Structured Light 3D Scanning in the Presence of Global Illumination

Gupta, M.; Agrawal, A.; Veeraraghavan, A.; Narasimhan, S.

Abstract

Global illumination effects such as inter-reflections, diffusion and sub-surface scattering severely degrade the performance of structured light-based 3D scanning. In this paper, we analyze the errors caused by global illumination in structured light-based shape recovery. Based on this analysis, we design structured light patterns that are resilient to individual global illumination effects using simple logical operations and tools from combinatorial mathematics. Scenes exhibiting multiple phenomena are handled by combining results from a small ensemble of such patterns. This combination also allows us to detect any residual errors that are corrected by acquiring a few additional images. Our techniques do not require explicit separation of the direct and global components of scene radiance and hence work even in scenarios where the separation fails or the direct component is too low. Our methods can be readily incorporated into existing scanning systems without significant overhead in terms of capture time or hardware. We show results on a variety of scenes with complex shape and material properties and challenging global illumination effects.

IEEE Computer Vision & Pattern Recognition (CVPR)

Structured Light 3D Scanning in the Presence of Global Illumination

Mohit Gupta[†], Amit Agrawal[‡], Ashok Veeraraghavan[‡] and Srinivasa G. Narasimhan[†]

[†] Robotics Institute, Carnegie Mellon University, Pittsburgh, USA

[‡] Mitsubishi Electric Research Labs, Cambridge, USA

Abstract

Global illumination effects such as inter-reflections, diffusion and sub-surface scattering severely degrade the performance of structured light-based 3D scanning. In this paper, we analyze the errors caused by global illumination in structured light-based shape recovery. Based on this analysis, we design structured light patterns that are resilient to individual global illumination effects using simple logical operations and tools from combinatorial mathematics. Scenes exhibiting multiple phenomena are handled by combining results from a small ensemble of such patterns. This combination also allows us to detect any residual errors that are corrected by acquiring a few additional images.

Our techniques do not require explicit separation of the direct and global components of scene radiance and hence work even in scenarios where the separation fails or the direct component is too low. Our methods can be readily incorporated into existing scanning systems without significant overhead in terms of capture time or hardware. We show results on a variety of scenes with complex shape and material properties and challenging global illumination effects.

1. Introduction

Structured light triangulation has become the method of choice for shape measurement in several applications including industrial automation, graphics, human-computer interaction and surgery. Since the early work in the field about 40 years ago [18, 12], research has been driven by two factors: reducing the acquisition time and increasing the depth resolution. Significant progress has been made on both fronts (see the survey by Salvi et al [16]) as demonstrated by systems which can recover shapes at close to 1000 Hz. [21] and at a depth resolution better than 30 microns [5].

Despite these advances, most structured light techniques make an important assumption: scene points receive illumination only directly from the light source. For many real world scenarios, this is not true. Imagine a robot trying to navigate an underground cave or an indoor scenario, a surgical instrument inside human body, a robotic arm sorting a heap of metallic machine parts, or a movie director wanting to image the face of an actor. In all these settings, scene points receive illumination indirectly in the form of inter-reflections, sub-surface or volumetric scattering. Such effects, collectively termed *global or indirect illumination*[1], often dominate the direct illumination and strongly depend on the shape and material properties of the scene. Not accounting for these effects results in large and systematic errors in the recovered shape (see Figure 1b).

The goal of this paper is to build an end-to-end system for structured light scanning under a broad range of global illumination effects. We begin by formally analyzing errors caused due to different global illumination effects. We show that the types and magnitude of errors depend on the region of influence of global illumination at any scene point. For instance, some scene points may receive global illumination only from a local neighborhood (sub-surface scattering). We call these short-range effects. Some points may receive global illumination from a larger region (inter-reflections or diffusion). We call these long range effects.

The key idea is to design patterns that modulate global illumination and prevent the errors at capture time itself. Short and long range effects place contrasting demands on the patterns. Whereas low spatial frequency patterns are best suited for short range effects, long range effects require the patterns to have high-frequencies. Since most currently used patterns (e.g., binary and sinusoidal codes) contain a combination of both low and high spatial frequencies, they are ill-equipped to prevent errors. We show that such patterns can be converted to those with only high frequencies by applying simple logical operations, making them resilient to long range effects. Similarly, we use tools from combinatorial mathematics to design patterns consisting solely of frequencies that are low enough to make them resilient to short range effects.

But how do we handle scenes that exhibit more than one type of global illumination effect (such as the one in Figure 1a)? To answer this, we observe that it is highly unlikely for two different patterns to produce the same erroneous decoding. This observation allows us to project a small ensemble of patterns and use a simple voting scheme to compute the correct decoding at ev-

---

[1] Global illumination should not be confused with the oft-used "ambient illumination" that is subtracted by capturing image with the structured light source turned off.

