International Journal of Computer Vision manuscript No.
(will be inserted by the editor)

A Practical Approach to 3D Scanning in the Presence of Interreflections, Subsurface Scattering and Defocus

Mohit Gupta · Amit Agrawal · Ashok Veeraraghavan · Srinivasa G. Narasimhan

Received: date / Accepted: date

Abstract Global illumination effects such as interreflections, diffusion and subsurface 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 present a practical 3D scanning system which works in the presence of a broad range of global illumination effects. First, we design binary 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.

The techniques presented in this paper do not require a priori knowledge of the light transport in the scene. They do not require explicit separation of the direct and global components of scene radiance, which reduces the number of required input images while also improving the reconstruction quality. Our techniques outperform many existing schemes while using significantly fewer images (12-42 versus 200-700) as compared to previous work dealing with global illumination. 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.

Keywords Structured light 3D scanning, interreflections, subsurface scattering, defocus, global illumination, projectors.

1 Introduction

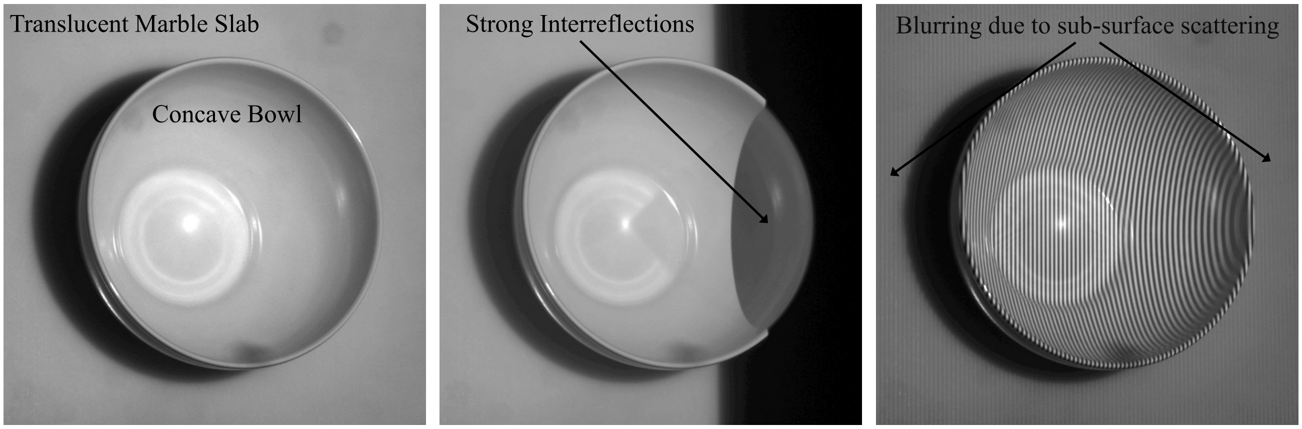
Structured light triangulation has become the method of choice for shape measurement in several applications including industrial automation, graphics, human-computer interaction and surgery. Since the early work in the field about 40 years ago [36, 24, 32], research has been driven by two factors: reducing the acquisition time and increasing the depth resolution. Significant progress has been made on both fronts (see the survey by Salvi *et al.* [33]) as demonstrated by systems which can recover shapes at close to 1000 Hz. [40] and at a depth resolution better than 30 microns [10].

Despite these advances, the applicability of most structured light techniques remains limited to *well behaved scenes*. It is assumed that 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 interreflections,

---

A preliminary version of this paper appeared in [15].

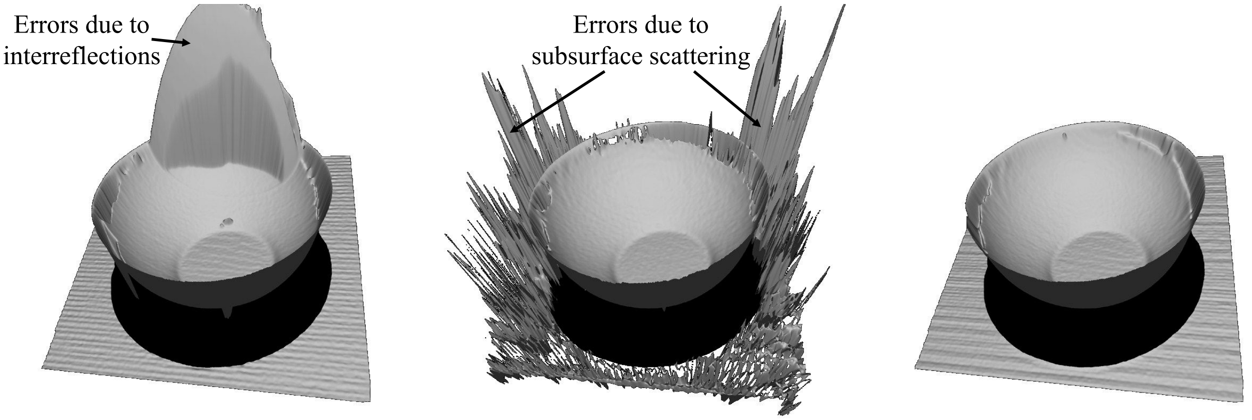
Mohit Gupta is with the Computer Science Department, Columbia University, New York, NY, 10027.
E-mail: mohitg@cs.columbia.edu
Amit Agrawal is with the Mitsubishi Electrical Research Labs, Cambridge, MA, 02139.
E-mail: agrawal@merl.com
Ashok Veeraraghavan is with the Electrical and Computer Engineering Department, Rice University, Houston, TX, 77005.
E-mail: vashok@rice.edu
Srinivasa G. Narasimhan is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 15213.
E-mail: srinivas@cs.cmu.edu



(a) Concave bowl on a translucent marble slab

(b) Input image under low frequency illumination

(c) Input image under high frequency illumination



(d) Shape recovered using Conventional Gray codes (11 images)

(e) Shape recovered using Modulated phase shifting [7] (162 images)

(f) Shape recovered using our ensemble codes (42 images)

Fig. 1 Measuring shape for the 'bowl on marble-slab' scene. This scene is challenging because of strong interreflections inside the concave bowl and subsurface scattering on the translucent marble slab. (b) Scene points inside the bowl which are not directly illuminated receive substantial irradiance due to interreflections. (d) This results in systematic errors in the recovered depth. (c) Due to subsurface scattering on the translucent marble slab, high-frequency illumination patterns are severely blurred. (e) This results in large depth errors on the marble-slab. (f) Our technique uses an ensemble of codes optimized for individual light transport effects, and results in an accurate shape reconstruction. Parentheses contain the number of input images. **For more results and detailed comparisons to existing techniques, please see the project web-page [1].**

subsurface 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 errors in the recovered shape (see Figure 1b). Because of the systematic nature of these errors [2], it is hard to correct them in post-processing.

The goal of this paper is to build an end-to-end system for structured light 3D scanning under a broad range of global illumination effects. The focus is on designing the projected patterns (coding) and decoding schemes. In particular, we consider binary structured light patterns, which are perhaps the simplest to implement and widely used in several research and commercial systems. The key observation is that different global illumination effects place contrasting constraints on the spatial frequencies of projected structured light patterns. In particular, interreflections result in errors for low frequency structured light patterns [3]. On the

---

[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.

[2] In photometric stereo, interreflections result in a shallow but smooth reconstruction [28,27]. In structured light 3D scanning, interreflections result in local errors.

[3] Strictly speaking, since all binary patterns have step edges, all of them have high spatial frequencies. For the analysis and

other hand, local effects such as subsurface scattering and defocus blur the high frequency patterns, making it hard to decode them reliably.

We design patterns that modulate global illumination and prevent the errors at capture time itself. We show that it is possible to construct codes with only high frequency binary patterns by introducing the concept of *logical coding and decoding.* The key idea is to express low frequency patterns as pixel-wise logical combinations of two high-frequency patterns. Because of high frequencies, these patterns are resilient to long range effects. In order to deal with short range effects, we use tools from combinatorial mathematics to design patterns consisting solely of low frequencies. In comparison, most currently used patterns (e.g., Gray codes) contain a combination of both low and high spatial frequencies, and thus are ill-equipped to deal with global illumination.

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. How do we handle scenes that exhibit more than one type of global illumination effect (such as the one in Figure 1(a)? To answer this, we observe that the probability of two different codes producing the same erroneous decoding is very low. This observation allows us to project a small ensemble of codes and use a simple voting scheme to compute the correct decoding at every pixel, without any prior knowledge about the scene (Figure 1(d)).

Finally, for highly challenging scenes, we present an error detection scheme based on a simple consistency check over the results of the individual codes in the ensemble. We then use an error correction scheme which further reduces the errors due to global illumination by selectively re-illuminating only the incorrectly reconstructed scene points [37]. 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 methods can be readily incorporated into existing systems without significant overhead in terms of acquisition time or hardware. We believe that these techniques are important steps towards making 3D scanning techniques applicable to a large class of complex, real world scenarios.

2 Related Work

Structured light 3D Scanning: 3D scanning using structured light is one of the oldest computer vision techniques. Since the first papers [36, 24, 32], a lot of progress has been made in terms of reconstruction speed, accuracy and resolution. Broadly, these techniques are divided into discrete [20] and continuous [39] coding schemes. For an exhaustive survey on structured light techniques, reader is referred to the survey by Salvi et al [33]. In addition, hybrid techniques that combine structured light with photometric stereo based techniques have been proposed as well [29, 2].

Shape recovery in the presence of global illumination: The seminal work of Nayar *et al.* [27] presented an iterative approach for reconstructing shape of Lambertian objects in the presence of interreflections. Liu *et al.* [23] proposed a method to estimate the geometry of a Lambertian scene by using the second bounce light transport matrix. Gupta *et al.* [17] presented methods for recovering depths using projector defocus [38] under global illumination effects. Chandraker *et al.* [5] use interreflections to resolve the bas-relief ambiguity inherent in shape-from-shading techniques. Holroyd *et al.* [19] 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.* [31, 30] move the camera or the scene to mitigate the errors due to global illumination in a structured light setup. Hermans *et al.* [18] 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 that are applicable for a wide range of scenes, and which require a single camera and a projector, without any moving parts.

