set the decoding error tolerance to be 0.5 column in average. As a result, the τ equals to 3.0 in our experiment.

3 Theoretical lower bound on acquisition time of structured light coding schemes

In this section, we discuss the theoretical lower bound on the aquisition time of existing structured light coding schemes. As introduced in Section 2, to achieve a required accuracy level, structured light coding schemes have to meet the decodability condition.

When the SNR is insufficient to hold the decodability condition, a common technique for increasing SNR is by capturing multiple frames per image. It will increase the SNR by a factor of $\sqrt{f}$ when f frames are averaged per image. To meet the decodability condition using frame-averaging, we get the expression for the minimal number of frames per image:

$$f \geq \left(\frac{\tau \cdot (L-1)}{\lambda R_l} \right)^2 R_a \ . \tag{12}$$

Let N_c be the number of images required by the particular structured light coding scheme used to encode all the projector columns uniquely, and f, as defined above, is the number of frames to be averaged per image. Then, the total number of measurements M is given as:

$$M = N_c \times f \, . \tag{13}$$

If we set a constant exposure time to all the captured frames, we will get a lower bound on the total number of measurements, which is also a lower bound on the aquisition time:

$$M \geq N_c \cdot \max(\left(\frac{\tau \cdot (L-1)}{\lambda R_l} \right)^2 R_a, 1) \ . \tag{14}$$

Eq. 14 is applicable to general existing structured light schemes. For example, we set the $L = 2$, then we get the lower bound for binary coding. We set the L to be the number of intensity levels in N-ary or color coding, then we also get the lower bound for them.

In our paper, we introduce the concept of light redistribution. Then we bring in a new dimensoin in the Eq. 14. If the total column number is C, and K columns are illuminated at a time, Eq. 14 can be re-written as:

$$M \geq \frac{C}{K} \cdot N_c^{'} \cdot \max(\left(\frac{K \cdot \tau \cdot (L-1)}{C \cdot \lambda R_l} \right)^2 R_a, 1) \ . \tag{15}$$

For more discussion on the light redistribution, please check the paper for more details.

References

[1] D. Caspi, N. Kiryati, and J. Shamir. Range imaging with adaptive color structured light. *IEEE PAMI*, 20(5), 1998.

[2] Samuel W. Hasinoff, Fredo Durand, and William T. Freeman. Noise-optimal capture for high dynamic range photography. In *CVPR*, 2010.

[3] J. Posdamer and M. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1), 1982.

[4] L. Zhang, B. Curless, and S.M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. *Proc. IEEE 3DPVT*, 2002.