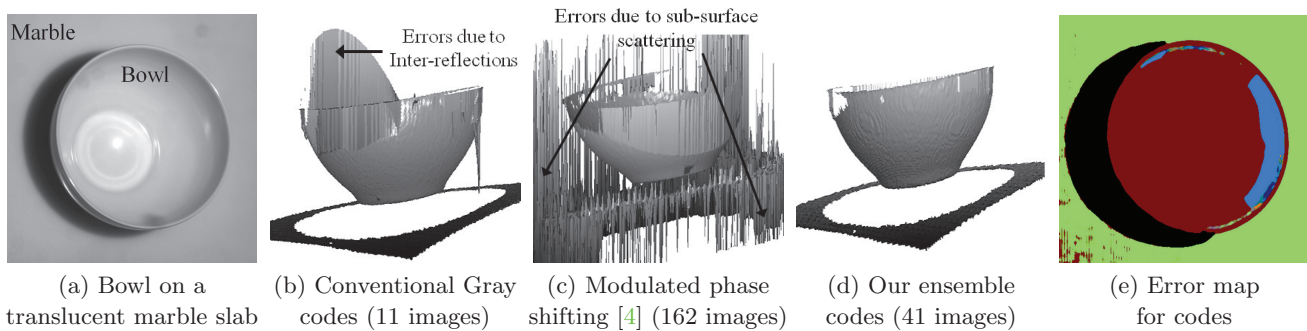| (a) Bowl on a translucent marble slab | (b) Conventional Gray codes (11 images) | (c) Modulated phase shifting [4] (162 images) | (d) Our ensemble codes (41 images) | (e) Error map for codes |

Figure 1. **Measuring shape for the 'bowl on marble-slab' scene.** This scene is challenging because of strong inter-reflections inside the concave bowl and sub-surface scattering on the translucent marble slab. (b-d) Shape reconstructions. Parentheses contain the number of input images. (b) Conventional Gray codes result in incorrect depths due to inter-reflections. (c) Modulated phase-shifting results in errors on the marble-slab because of low direct component. (d) Our technique uses an ensemble of codes optimized for individual light transport effects, and results in the best shape reconstruction. (e) By analyzing the errors made by the individual codes, we can infer qualitative information about light-transport. Points marked in green correspond to translucent materials. Points marked in light-blue receive heavy inter-reflections. Maroon points do not receive much global illumination. **For more results and detailed comparisons to existing techniques, please see the project web-page** [1].

ery pixel, without any prior knowledge about the types of effects in the scene (Figure 1d). For very challenging scenes, we present an error detection scheme based on a simple consistency check over the results of the individual codes in the ensemble. Finally, we present an error correction scheme by collecting a few additional images. We demonstrate accurate reconstructions on scenes with complex geometry and material properties, such as shiny brushed metal, translucent wax and marble and thick plastic diffusers (like shower curtains).

Our techniques do not require explicit separation of the direct and global components of scene radiance and hence work even in scenarios where the separation fails (e.g., strong inter-reflections among metallic objects) or where the direct component is too low and noisy (e.g., translucent objects or in the presence of defocus). Our techniques consistently outperform many traditional coding schemes and techniques which require explicit separation of the global component, such as modulated phase-shifting [4]. Our methods are simple to implement and can be readily incorporated into existing systems without significant overhead in terms of acquisition time or hardware.

2. Related Work

In this section, we summarize the works that address the problem of shape recovery under global illumination. The seminal work of Nayar *et al*. [13] presented an iterative approach for reconstructing shape of Lambertian objects in the presence of inter-reflections. Gupta *et al*. [8] presented methods for recovering depths using projector defocus [20] under global illumination effects. Chandraker *et al*. [2] use inter-reflections to resolve the bas-relief ambiguity inherent in shape-from-shading techniques. Holroyd et al [10] proposed an active multi-view stereo technique where high-frequency

illumination is used as scene texture that is invariant to global illumination. Park *et al*. [15] move the camera or the scene to mitigate the errors due to global illumination in a structured light setup. Hermans et al [9] use a moving projector in a variant of structured light triangulation. The depth measure used in this technique (frequency of the intensity profile at each pixel) is invariant to global light transport effects. In this paper, our focus is on designing structured light systems while avoiding the overhead due to moving components.

Recently, it was shown that the direct and global components of scene radiance could be efficiently separated [14] using high-frequency illumination patterns. This has led to several attempts to perform structured light scanning under global illumination [3, 4]. All these techniques rely on *subtracting or reducing* the global component and apply conventional approaches on the residual direct component. While these approaches have shown promise, there are three issues that prevent them from being applicable broadly: (a) the direct component estimation may fail due to strong inter-reflections (as with shiny metallic parts), (b) the residual direct component may be too low and noisy (as with translucent surfaces, milk and murky water), and (c) they require significantly higher number of images than traditional approaches, or rely on weak cues like polarization. In contrast, we explicitly design ensembles of illumination patterns that are resilient to a broader range of global illumination effects, using significantly less number of images.

3. Errors due to Global Illumination

The type and magnitude of errors due to global illumination depends on the spatial frequencies of the patterns and the global illumination effect. As shown in Figures 2 and 3, long range effects and short range

Conventional Coding and Decoding

(a)   (b)   (c)   (d)   (e)

Logical Coding and Decoding

(f)   (g)   (h)   (i)   (j)