Nayar *et al.* showed that the direct and global components of scene radiance could be efficiently separated [28] using high-frequency illumination patterns. This has led to several attempts to perform structured light scanning under global illumination [6, 7, 13]. 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 interreflections (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

---

discussion in this paper, low frequency patterns implies patterns with thick stripes. Similarly, high frequency patterns mean patterns with only thin stripes.

approaches, or rely on weak cues like polarization. Recently, Couture *et al.* [8] proposed using band-pass unstructured patterns to handle interreflections. Their approach involves capturing a large number (200) of images with random high-frequency patterns projected on the scene. In contrast, we explicitly design ensembles of illumination patterns that are resilient to a broader range of global illumination effects (interreflections, subsurface scattering, defocus, diffusion, and combinations of multiple effects), while using significantly fewer images.

Shape recovery in other optically challenging scenarios: Active illumination has also been used to measure density distribution of volumetric media [3, 14] and reconstruct transparent objects [34,25]. For a detailed survey on techniques for reconstructing transparent and specular surfaces, please refer to the state of the art report by Ihrke et al [21]. There have also been techniques for performing 3D scanning in the presence of volumetric media using light striping [26,16]. Our techniques can not handle volumetric scattering. The focus of this work is on reconstructing opaque and translucent surfaces with complex shapes.

3 Errors due to Global Illumination

In this section, we analyze errors in structured light based depth recovery caused due to different global illumination effects. The basic principle behind shape from structured light techniques is triangulation. Each projector row/column is encoded with a unique spatial or temporal code. Projector illuminates the scene with the assigned code and camera takes a sequence of images, one for each projected pattern. For each camera pixel, the corresponding projector row/column is found by decoding the measured intensity values. The depths are then computed by intersecting the camera ray with the plane containing the corresponding projector row/column and projector center.

The resulting depth estimate is incorrect is there is an error in estimating the correspondence. The form 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 (subsurface scattering). We call these short-range effects. Some points may receive global illumination from a larger region (interreflections or diffusion). We call these long range effects. As shown in Figures 2 and 3, long range effects and short range effects result in incorrect decoding of low and high spatial frequency patterns, respectively. We analyze these errors for the case of binary structured light patterns.

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 [4]. 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:

$$L^i = L_d^i + \beta\, L_g^i \ , \tag{1}$$

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

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 the pattern P.

In the presence of defocus (projector or camera), the projected patterns and the captured image is blurred. Similarly, aberrations due to imperfect projector optics also result in blurring of the projected patterns. The blur influences the highest frequency patterns, often completely blurring them out [5]. Defocus, unlike global illumination effects, modulates the direct component as well, as shown in [17]:

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

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

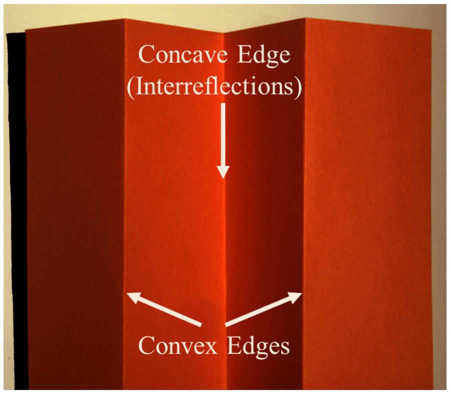
The fractions (α and $1-\alpha$) depend on the projected pattern and the amount of defocus. In the absence of defocus, $\alpha = 1$. **For correct binarization, it is required that** $L_i > \overline{L_i}$, i.e.

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

---

[4] The inverse pattern can be generated by subtracting the image from image of the fully lit scene.

[5] For example, pico-projectors are increasingly getting popular for structured light applications in industrial assembly lines. However, due to imperfect optics, they can not resolve patterns with thin stripes, for example, a striped pattern of 2-pixel width.



Scene with interreflections: Concave V-groove

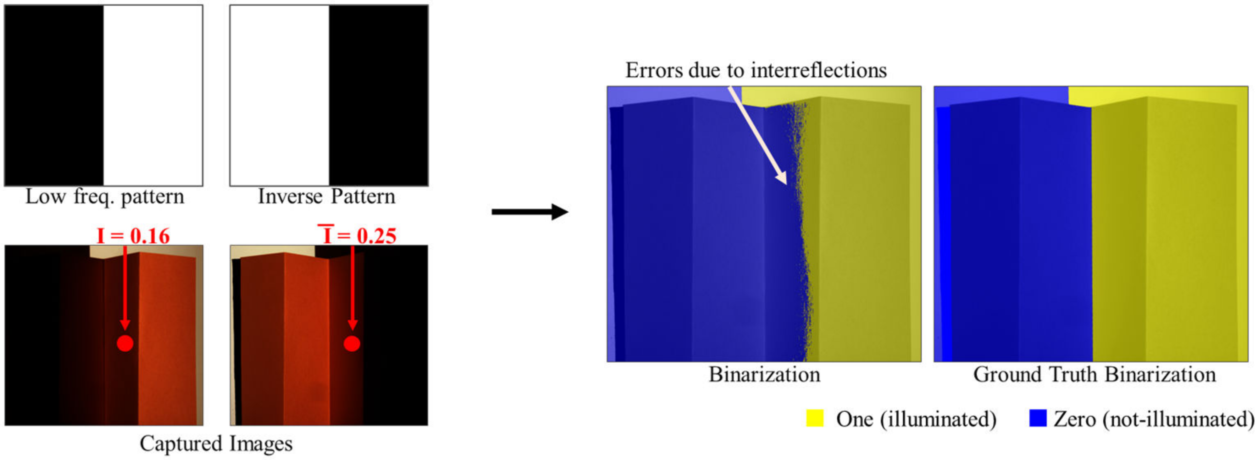
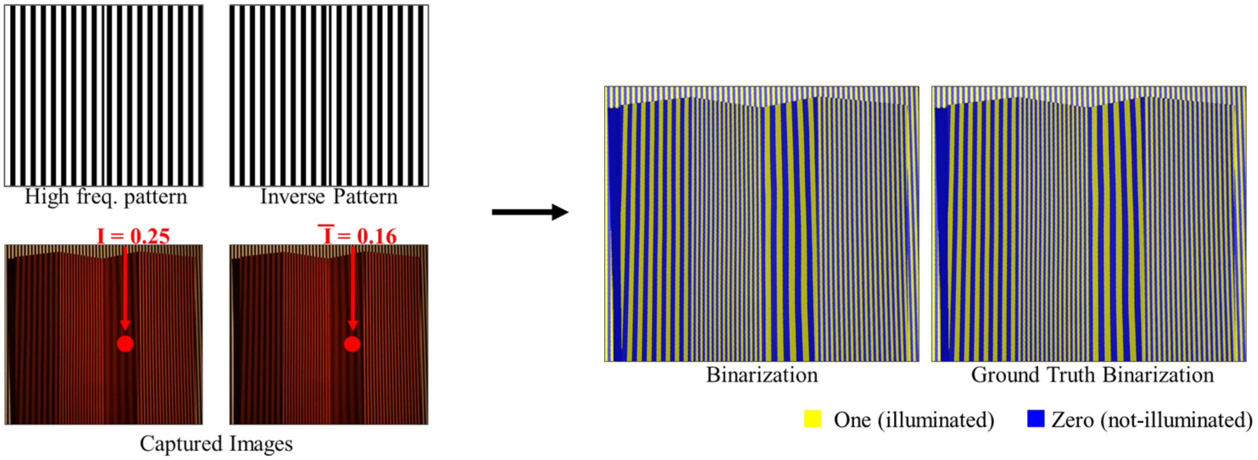


Illustration of errors due to interreflections: The scene point marked with red is directly illuminated when the low frequency pattern is projected on the scene. It is not directly lit when the inverse pattern is projected. However, due to interreflections, its intensity is higher when it is not directly lit (0.25) as compared to when it is directly lit (0.16). This results in a decoding (binarization) error, as shown on the right. Scene points decoded as one (directly illuminated) are marked in yellow and points decoded as zero (not illuminated) are marked in blue. In the correct decoding, only the points to the left of the concave edge should be zero.



Decoding for high-frequency patterns: High frequency patterns are decoded correctly even in the presence of interreflections. See Section 3.1 for a detailed explanation.

Fig. 2 Structured light decoding in the presence of interreflections. Top: A concave v-groove. **Middle:** Illustration of structured light decoding errors due to interreflections. Interreflections result in low-frequency patterns being decoded incorrectly. **Bottom** High-frequency patterns are decoded correctly.

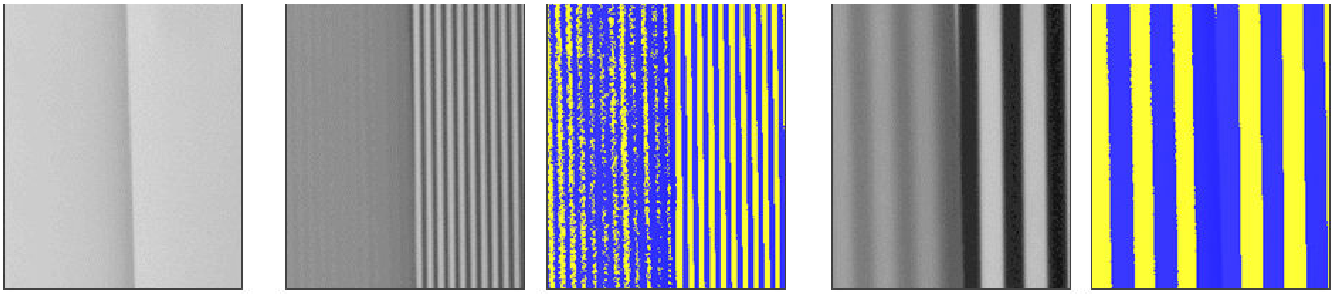

(a) (b) (c) (d) (e)

Fig. 3 Structured light decoding in the presence of subsurface 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. Consequently, binarization information can not be extracted reliably on the marble slab (c). In contrast, the image captured (d) under a low-frequency pattern is binarized (e) more accurately.

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[6].