Results and Comparison

(k) Depth (conventional Gray codes)   (l) Depth (our XOR04 codes)   (m) Comparison with the
Mean absolute error $= 28.8mm$ Mean absolute error $= 1.4mm$ ground-truth
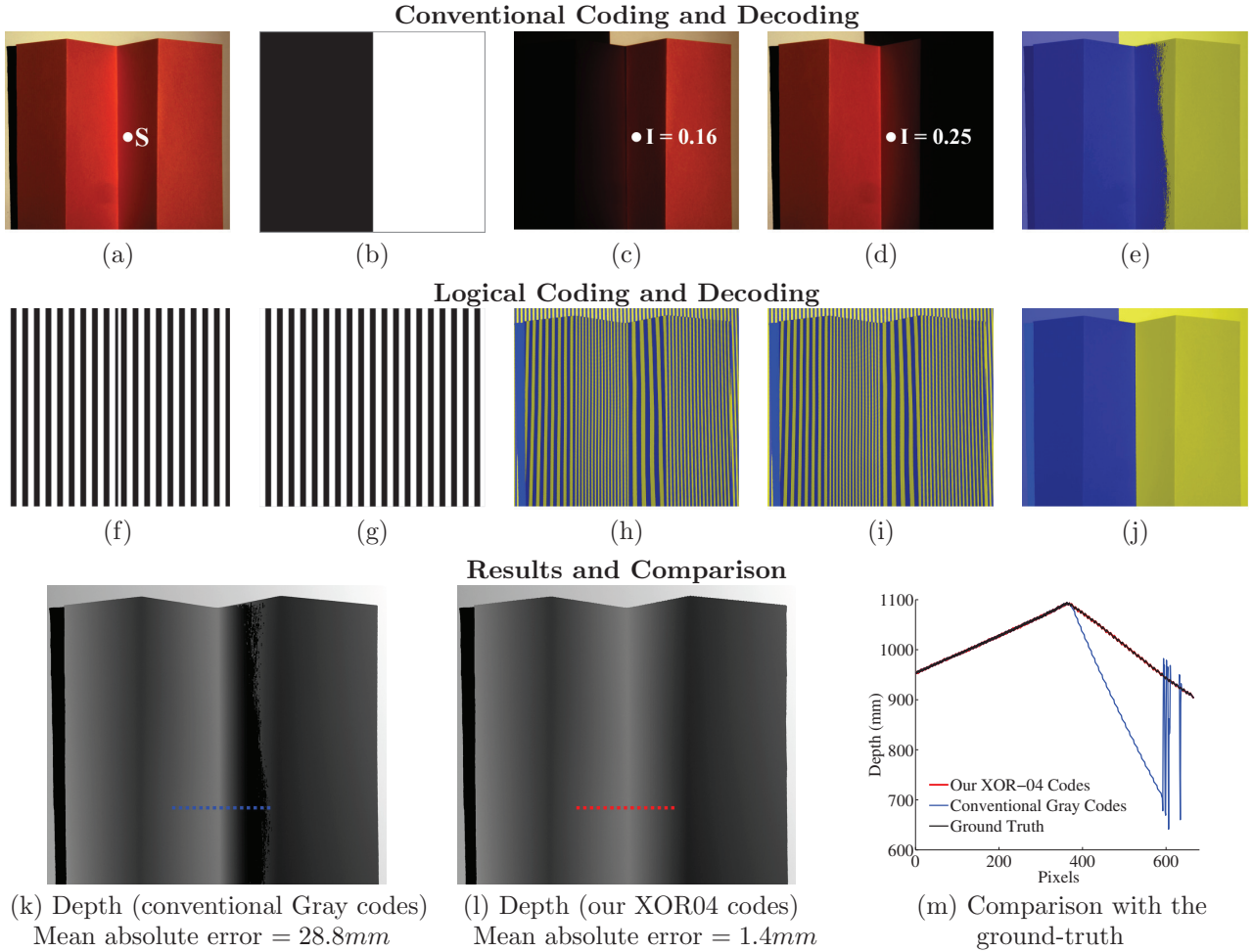
Figure 2. **Errors due to inter-reflections: First row:** Conventional coding and decoding. (a) A concave V-groove. The center edge is concave. (b) Low frequency pattern. (c-d) Images captured with pattern (b) and its inverse respectively. Point S is directly illuminated in (c). However, because of inter-reflections, its intensity is higher in (d), resulting in a decoding error. (e) Decoded bit plane. Points decoded as one (directly illuminated) and zero (not illuminated) are marked in yellow and blue respectively. In the correct decoding, only the points to the right of the concave edge should be one, and the rest zero. (k) Depth map computed with the conventional codes. Because of incorrect binarization of the low frequency patterns (higher-order bits), depth map has large errors. **Second row: Logical coding and decoding (Section 4.1).** (f-g) Pattern in (b) is decomposed into two high-frequency patterns. (h-i) Binarization of images captured with (f-g) respectively. (j) Binary decoding under (b) computed by taking pixel-wise XOR of (h) and (i). (l) Depth map computed using logical coding-decoding. The errors have been nearly completely removed. (m) Comparison with the ground-truth along the dotted lines in (k-l). Ground truth was computed by manually binarizing the captured images.

effects result in incorrect decoding of low and high spatial frequency patterns, respectively. In this section, we formally analyze these errors. For ease of exposition, we have focused on binary patterns. The analysis and techniques are easily extended to N-ary codes.

Binary patterns are decoded by binarizing the captured images into projector-illuminated vs. non-illuminated pixels. A robust way to do this is to capture two images L and $\overline{L}$, under the pattern P and the inverse pattern $\overline{P}$, respectively. For a scene point S^i, its irradiances L^i and $\overline{L^i}$ are compared. If, $L^i > \overline{L^i}$, then the point is classified as directly lit. A fundamental assumption for correct binarization is that each scene point receives irradiance from only a single illumination

element (light stripe or a projector pixel). However, due to global illumination effects and projector defocus, a scene point can receive irradiance from multiple projector pixels, resulting in incorrect binarization.

In the following, we derive the condition for correct binarization in the presence of global illumination and defocus. Suppose S^i is directly lit under a pattern P. The irradiances L^i and $\overline{L^i}$ are given as:

$$L^i = L_d^i + \beta L_g^i , \qquad (1)$$

$$\overline{L^i} = (1 - \beta) L_g^i , \qquad (2)$$

where L_d^i and L_g^i are the direct and global components of the irradiance at S^i when the scene is fully lit. β is the fraction of the global component under

715

the pattern P. In the presence of projector defocus, S^i receives fractions of the direct component, both under the pattern and its inverse [8].

$$L^i = \alpha L_d^i + \beta L_g^i \ , \tag{3}$$

$$\overline{L^i} = (1-\alpha) L_d^i + (1-\beta) L_g^i \ . \tag{4}$$

The fractions (α and $1-\alpha$) depend on the projected pattern and the amount of defocus. In the absence of defocus, $\alpha = 1$. For correct binarization, it is required that $L_i > \overline{L_i}$, i.e.

$$\alpha L_d^i + \beta L_g^i > (1-\alpha) L_d^i + (1-\beta) L_g^i \tag{5}$$

This condition is satisfied in the absence of global illumination ($L_g^i = 0$) and defocus ($\alpha = 1$). In the following, we analyze the errors in the binarization process due to various global illumination effects and defocus, leading to systematic errors[2].

Long range effects (diffuse and specular inter-reflections): Consider the scenario when S^i receives a major fraction of the global component when it is not directly lit ($\beta \approx 0$), and the global component is larger than the direct component ($L_d^i < L_g^i$) as well. Substituting in the binarization condition (Eqn. 5), we get $L^i < \overline{L^i}$, which results in a binarization error. Such a situation can commonly arise due to long-range inter-reflections, when scenes are illuminated with low-frequency patterns. For example, consider the v-groove concavity as shown in Figure 2. Under a low frequency pattern, several scene points in the concavity are brighter when they are not directly lit, resulting in a binarization error. Since the low frequency patterns correspond to the higher-order bits, this results in a large error in the recovered shape.

Short-range effects (sub-surface scattering and defocus): Short range effects result in low-pass filtering of the incident illumination. In the context of structured light, these effects can severely blur the high-frequency patterns, making it hard to correctly binarize them. This can be explained in terms of the binarization condition in Eqn 5. For high frequency patterns, $\beta \approx 0.5$ [14]. If the difference in the direct terms $|\alpha L_d^i - (1-\alpha) L_d^i|$ is small, either because the direct component is low due to sub-surface scattering ($L_d^i \approx 0$) or because of severe defocus ($\alpha \approx 0.5$), the pattern can not be robustly binarized due to low signal-to-noise-ratio (SNR). An example is shown in Figure 3. For conventional Gray codes, this results in a loss of depth resolution, as illustrated in Figure 4.

4. Patterns for Error Prevention

Errors due to global illumination are systematic, scene-dependent errors that are hard to eliminate in

---

[2]Errors for the particular case of laser range scanning of translucent materials are analyzed in [7]. Errors due to sensor noise and spatial mis-alignment of projector-camera pixels were analyzed in [17].
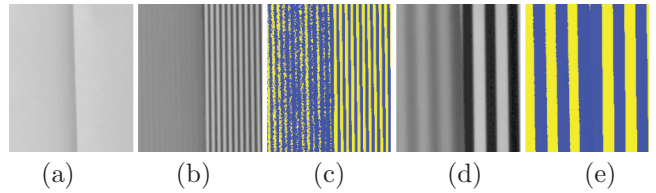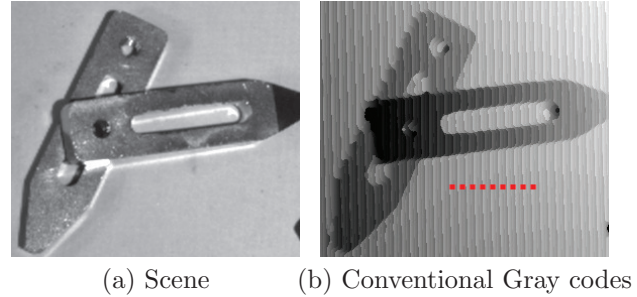

(a)   (b)   (c)   (d)   (e)

Figure 3. **Errors due to sub-surface scattering:** (a) This scene consists of a translucent slab of marble on the left and an opaque plane on the right. (b) A high frequency pattern is severely blurred on the marble, and can not be binarized correctly (c). Image captured (d) under a low-frequency pattern can be binarized correctly (e).

(a) Scene   (b) Conventional Gray codes

(c) Large min-SW Gray codes   (d) Comparison

Figure 4. **Depth computation under defocus:** (a) A scene consisting of industrial parts. (b) Due to defocus, the high frequency patterns in the conventional Gray codes can not be decoded, resulting in a loss of depth resolution. Notice the quantization artifacts. (c) Depth map computed using Gray codes with large minimum stripe-width (min-SW) does not suffer from loss of depth resolution.

post-processing. In this section, we design patterns that modulate global illumination and prevent errors from happening at capture time itself. In the presence of only long range effects and no short-range effects, high-frequency binary patterns (with equal off and on pixels) are decoded correctly because $\beta \approx 0.5$ [14], as shown in Figures 2(f-i). On the other hand, in the presence of short-range effects, most of the global illumination comes from a local neighborhood. Thus, for low frequency patterns, when a scene point is directly illuminated, most of its local neighborhood is directly illuminated as well. Hence, $\alpha \geq 0.5$ and $\beta \geq 0.5$. Thus, if we use low frequency patterns for short-range effects, the global component actually helps in correct decoding even when the direct component is low (Figure 3).

Because of the contrasting requirements on spatial

frequencies, it is clear that we need different codes for different effects. For long range effects, we want patterns with only high frequencies (low maximum stripe-widths). For short-range effects, we want patterns with only low frequencies (high minimum stripe-widths). But most currently used patterns contain a combination of both low and high spatial frequencies. How do we design patterns with only low or only high frequencies? We show that by performing simple logical operations, it is possible to design codes with only high frequency patterns. For short range effects, we draw on tools from the combinatorial maths literature to design patterns with large minimum stripe-widths.