3.1 Long range effects (diffuse and specular interreflections)

Consider the scenario when a scene point 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 arise due to long-range interreflections, when scenes are illuminated with low-frequency patterns. This is because low frequency patterns illuminate the scene asymmetrically. 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.

In contrast, if the scene is illuminated with a high-frequency pattern, the corresponding captured image is binarized correctly even in the presence of interreflections. This is explained as follows. If a high-frequency pattern (with equal off and on pixels) is projected on the scene, scene points receive approximately half the global component, i.e., $\beta \approx 0.5$ [28]. Thus, for a scene point S_i, $L^i \approx L_d^i + 0.5 L_g^i$ and $\overline{L^i} \approx 0.5 L_g^i$. Consequently, S_i, $L^i \geq \overline{L^i}$, and the condition for correct binarization is satisfied. An example is shown in Figure 2.

3.2 Short-range effects (subsurface 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, $\beta \approx 0.5$ [28]. 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 subsurface scattering ($L_d^i \approx 0$) or because of severe defocus ($\alpha \approx 0.5$), the pattern can not be binarized robustly. An example is shown in Figure 3.

In the presence of short range effects, most of the global illumination at a scene point 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. As a result, an image captured under low-frequency pattern is binarized more reliably.

For conventional Gray codes, the high-frequency patterns correspond to the lower significance bits. Loss of information in the high-frequency patterns results in a loss of depth resolution. For example, if patterns of width less than 5 pixels can not be resolved, last 2 bits of information will be lost if using the conventional Gray codes. An example is shown in Figure 6.

In summary, long and short range effects respond differently to the spatial frequencies of the incident il-

---

[6] Errors for the particular case of laser range scanning of translucent materials are analyzed in [12]. Errors due to sensor noise and spatial mis-alignment of projector-camera pixels were analyzed in [35].

lumination. In the presence of long-range effects, low-frequency patterns are susceptible to incorrect binarization, whereas high-frequency patterns are decoded correctly. On the other hand, for short-range effects, high-frequency patterns are susceptible to coding errors while the low-frequency patterns are decoded accurately.

4 Patterns for Error Prevention

Errors due to global illumination are systematic, scene-dependent errors that are hard to eliminate in post-processing. In this section, we design patterns that modulate global illumination and prevent errors from happening at capture time itself. As discussed in the previous section, 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. On the other hand, in the presence of short-range effects, low-frequency patterns are decoded more reliably.

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). However, 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? In this section, 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 binary codes with large minimum stripe-widths, resulting in patterns with low spatial frequencies.

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 4. 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 ($\mathbb{B}$):

$$f : \mathbb{P} \Rightarrow \mathbb{B} . \tag{6}$$

For a given pattern $P \in \mathbb{P}$, this function returns a binarization of the captured image if the scene is illuminated by P. As we saw earlier, under interreflections, 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^1_{hf} and P^2_{hf} using a pixel-wise binary operator $\odot$ such that:

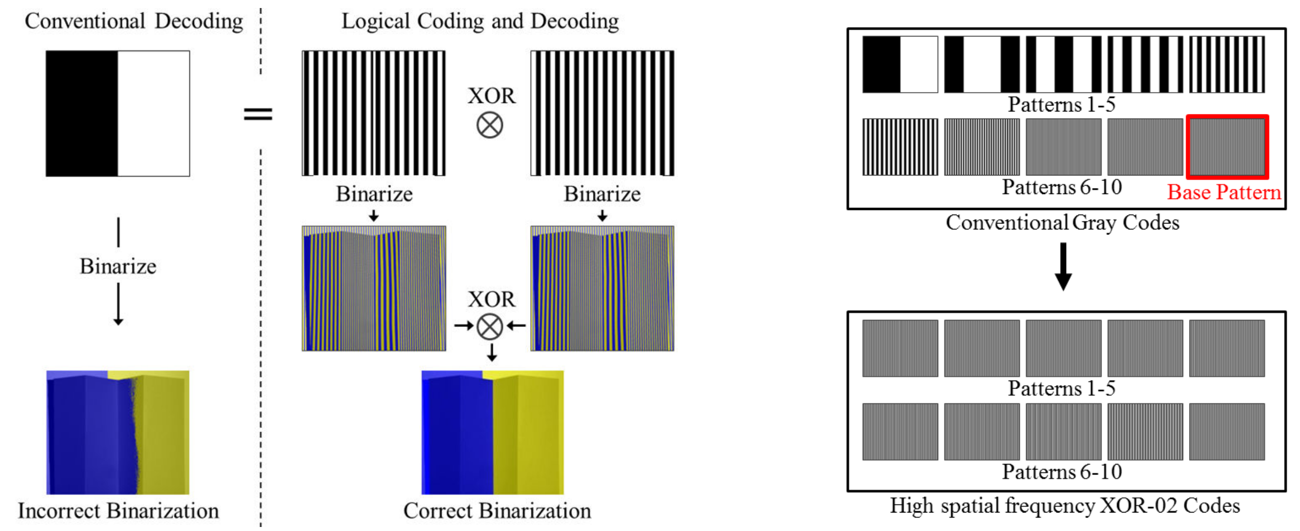
$$f(P_{lf}) = f\left(P^1_{hf} \odot P^2_{hf}\right) = f\left(P^1_{hf}\right) \odot f\left(P^2_{hf}\right) \tag{7}$$

If we find such a decomposition, we can robustly compute the binarizations $f\left(P^1_{hf}\right)$ and $f\left(P^2_{hf}\right)$ under the two high frequency patterns, and compose these to achieve the correct binarization $f\left(P_{lf}\right)$ 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 ($\otimes$) because it has the following property:

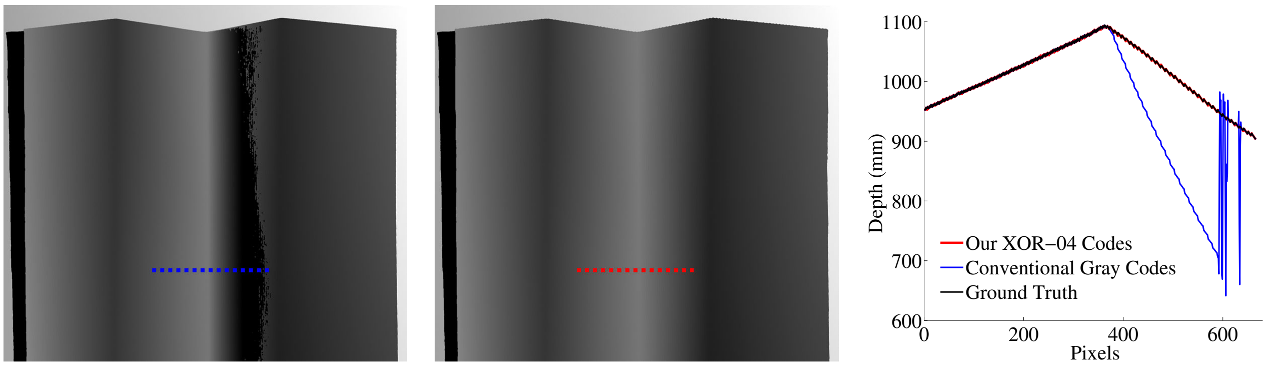
$$P^2_{hf} \otimes P^1_{hf} = P_{lf} \Rightarrow P^2_{hf} = P_{lf} \otimes P^1_{hf} \tag{8}$$

This choice of operator provides a simple means to decompose P_{lf}. We first choose a high-frequency pattern P^1_{hf}. The second pattern P^2_{hf} is then computed by simply taking the pixel-wise logical XOR of P_{lf} and P^1_{hf}. We call the first high frequency pattern the *base* pattern. Instead of the original low frequency pattern, the two high-frequency patterns P^1_{hf} and P^2_{hf} are projected on the scene. The corresponding captured images are binarized. The two binarizations are then combined by performing another pixel-wise logical XOR operation. This produces the correct binarization as if the scene was illuminated by the original low frequency pattern. An example is shown in Figure 4.

The logical patterns are constructed by taking the pixel-wise logical XOR of a high-frequency pattern (base plane or base pattern) in the conventional Gray codes with all other patterns. This is illustrated in Figure 4. The resulting patterns have only high spatial frequencies. Note that there is no overhead introduced; the number of projected patterns remains the same as the conventional codes. If the last Gray code pattern is chosen as the base plane, the resulting codes are called logical XOR-02 codes. All the projected patterns have a maximum stripe width of 2 pixels. In contrast, the original Gray codes have a maximum stripe-width of 512 pixels.



(a) Conventional Decoding vs. Logical Coding and Decoding **(b) Generation of Logical XOR Patterns**



(c) Depth (conventional Gray codes) Mean absolute error = 28.8mm (d) Depth (our Logical XOR04 codes) Mean absolute error = 1.4mm (e) Comparison with the ground truth

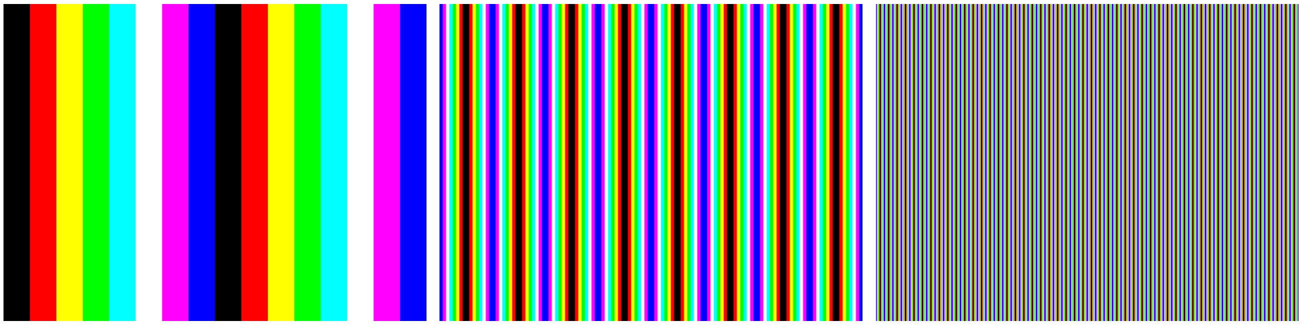
Fig. 4 Logical coding and decoding for long range global light transport: (a) In logical coding and decoding, a low frequency pattern is expressed as a pixel-wise logical combination (e.g., XOR) of two high frequency patterns. The high-frequency patterns are projected on the scene and the captured images are binarized. The two binarizations are then combined by performing another pixel-wise logical operation (XOR). This produces the correct binarization as if the scene was illuminated by the original low frequency pattern. **(b)** The logical patterns can be constructed by taking the pixel-wise logical XOR of a high-frequency pattern (base plane) in the conventional Gray codes with all other patterns. The resulting patterns have only high spatial frequencies. The number of projected images remains the *same*. If the last pattern is chosen as the base plane, the resulting codes are called logical XOR-02 codes. If the second-to-last pattern is used as the base plane, the resulting codes are called logical XOR-04 codes. **(c)** Depth map computed with the conventional codes. Because of incorrect binarization of the low frequency patterns (higher-order bits), depth map has large and systematic errors. Because of their systematic nature, these cannot be removed by simple smoothing in post-processing. **(d)** Depth map computed using our logical XOR-04 codes. The errors due to interreflections have been significantly reduced. **(e)** Comparison with the ground-truth. Ground truth was computed by manually binarizing the captured images.

If the second-to-last pattern is used as the base plane, the resulting codes are called logical XOR-04 codes. The last pattern is projected unmodified. In these codes, all the projected patterns have a maximum stripe-width of 4 pixels. In general, if the $(n-k)^{th}$ pattern is used as the base plane, the resulting codes would be logical XOR-2^{k+1} codes. The maximum stripe width would be 2^{k+1} pixels and the last $k-1$ planes would be projected unmodified. The patterns for logical XOR-02 and XOR-04 codes are shown in Figure 7. The pattern images can be downloaded from the project web-page [1].

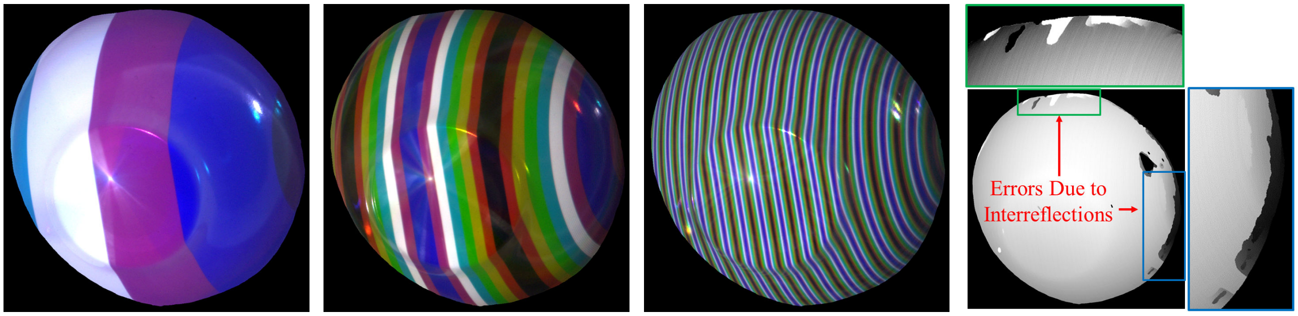
Color Logical XOR Codes: Next, we discuss the use of color to design patterns. Color patterns reduce the number of required input images [7] as compared to binary patterns. We show that it is possible to construct

---

[7] The color of the incident illumination can be decoded from the image of the illuminated scenes on a per-pixel basis, even for non-white scenes [4]. It is not required to assume spatial smoothness or color neutrality of the scene.

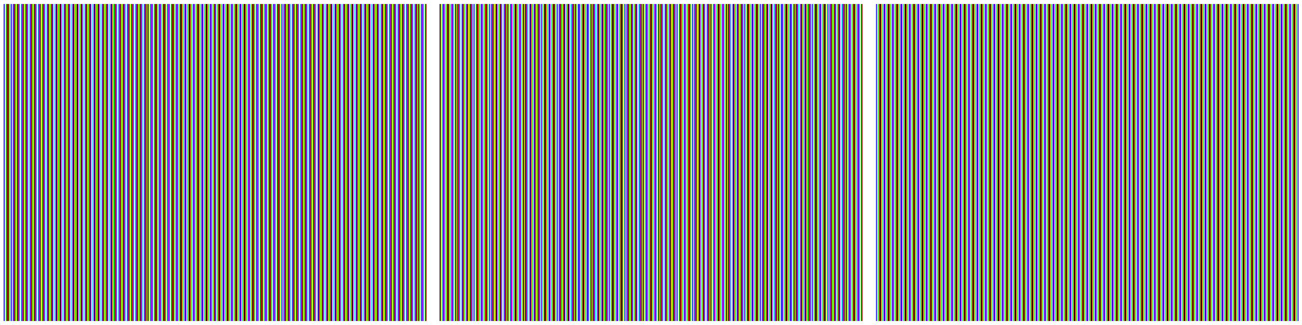


(a-c) Color Gray codes

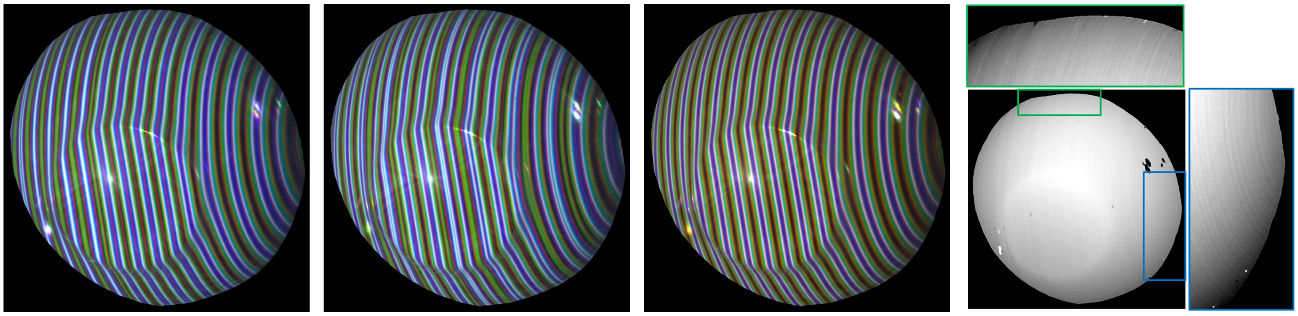


(d-f) Input images for concave bowl using color Gray codes

(g) Depth map



(h-j) Color Logical XOR codes



(k-m) Input images for concave bowl using color Logical XOR codes

(n) Depth map

Fig. 5 Color Gray codes vs. color Logical XOR codes. (a-c) Color Gray codes for a projector with 512 columns. Since the number of possible distinct color values at each projector pixel is 8, only 3 patterns are required to encode 512 distinct columns. In contrast, binary coding requires 9 patterns. **(d-f)** Input images of a concave bowl for projected patterns (a-c). **(g)** Computed depth map. Due to low frequencies in the projected patterns, interreflections result in erroneous reconstruction near the periphery of the bowl. Please zoom in for details. **(h-j)** Color Logical XOR codes, constructed by performing logical XOR operations on the color Gray codes. All the patterns have high spatial frequencies. **(k-m)** Input images of a concave bowl for projected patterns (h-j). **(n)** Computed depth map. Errors due to interreflections have been significantly mitigated. Most of the residual errors result from pixel saturation due to specularities.

color logical XOR codes as well by performing logical operations, similar to the binary case.

Given 3 color channels, we consider the case where each color channel at a projector pixel has a binary value. Thus, each projector pixel can take 8 possible color values - {RGB} = {000, 001, 010, 011, 100, 101, 110, 111}. For example, if a projector pixel is encoded as {100}, its red channel will be 1, and the green and blue channels will be 0. In case of color codes, $N_{color} = \lceil \log_8(M) \rceil$ patterns are required to uniquely encode M projector columns. In contrast, under binary coding schemes, $N_{bin} = \lceil \log_2(M) \rceil$ patterns are required to encode M different projector columns. For example, if $M = 512$, $N_{color} = 3$. In contrast, $N_{bin} = 9$.

Figure 5 (a-c) shows color Gray codes for a projector with 512 columns. These codes were generated using the K-ary ($K = 8$) reflected Gray code construction [9]. Figures 5 (d-f) show input images of a concave bowl under the color Gray codes. Due to low frequencies in the projected patterns, interreflections result in erroneous reconstruction near the periphery of the bowl.

In order to construct color logical XOR codes, we start with color Gray codes. It has been shown that by performing a color calibration between projector and camera [4], the color transfer matrix between the projector and the camera can be made a diagonal matrix, and each color channel can be treated independently. With this observation, the color logical XOR codes can be constructed in a similar way as binary codes. First, we choose a base plane. In our experiments, we chose the highest frequency pattern from color Gray codes as the base plane. The remaining color XOR codes are made by taking the pixel-wise logical XOR of the base plane with other color Gray code patterns, **for each color channel independently:**

$$X_c^i = G_c^i \otimes G_c^1 \ , \qquad (9)$$

for $c = \{R, G, B\}$, $i = \{2 : N_{color}\}$. X_c^i is the c^{th} color channel of the i^{th} pattern of the color Logical XOR codes. G_c^i is the c^{th} color channel of the i^{th} pattern of the color Gray codes. G_c^1 is the c^{th} color channel of the base plane. The captured images are first binarized in each color channel independently [8] and then combined by performing a pixel-wise logical XOR operation in each color channel. This produces the $K-ary$ (in this case, $K = 8$) decoding as if the scene was illuminated by the original low-frequency patterns. Figure 5 (h-j) show the color Logical XOR codes constructed using the algorithm described above. All the patterns have high spatial frequencies. Figures 5 (k-n) show the corresponding input images of the concave bowl and the computed depth map. Errors due to interreflections have been significantly mitigated. The MATLAB code for generating the patterns and decoding the input images is provided on the project web-site [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. In order to be resistant to local blurring due to short-range effects, patterns with low spatial frequencies must be designed. For binary patterns, this corresponds to designing patterns with large minimum stripe-width. In general, it is not feasible to find such codes with a brute-force search as these codes are extremely rare[9].