4.1. Logical coding-decoding for long range effects

We introduce the concept of logical coding and decoding to design patterns with only high frequencies. An example of logical coding-decoding is given in Figure 2. The important observation is that for structured light decoding, *the direct component is just an intermediate representation*, with the eventual goal being the correct binarization of the captured image. Thus, we can bypass explicitly computing the direct component. Instead, we can model the binarization process as a scene-dependent function from the set of binary projected patterns ($\mathbb{P}$) to the set of binary classifications of the captured image ($\mathbb{B}$):

$$f : \mathbb{P} \Rightarrow \mathbb{B} . \quad (6)$$

For a given pattern $P \in \mathbb{P}$, this function returns a binarization of the captured image if the scene is illuminated by P. Under inter-reflections, this function can be computed robustly for high-frequency patterns but not for low-frequency patterns. For a low frequency pattern P_{lf}, we would like to decompose it into two high-frequency patterns P_{hf}^1 and P_{hf}^2 using a pixel-wise binary operator $\odot$ such that:

$$f(P_{lf}) = f\left(P_{hf}^1 \odot P_{hf}^2\right) = f\left(P_{hf}^1\right) \odot f\left(P_{hf}^2\right) \quad (7)$$

If we find such a decomposition, we can robustly compute the binarizations $f\left(P_{hf}^1\right)$ and $f\left(P_{hf}^2\right)$ under the two high frequency patterns, and compose these to achieve the correct binarization $f(P_{lf})$ under the low frequency pattern. Two questions remain: (a) What binary operator can be used? (b) How can we decompose a low frequency pattern into two high frequency patterns? For the binary operator, we choose the logical XOR ($\oplus$) because it has the following property:

$$P_{hf}^2 \oplus P_{hf}^1 = P_{lf} \Rightarrow P_{hf}^2 = P_{lf} \oplus P_{hf}^1 \quad (8)$$

This choice of operator provides a simple means to decompose P_{lf}. We first choose a high-frequency pattern P_{hf}^1. The second pattern P_{hf}^2 is then computed
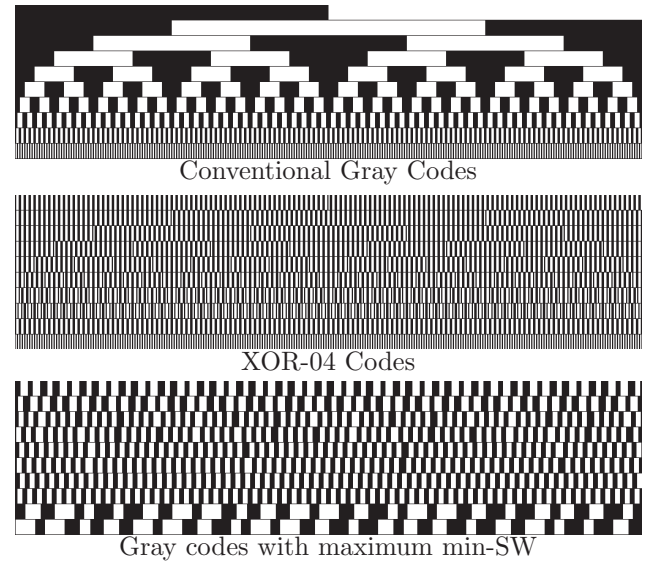

Figure 5. **Different codes:** The range of stripe-widths for conventional Gray codes is [2, 512]. For XOR-04 codes (optimized for long range effects) and Gray codes with maximized min-SW (optimized for short-range effects), the ranges are [2, 4] and [8, 32] respectively.

by simply taking the pixel-wise logical XOR of P_{lf} and P_{hf}^1. We call the first high frequency pattern the *base* pattern. Instead of projecting the original low frequency patterns, we project the base pattern P_{hf}^1 and the second high-frequency patterns P_{hf}^2. For example, if we use the last Gray code pattern (stripe width of 2) as the base pattern, all the projected patterns have a maximum width of 2. We call these the *XOR-02* codes. In contrast, the original Gray codes have a maximum stripe-width of 512. Note that there is no overhead introduced; the number of projected patterns remains the same as the conventional codes. Similarly, if we use the second-last pattern as the base-plane, we get the *XOR-04* codes (Figure 5). The pattern images can be downloaded from the project web-page [1].

4.2. Maximizing the minimum stripe-widths for short-range effects

Short-range effects can severely blur the high-frequency base plane of the logical XOR codes. The resulting binarization error will propagate to all the decoded patterns. Thus, for short-range effects, we need to design codes with large minimum stripe-width (min-SW). It is not feasible to find such codes with a brute-force search as these codes are extremely rare[3].