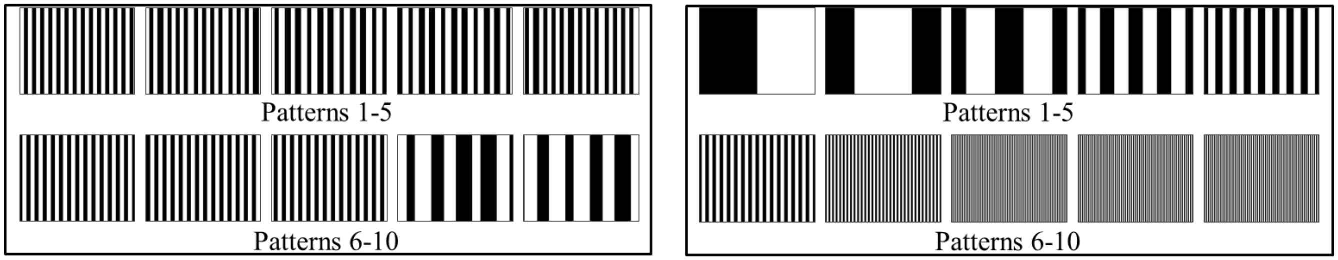
Fortunately, this problem has been well studied in combinatorial mathematics. There are constructions available to generate codes with large minimum stripe-widths (min-SW). Kim et al [22] used a variant of Gray codes with large min-SW called the antipodal Gray codes to mitigate errors due to defocus. The 10-bit binary Gray code with the maximum known min-SW (8) is given by Goddyn *et al.* [11]. We call these codes the maximum min-SW Gray codes. These codes are shown in Figure 6 and 7. The algorithm to construct these codes is given in [11]. A limitation is that these algorithms do not generalize well for general K-ary codes. For the remaining paper, we will consider only binary codes. The MATLAB code to generate these codes can be downloaded from the project web-site [1].

In comparison, conventional Gray codes have a min-SW of 2. For Gray codes, increasing the minimum stripe-width also serves the dual purpose of reducing the maximum stripe-width. Thus, maximum min-SW Gray codes have a maximum stripe width of 32 pixels. Consequently, these codes, while being resistant to short-range effects, are also more resistant to long range effects as compared to the conventional Gray codes. Figure 6 shows a scene consisting of industrial parts. A pico-projector was used to illuminate the scene. Due to defocus, the high frequency patterns in the conventional Gray codes can not be decoded reliably, resulting in a loss of depth resolution. In contrast, depth map computed using maximum

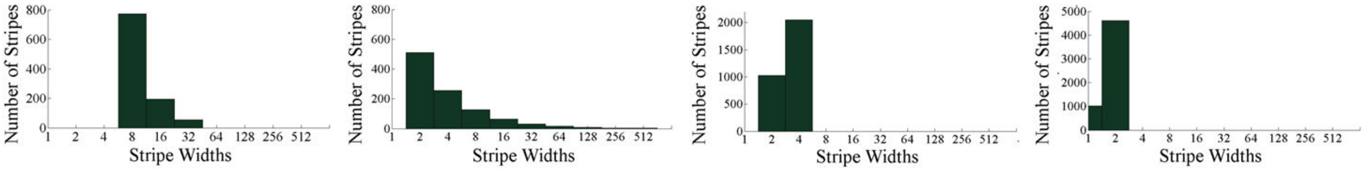
---

[8] Two additional images of the scene, one under all white illumination, and one under all black illumination were acquired to establish the per-pixel intensity thresholds for binarization.

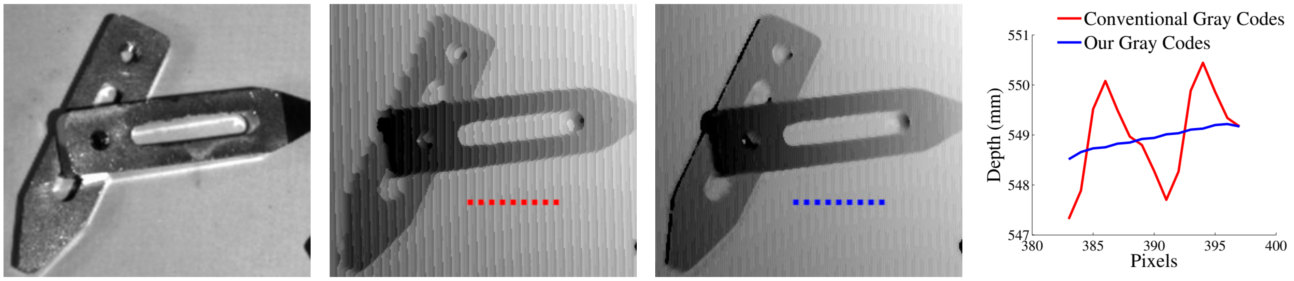
[9] It is relatively easy to generate codes with small maximum stripe-width (9), as compared to 512 for the conventional Gray codes, by performing a brute-force search.



(a) Maximum min-SW Gray codes
(b) Conventional Gray codes



(c) Histogram of stripe-widths Maximum min-SW Gray codes
(d) Histogram of stripe-widths Conventional Gray codes
(e) Histogram of stripe-widths Logical XOR-04 codes
(f) Histogram of stripe-widths Logical XOR-02 codes



(g) Scene consisting of industrial parts
(h) Depth map using conventional Gray codes
(i) Depth map using maximum min-SW Gray codes
(j) Depth map comparison

Fig. 6 Designing patterns for local effects: (a) Local effects such as subsurface scattering and defocus result in blurring of incident illumination. For such effects, patterns with low spatial frequencies must be designed. We used tools from combinatorial mathematics literature to design binary patterns which maximize the minimum stripe width. These patterns are called maximum min-SW Gray codes. (b) Conventional Gray codes. (c-d) Histograms of stripe-widths for different patterns. For the maximum min-SW Gray codes, all the stripes have widths in the range [8, 32] pixels. In contrast, the range of stripe-widths for conventional Gray codes, [2, 512] pixels, is significantly larger. For XOR-04 and XOR-02 codes, the ranges are [2, 4] and [1, 2] pixels respectively. (g) A scene consisting of industrial parts. (h) Due to defocus, the high frequency patterns in the conventional Gray codes can not be decoded reliably, resulting in a loss of depth resolution. Notice the quantization artifacts. (i) Depth map computed using Gray codes with large minimum stripe-width (min-SW) does not suffer from loss of depth resolution.

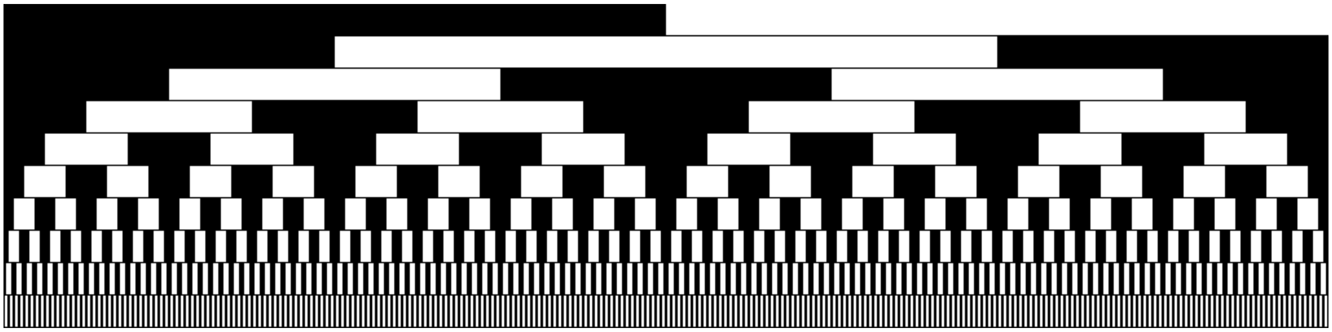
min-SW Gray codes does not suffer from loss of depth resolution.

5 Ensemble of codes for general scenes

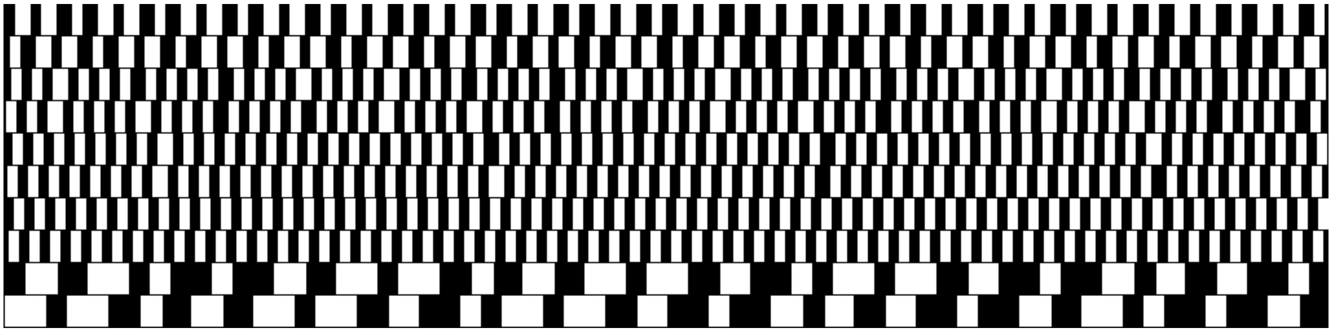
So far, we have designed codes optimized for long or short range effects. In general though, it is not straightforward to identify which code to use without knowing the dominant error-inducing mode of light transport. This, in turn, requires a priori knowledge about scene. Moreover, global illumination 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 this section, we address the question: how can we handle general real world scenes which can have both short and long range global illumination effects?

5.1 Depth recovery algorithm using ensemble of codes

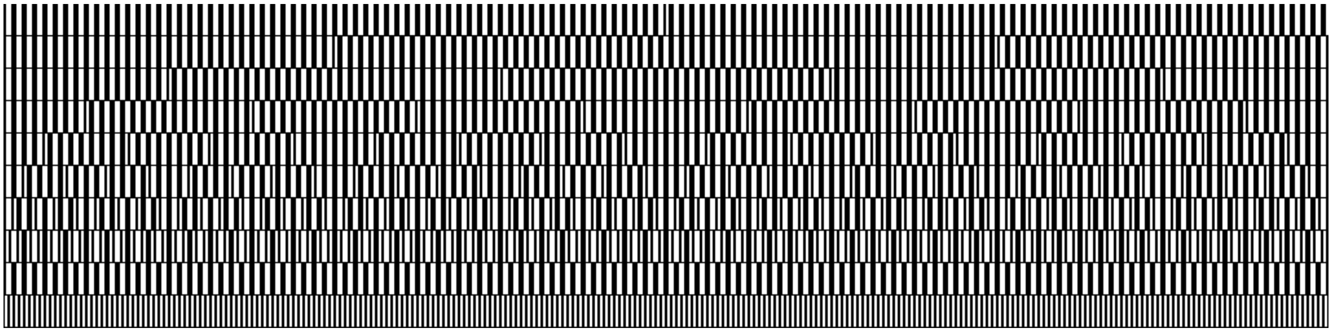
We show that by projecting a small ensemble of codes optimized for different effects, it is possible to handle a large class of optically challenging scenes, without a priori knowledge about scene properties. The *key idea* is that errors made by different codes are nearly random. Thus, if the depth values computed using two different codes is the same, with a very high probability, it must be the correct value. Using this observation, we



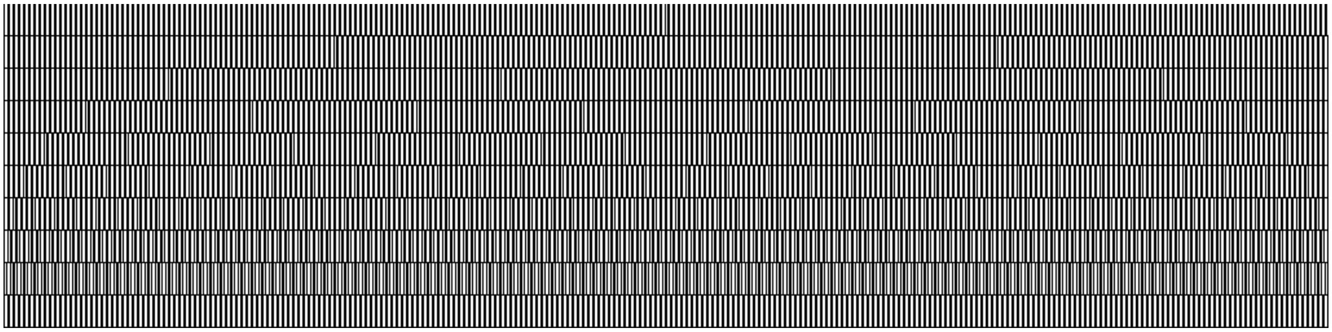
Conventional Gray Codes



Maximum Min-SW Gray Codes



Logical XOR-04 Codes



Logical XOR-02 Codes

Fig. 7 Visualization of different binary coding schemes: The patterns are for a projector with resolution 768 × 1024; thus, each scheme has 10 patterns. For each scheme, each row in the figure represents one pattern. Conventional Gray codes have a wide range of stripe-widths- $[2, 512]$ pixels. The range for maximum min-SW Gray codes is $[8, 32]$ pixels. For logical XOR-04 and XOR-02 codes (optimized for long range effects), the ranges are $[2, 4]$ and $[1, 2]$ pixels respectively. Pattern images are available for download from the project web-site [1].

propose a simple depth recovery algorithm which uses an ensemble of codes.

In particular, we project four different codes: two optimized for long-range effects (the *XOR-04* and the

XOR-02 codes), and two codes for short-range effects (the Gray codes with maximum min-SW and the conventional Gray codes). Each code returns a depth map of the scene, as shown in Figure 8 (a-d). The final depth value is computed by performing a simple consistency check across the depth values computed using the individual codes. If any two depth values are within a small threshold, that value is returned [10].

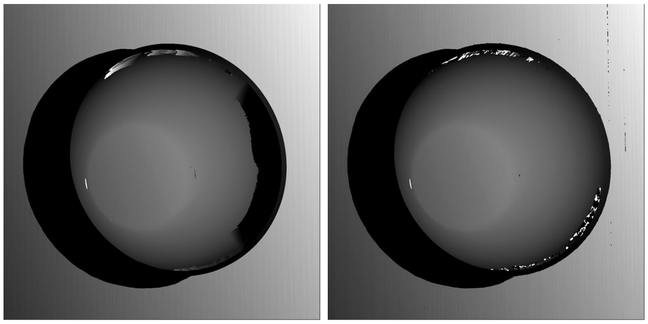
Intuitively, the two long-range codes produce the correct depth value in the presence of long-range effects, and the short-range codes produce the correct value in the presence of short-range effects. Since there are two codes each for long and short-range effects, the consistency check will *pick* the correct depth value. Note that conventional Gray codes might lose depth resolution in the presence of defocus of subsurface scattering. Therefore, if only the two Gray codes agree, we return the value computed by the maximum min-SW Gray codes, because it has a higher depth resolution.

Figure 8 (e) shows the depth map computed using the above algorithm. While the individual codes have significant errors due to either interreflections or subsurface scattering, the final depth map is nearly error-free. The 3D reconstruction of the scene is shown in Figure 1. **The pseudo-code for the method** is given in Algorithm 1. MATLAB code can be downloaded from the project web-page [1].

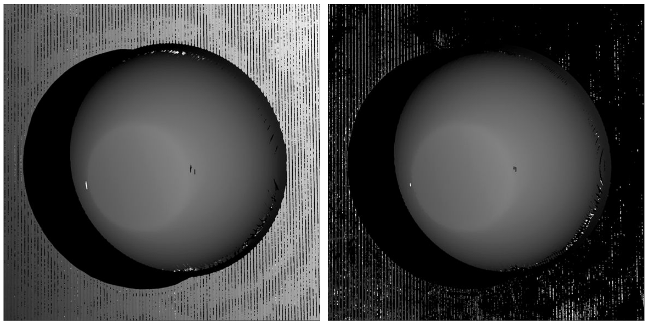
In the following, we show that the probability of two different codes making the same error, i.e., two different codes producing the same incorrect depth value, is very low. Readers not interested in the detailed error analysis can skip Section 5.2 and go directly to results in Section 6.

5.2 Error analysis of the code ensemble algorithm

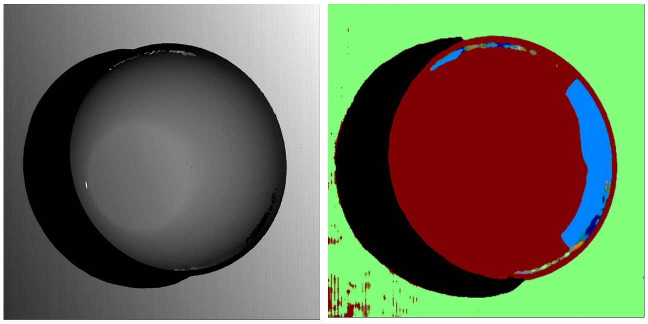
Assume, without loss of generality, that the intensity coding is along the x-dimension of the projector image plane, i.e., vertical stripes are projected. Therefore, each projector column has a unique code. For binary patterns, the code is binary. In this case, if the total number of projector columns is M, the code needs to have N bits, where $N = \lceil \log_2(M) \rceil$. N binary patterns are projected on the scene and the camera captures N images, one for each projected pattern.



(a) Conventional Gray codes (b) Maximum min-SW Gray codes



(c) XOR-04 codes (d) XOR-02 codes



(e) Code ensemble algorithm (f) Qualitative light transport

Fig. 8 Code ensemble algorithm to reconstruct scenes with multiple global illumination effects: The scene is the same as shown in Figure 1. We project four different codes - two logical XOR codes and the two Gray codes. Depth estimates using individual codes (a-d) have errors due to global illumination effects. (e) The code ensemble algorithm performs a simple consistency check to compute a depth map with significantly fewer errors. (f) 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 strong interreflections.

Let the projector column number a has an N bit binary code C_a^S. S denotes the coding scheme. In this paper, $S \in \{CG, MM-SW, XOR02, XOR04\}$, corresponding to conventional Gray, maximum min-SW Gray, logical XOR02 and logical XOR04 codes, respectively. Suppose a pixel in the column a directly illuminates the camera pixel x. Let the vector of intensity values at x be I_x^S. In order for the correct correspondence

---

[10] Due to imperfect projector optics, insufficient camera/projector resolution or misalignment between projector and camera pixels, the depth results from individual codes might suffer from spatial aliasing. This problem is more pronounced for the high-frequency XOR codes. To prevent aliasing from affecting the final depth estimate, we apply a median filter (typically 3×3 or 5×5) to the individual correspondence maps before performing the consistency check.

to be established, the code C_a^S should be recovered from I_x^S. However, various factors such as sensor noise, or illumination fluctuations or defocus and global illumination effects can result in some of the bits *flipping* from 0 to 1 or vice versa. This results in a decoding error. Let the recovered code be C_b^S. We assume that the observation noise in different observations is independent. Thus, flipping of each bit in the code is independent of other bits. Then, the probability of code C_a^S getting decoded incorrectly as C_b^S is

$$Pr[C_a^S \rightarrow C_b^S] = p^d \ , \tag{10}$$

where p is the probability of one bit flipping and d is the hamming distance between C_a^S and C_b^S, $0 \leq d \leq N$. p is a function of sensor noise characteristics, illumination levels, scene albedos and light transport in the scene. A small value of p implies that the decoding is reliable. On the other hand, a large value of p indicates unreliable decoding. If we pose the problem of structured light as a communication problem, p would denote the reliability of the *communication channel* between the projector and the camera.