Fortunately, this problem has been well studied in combinatorial mathematics. There are constructions available to generate codes with large min-SW. For instance, the 10-bit Gray code with the maximum known

---

[3]On the contrary, it is easy to generate codes with small maximum stripe-width (9), as compared to 512 for the conventional Gray codes, by performing a brute-force search

(a) Fruit-basket   (b) Conventional Gray (11 images)

(c) Phase-shifting
(18 images)

(d) Modulated phase
shifting [4] (162 images)

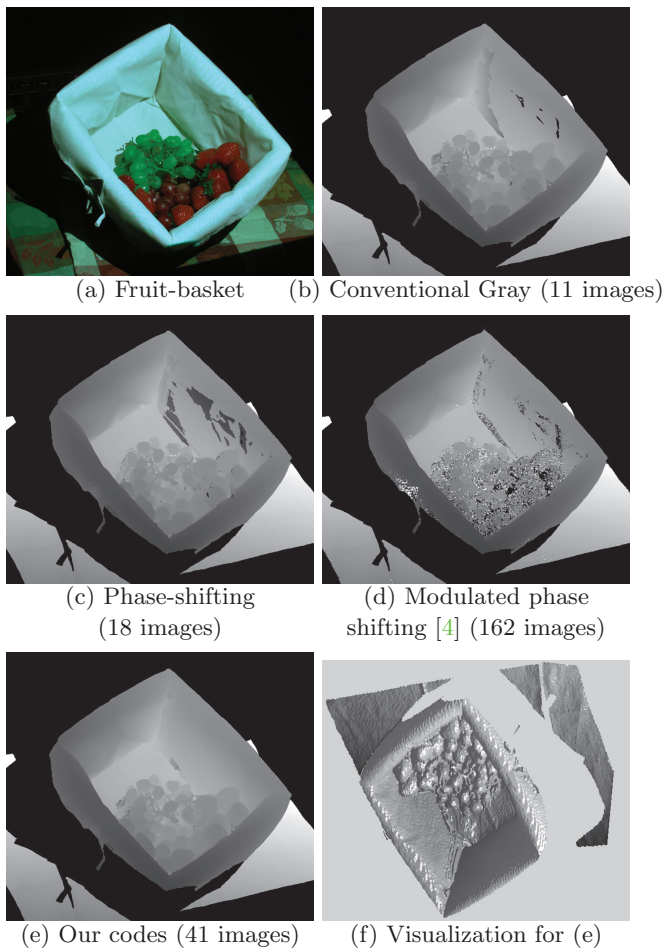(e) Our codes (41 images)   (f) Visualization for (e)

Figure 6. **Depth map computation for the fruit-basket scene.** Parentheses contain the number of input images. Conventional Gray codes (b) and phase-shifting (c) result in errors at points receiving inter-reflections. Modulated phase-shifting (d) produces errors on translucent fruits, due to low direct component. (e) Our result.

min-SW (8) is given by Goddyn *et al.* [6] (Figure 5). In comparison, conventional Gray codes have a min-SW of 2. Kim *et al.* [11] used a variant of Gray codes with large min-SW called the antipodal Gray codes to mitigate errors due to defocus. For conventional Gray codes, although short-range effects might result in incorrect binarization of the lower-order bits, the higher-order bits are decoded correctly. Thus, these codes can be used in the presence of short-range effects as well.

4.3. Ensemble of codes for general scenes

Global illumination in most real world scenes is not limited to either short or long range effects. Codes optimized for long-range effects would make errors in the presence of short-range effects and vice versa. In general, it is not straight-forward to identify which code to use without knowing the dominant error-inducing mode of light transport, which in turn requires a priori knowledge about scene.

We show that by projecting a small ensemble of codes optimized for different effects, we can handle a large class of optically challenging scenes, without a priori knowledge about scene properties. The key idea is that errors made by different codes are nearly random. Thus, if two different codes produce the same depth value, it must be the correct value with a very high probability. We project four codes optimized for different effects: two for long-range (the *XOR-04* codes and the *XOR-02* codes), and two for short-range (the Gray codes with maximum min-SW and the conventional Gray codes). We can find the correct depth value by comparing the depth values computed using the individual codes. If any two agree within a small threshold, that value is returned. If only the two Gray codes agree, we return the value computed by the Gray code with maximum min-SW, because they result in a better depth resolution. **The pseudo-code for the method** is given in Algorithm 1. MATLAB code can be downloaded from the project web-page [1].

Results: Figure 1 shows a scene consisting of a bowl on a marble slab. For phase-shifting, we project 18 patterns (3 frequencies, 6 shifts for each frequency). For modulated phase-shifting [4], we project 162 patterns (9 modulated patterns for each phase-shifting pattern). For our ensemble codes, we project a total of 41 patterns - 10 patterns for each of the 4 codes and 1 all-white pattern. Interestingly, by analyzing the errors made by the individual codes, we can get qualitative information about light-transport. Scene points where only the logical codes agree (marked in light-blue) indicate strong inter-reflections. On the other hand, scene points where only the two Gray codes agree (green) correspond to translucent materials (sub-surface scattering). Scene points where all the codes agree (maroon) do not receive much global illumination.

The scene in Figure 6 has inter-reflections inside the fruit basket and strong sub-surface scattering on the fruits. Modulated phase-shifting performs poorly on translucent materials, whereas conventional Gray codes and phase-shifting produce errors in the presence of inter-reflections. Reconstruction produced using our ensemble of codes has significantly reduced errors. In Figure 7, the goal is to reconstruct the shower curtain. The correct shape of the curtain is planar, without any ripples. Light diffuses through the curtain and is reflected from the background, creating long-range interactions. Conventional Gray codes and phase-shifting result in large errors. In this case, only the logical codes (optimized for long-range interactions) are sufficient to achieve a nearly error-free reconstruction, instead of the full ensemble.

For more results on a range of scenes, **please see the project web-page** [1]. Our techniques consistently outperform many existing schemes (Gray codes, phase-shifting, and modulated phase-shifting [4]).

5. Error detection and correction

The patterns presented in the previous section can successfully prevent a large fraction of errors. For highly challenging scenes, however, some errors might still be made (for example, see Figure 8). For building a reliable shape measurement system, it is critical to detect and correct these residual errors. Traditionally, error detection and correction strategies from communication theory have been adopted in the context of structured light. An example is the Hamming error correcting codes used by Minou *et al.* [12]. These techniques treat structured light coding-decoding as a signal transmission problem. Although good for handling random sensor/illumination noise, these codes can not handle the systematic errors made due to global illumination. In this section, we present strategies for detecting and correcting such errors.