We have assumed p to be constant for all bit positions. In general, since the errors due to global illumination are structured, p is different for different bit positions. For example, for conventional Gray codes in the presence of interreflections, since low frequency patterns (higher significance bits) are more likely to be decoded incorrectly, p is more for higher significance bit positions as compared to lower significance bits. Computing the p values for different codes would require knowing the scene structure a priori. One possibility is to simulate the structured light decoding process by rendering several scenes with light transport effects and sensor noise models. While such an approach can provide estimates of the value of p, it is beyond the scope of this paper. The goal of our analysis is to show that the probability of two different coding schemes making the same error is very low, for a wide range of values of p. If different p values are estimated for different bit positions, the analysis presented here can be done in a similar way.

We now define the confusion matrix M^S for a coding scheme S as $M^S(a,b) = Pr[C_a^S \rightarrow C_b^S]$, where a and b are two projector columns. $M^S(a,b)$ is the probability of C_a^S being decoded incorrectly as C_b^S. This matrix is a measure of error resilience of a given coding scheme. In order to be the most error resistant, the confusion matrix should be a diagonal matrix. Note that the confusion matrix is a function of p, the probability of a single bit-flip.

Figure 9 shows the confusion matrices for the four coding schemes that we consider, for two different values of p. As expected, for a low value of p, the matrix is nearly diagonal for all the schemes. However, for a large value of p, the off-diagonal terms are comparable to the near-diagonal terms. This can result in large decoding errors. Note that the structure of the confusion matrices for the logical XOR codes is similar to the conventional Gray codes as the former are derived from the latter.

The code ensemble algorithm (Section 5.1) results in an error if the *same* decoding error happens for two different schemes. For the camera pixel x, suppose the correct corresponding projector column is a. The joint probability of the column a being incorrectly decoded as the column b, for two different coding schemes $S1$ and $S2$ is

$$\begin{aligned} &Pr[(C^S 1_a \rightarrow C^S 1_b) \ \& \ (C^S 1_a \rightarrow C^S 1_b)] = \\ &Pr[C^S 1_a \rightarrow C^S 1_b] \ . \ Pr[C^S 1_a \rightarrow C^S 1_b] \, . \end{aligned} \tag{11}$$

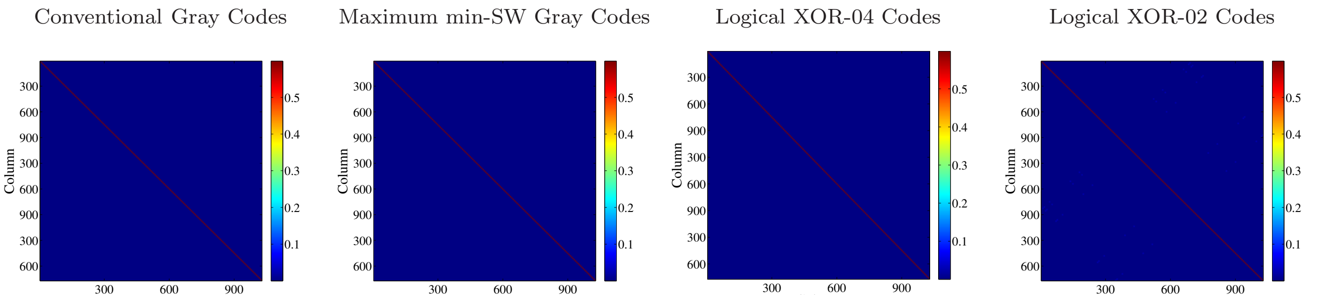
This follows from the independence of the image acquisition and the decoding process for the two schemes. These probabilities form the joint error probability matrix $P^{(S1,S2)}$, where $P^{(S1,S2)}(a,b) = M^{S1}(a,b) \times M^{S2}(a,b)$. Figure 10 shows the matrices for 6 pairs of schemes. The off-diagonal values are small; most of them being of the order of 10^{-6}. Finally, we note that a column a can be incorrectly decoded as any other column b. So, the probability that the code ensemble algorithm will result in a decoding error for the column a is the sum of the a^{th} row of the matrix $P^{(S1,S2)}$

$$P^{(S1,S2)}(a) = \sum_b P^{(S1,S2)}(a,b) \, . \tag{12}$$

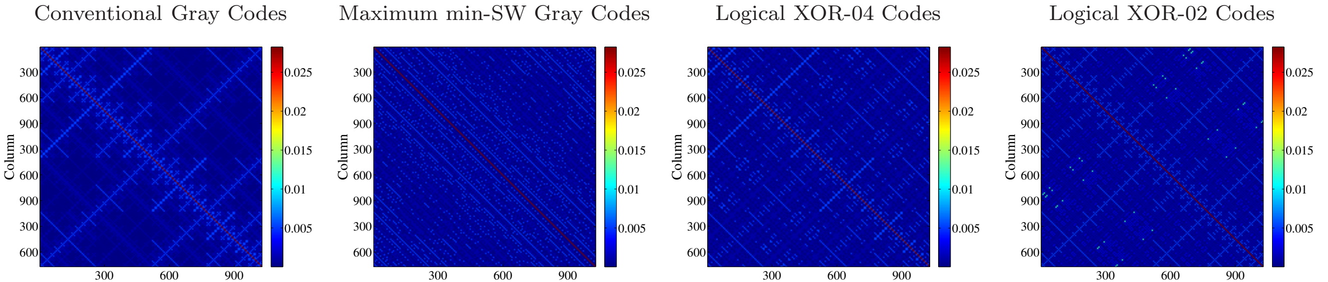
Figure 10 shows the plots for $P^{(S1,S2)}(a)$ with respect to a for different pairs of schemes. Note that most of the probability values are less than 1%. Figure 11 shows the mean probability of error for different pairs of schemes, where the mean is taken over all the projector columns. Most of the values are less than 1%, with the maximum being 1.4%.

Mean depth error: A decoding error results in an incorrect depth estimate. The magnitude of the depth error is directly proportional to the column error $|a-b|$, where a is the correct column number and b is the decoded (incorrect) column number. The expected column error $E^{(S1,S2)}$ for a pair of schemes $S1$ and $S2$ is

$$E^{(S1,S2)} = \frac{1}{M} \sum_{a,b} |a-b| \, P^{(S1,S2)}(a,b) \ , \tag{13}$$



p = 0.05



p = 0.3

Fig. 9 Confusion matrices for different coding schemes. Confusion matrix gives the probabilities of a projector column being decoded incorrectly as another projector column, under a given coding scheme. In order for the scheme to be the most error resistant, the confusion matrix should be a diagonal matrix. We use the confusion matrices of individual coding schemes to perform error analysis of our code ensemble algorithm (Section 5). **Top row:** Confusion matrices for $p = 0.05$, where p is the probability of a single binary bit (of the N bit code) flipping. p is a function of the noise of the imaging and illumination system, scene albedos and light transport in the scene. For a low value of p, the confusion matrices for all the schemes are nearly diagonal. **Bottom row:** Confusion matrices for $p = 0.3$. Because of a high value of p, the off-diagonal terms are comparable to the diagonal terms.

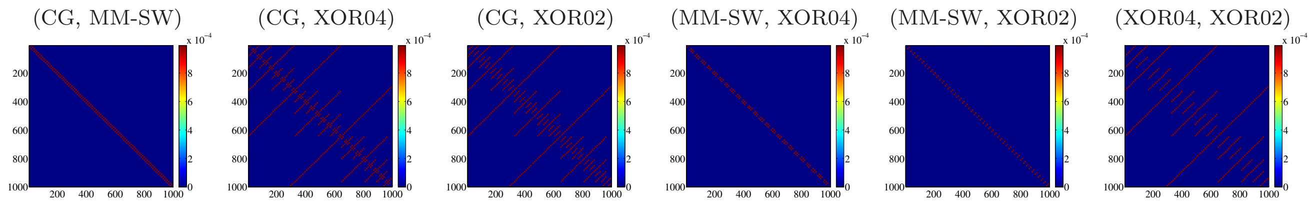
where M is the total number of projector columns. Figure 12 shows the mean column decoding error for different pairs of schemes, under different noise levels. Most of the errors are less than 1 pixel, with the maximum being 1.67 pixels. While this analysis was done for a projector with 1024 columns, it can be extended in a similar way for a different number of columns.

6 Experiments and Results

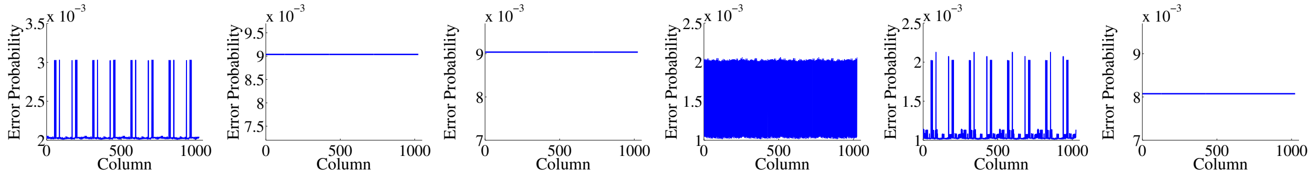
In our experiments, for phase-shifting, we project 18 patterns (3 frequencies, 6 shifts for each frequency). For modulated phase-shifting [7], we project 162 patterns (9 modulated patterns for each phase-shifting pattern). For our ensemble codes, we project a total of 42 patterns - 10 patterns for each of the 4 codes, 1 all-white pattern and 1 all-black patterns. Images captured under the all-white and all-black illumination patterns are used to establish per-pixel intensity thresholds for binarization.

Scenes with subsurface scattering and defocus: Figure 6 shows a scene consisting of industrial parts. Due to defocus, the high frequency patterns in the conventional Gray codes can not be decoded reliably, resulting in a loss of depth resolution. Depth map computed using maximum min-SW Gray codes does not suffer from loss of depth resolution. Figure 13 and 14 shows objects and scenes with strong subsurface scattering. Translucent materials are often characterized by low direct component. Since modulated phase shifting [7] relies on explicitly separating the direct and the global components, it suffers from low signal-to-noise-ratio for highly translucent materials. The resulting depth maps are severely degraded due to noisy. On the other hand, our code ensemble does not rely on explicit direct-global separation, resulting in significantly better reconstructions.