Error detection The consistency check proposed in the previous section, in addition to preventing errors, can also be used for detecting errors. For a pixel, if none of the four codes agree, it is marked as an *error* pixel. It is possible that one of the four values might be the correct value. However, as there is an error correction stage, we take a conservative approach and classify such pixels as error pixels. Since no extra patterns need to be projected, the error detection stage does not place any overhead in terms of acquisition time.

Error correction To correct the errors, we iteratively collect additional images while illuminating only the scene points corresponding to the error pixels. This technique, based on the work of Xu *et al.* [19], progressively reduces the amount of global illumination, resulting in reduction of the error pixels. Consider the concave lamp made of shiny brushed metal given in Figure 8. This is a challenging object because of high-frequency specular inter-reflections. Conventional Gray codes can not reconstruct a large portion of the object. Separation based modulated phase-shifting [4] can not remove the high-frequency inter-reflections, resulting in large errors. Our ensemble of codes, while reducing the errors, can not reconstruct the object completely. By acquiring images in 2 extra iterations [4], we achieve a nearly perfect reconstruction. The mean absolute errors as compared to the ground truth for our result, conventional Gray codes and modulated phase-shifting are $1.2mm$, $29.8mm$ and $43.9mm$ respectively (height of the lamp $= 250mm$). The ground truth was acquired by manually binarizing the captured images.

It is important to note that for this error correction strategy to be effective, the error prevention and detection stages are critical. Since our techniques correctly reconstruct a large fraction of the scene in the

---

Algorithm 1 Structured Light Scanning in the Presence of Global Illumination

1. Project patterns and capture images for the 4 codes - two Gray codes (Conventional Gray and Gray codes with maximum min-SW), and the two logical codes (XOR02 and XOR04).

2. Compute depth values under the two Gray codes using conventional decoding and the two logical codes using the logical decoding (Section 4.1).

3. Compare the depth values. If any two codes agree, return that value as the correct depth. If the two Gray codes agree, return the value computed by the Gray codes with min-SW (Section 4.3).

4. **Error detection:** Mark the camera pixels where no two codes agree as error pixels (Section 5).

5. **Error correction:** Mask the patterns so that only the scene points corresponding to the error pixels are lit [19]. Repeat steps $1-5$ to progressively reduce the residual errors (Section 5).

---

first iteration itself, we require only a small number of extra iterations (typically 1-2) even for challenging scenes. In comparison, Ref. [19] requires a large number of iterations (10-20) and images (500-800). This is because it uses conventional Gray codes, which do not prevent errors in the first place. Secondly, its error detection technique, based on direct-global separation, is conservative. Consequently, if the direct component is low (for example, in the presence of sub-surface scattering), this technique may not converge.

6. Limitations

Our methods assume a single dominant mode of light transport for every scene point. If a scene point receives both large short-range and long-range effects, for example, inside of a translucent concave bowl, none of the codes will produce the correct result. Consequently, the combination technique and further error correction steps will not be able to retrieve the correct result. We have not considered the effects of volumetric scattering as it results in both short-range and long-range interactions. A future direction would be to design systems for reconstructing shape under volumetric scattering.

---

[4]In this case, we projected only the logical codes in subsequent iterations, thus requiring 81 images in total.

(a) Shower curtain   (b) Conventional Gray (11 images)

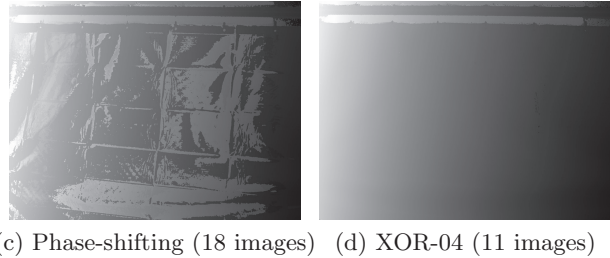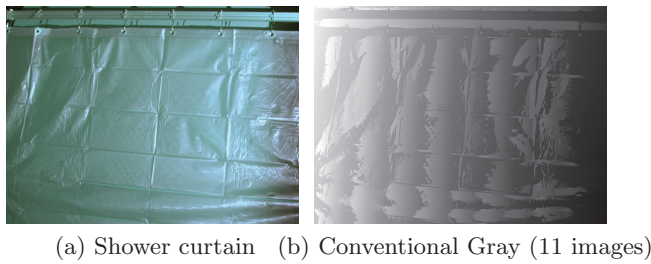(c) Phase-shifting (18 images)   (d) XOR-04 (11 images)

Figure 7. **Shower-curtain:** The correct shape of the curtain is planar, without ripples. Light diffuses through the curtain and is reflected from the background. (b-c) Conventional Gray codes and phase-shifting result in large errors. (d) Reconstruction using logical codes is nearly error-free, with same number of input images as conventional codes.

References

[1] Webpage. *http://graphics.cs.cmu.edu/projects/ StructuredLight3DScanning/.* 714, 717, 718

[2] M. K. Chandraker, F. Kahl, and D. J. Kriegman. Reflections on the generalized bas-relief ambiguity. In *CVPR*, 2005. 714

[3] T. Chen, H. P. A. Lensch, C. Fuchs, and H. peter Seidel. Polarization and phase-shifting for 3D scanning of translucent objects. In *CVPR*, 2007. 714

[4] T. Chen, H.-P. Seidel, and H. P. A. Lensch. Modulated phase-shifting for 3D scanning. In *CVPR*, 2008. 714, 718, 719, 720

[5] J. Ghring. Dense 3-D surface acquisition by structured light using off-the-shelf components. *Videometrics and Optical Methods for 3D Shape Measurement*, 4309, 2001. 713

[6] L. Goddyn and P. Gvozdjak. Binary gray codes with long bit runs. *The electronic journal of combinatorics*, 2003. 718

[7] G. Godin, J.-A. Beraldin, M. Rioux, M. Levoy, L. Cournoyer, and F. Blais. An assessment of laser range measurement of marble surfaces. In *Proc. of Fifth Conference on optical 3D measurement techniques*, 2001. 716

[8] M. Gupta, Y. Tian, S. G. Narasimhan, and L. Zhang. (De) Focusing on global light transport for active scene recovery. In *CVPR*, 2009. 714, 716

[9] C. Hermans, Y. Francken, T. Cuypers, and P.Bekaert. Depth from sliding projections. In *CVPR*, 2009. 714

[10] M. Holroyd, J. Lawrence, and T. Zickler. A coaxial optical scanner for synchronous acquisition of 3d geometry and surface reflectance. *ACM Trans. Graph.*, 29(3), 2010. 714

[11] D. Kim, M. Ryu, and S. Lee. Antipodal gray codes for structured light. In *ICRA*, 2008. 718

[12] M. Minou, T. Kanade, and T. Sakai. A method of time-coded parallel planes of light for depth measurement. *Transactions of IECE Japan*, 64(8), 1981. 713, 719

[13] S. Nayar, K. Ikeuchi, and T. Kanade. Shape from Inter-reflections. *IJCV*, 6(3), 1991. 714

[14] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Trans. Graph.*, 25(3), 2006. 714, 716

(a) Metallic lamp   (b) Conventional Gray (11 images)

(c) Modulated phase shifting [4] (162 images)   (d) Our ensemble codes (41 images)

(e) After error detection and correction (81 images)   (f) 3D visualization of (e)
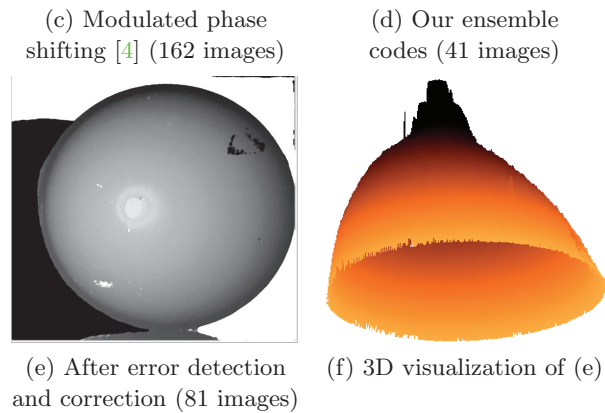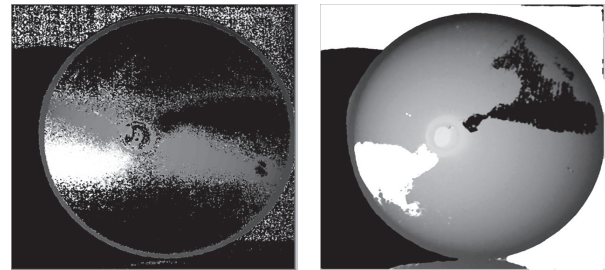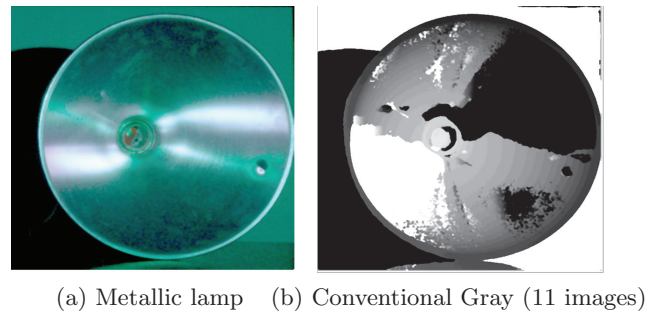
Figure 8. **Shiny metal lamp.** (a) A hemispherical lamp made of brushed shiny metal. (b-c) Both conventional Gray codes and modulated phase-shifting perform poorly due to strong and high-frequency inter-reflections. (d) Our ensemble codes reduce the errors significantly. (e) Using our error detection and correction, we get nearly perfect reconstruction. Please see text for ground truth comparisons.

[15] J. Park and A. C. Kak. 3D modeling of optically challenging objects. *IEEE Transactions on Visualization and Computer Graphics*, 14(2), 2008. 714

[16] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43, 2010. 713

[17] M. Trobina. Error model of a coded-light range sensor. *Technical Report*, 1995. 716

[18] P. M. Will and K. S. Pennington. Grid coding: A preprocessing technique for robot and machine vision. *Artificial Intelligence*, 2(3-4), 1971. 713

[19] Y. Xu and D. Aliaga. An adaptive correspondence algorithm for modeling scenes with strong interreflections. *IEEE TVCG*, 2009. 719

[20] L. Zhang and S. K. Nayar. Projection Defocus Analysis for Scene Capture and Image Display. *ACM Trans. Graph.*, 25(3), 2006. 714

[21] S. Zhang, D. V. D. Weide, and J. Oliver. Superfast phase-shifting method for 3-D shape measurement. *Optics Express*, 18(9), 2010. 713