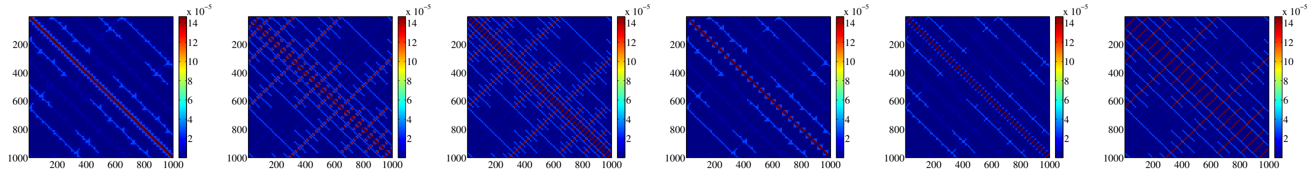
Scenes with diffusion: Next, we consider scenes which have only long range effects. Figures 15 and 16 show scenes comprising thin, nearly transparent surfaces. In both cases, light diffuses through the material and is reflected from the background/interior, creating long-range optical interactions. Consequently, conventional Gray codes and phase-shifting result in large errors in the reconstructed shape. In contrast, our codes achieve high quality reconstructions. For some moderately difficult scenes, such as the shower curtain in Figure 15,



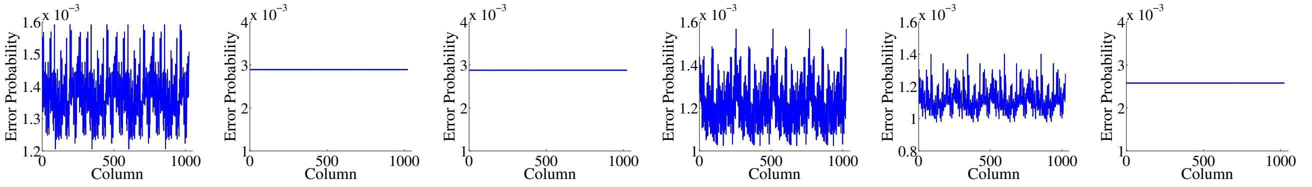
Joint error probability matrices for different pairs of coding schemes **(p=0.05)**



Error probabilities for our code ensemble algorithm **(p=0.05)**



Joint error probability matrices for different pairs of coding schemes **(p=0.3)**



Error probabilities for the code ensemble algorithm **(p=0.3)**

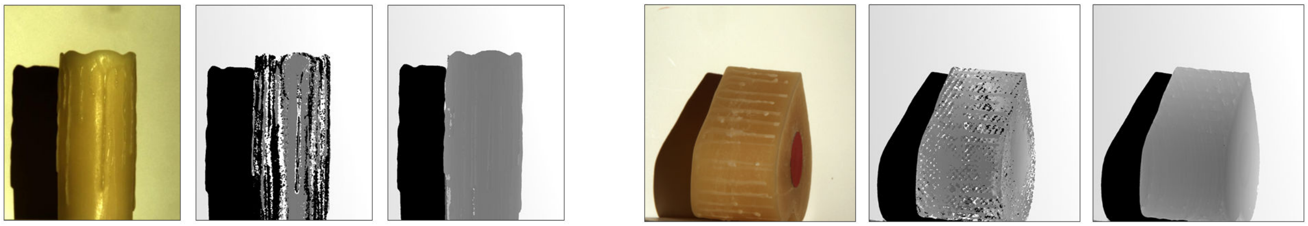
Fig. 10 **Error analysis of our code ensemble algorithm (Section 5.1). First and third rows:** Joint error probability matrices for different pairs of schemes, for different values of p. Most of the off-diagonal values are of the order of 10^{-6}. **Second and fourth rows:** Sum of rows of the joint error probability matrices. The resulting plots are the probabilities that the code ensemble algorithm will result in a decoding error for the corresponding projector column. Most of the probability values are less than 1%.

	(CG, MM-SW)	(CG, XOR04)	(CG, XOR02)	(MM-SW, XOR04)	(MM-SW, XOR02)	(XOR04, XOR02)
p=0.05	0.2%	0.9%	0.9%	0.2%	0.1%	0.8%
p=0.1	0.3%	1.4%	1.4%	0.3%	0.2%	1.2%
p=0.3	1.4%	0.3%	0.3%	0.1%	0.1%	0.3%
p=0.5	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%

Fig. 11 **Mean error probabilities for the code ensemble algorithm.** This table gives the mean probabilities of a pair of schemes making the same decoding error. These are computed by taking the mean of the error probabilities for all the projector columns (Figure 10, second and fourth rows). Most of the values are less than 1%, with the maximum being 1.4%.

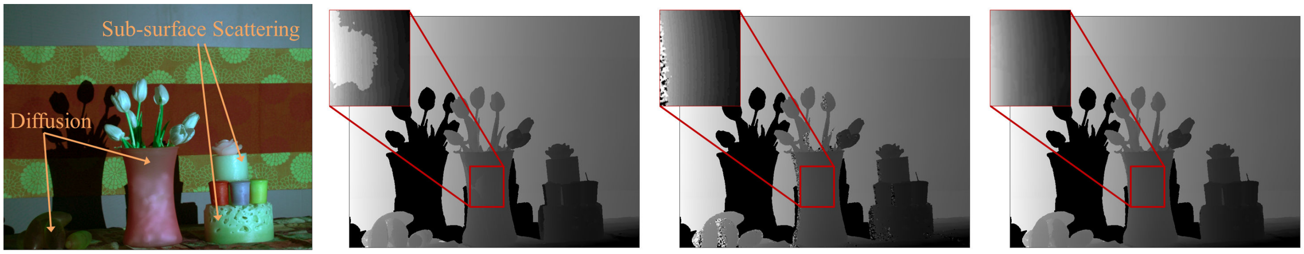
	(CG, MM-SW)	(CG, XOR04)	(CG, XOR02)	(MM-SW, XOR04)	(MM-SW, XOR02)	(XOR04, XOR02)
p=0.05	0.01	1.03	1.03	0.02	0.02	1.06
p=0.1	0.06	1.67	1.67	0.07	0.07	1.74
p=0.3	0.25	0.52	0.51	0.29	0.28	0.56
p=0.5	0.33	0.33	0.33	0.33	0.33	0.33

Fig. 12 **Mean decoding error (in pixels) for the code ensemble algorithm.** A decoding error results in an incorrect depth estimate. The magnitude of the depth error is directly proportional to the column error $|a - b|$, where a is the correct column number and b is the decoded (incorrect) column number. This table gives the mean column decoding errors for different pairs of schemes. Most of the errors are less than 1 pixel, with the maximum being 1.67 pixels.



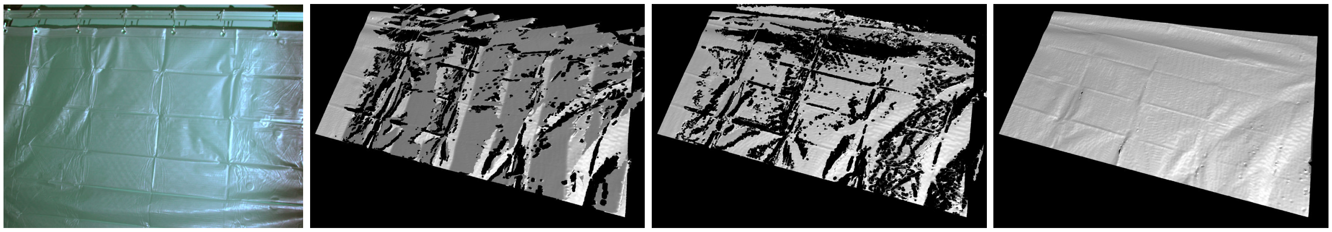
(a) Object (b) Modulated PS (c) Code ensemble (d) Object (e) Modulated PS (f) Code ensemble

Fig. 13 Reconstructing translucent wax objects: Translucent materials (a,d) are often characterized by low direct component. Since modulated phase shifting relies on explicitly separating the direct and the global components, it suffers from low signal-to-noise-ratio for highly translucent materials. The resulting depth maps are very noisy (b,e). On the other hand, our code ensemble does not rely on explicit direct-global separation, resulting in better quality reconstructions (c,f).



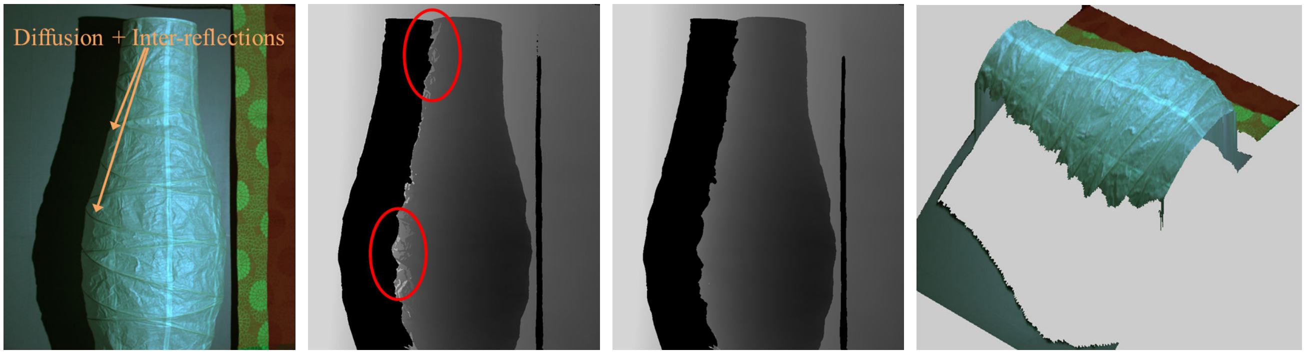
(a) Candles and flower vase (b) Conventional Gray codes (c) Modulated phase shifting (d) Our code ensemble

Fig. 14 Candles and flower-vase: (a) The scene consists of a flower vase and some wax candles. The flower vase is made of stained glass, resulting in diffusion of light. The wax candles have subsurface scattering. Depth from phase shifting (b) has errors on the flower vase, while modulated phase shifting results in errors on the candles (c). Depth map using our code ensemble (d) is nearly error free. **Please zoom in for details.**



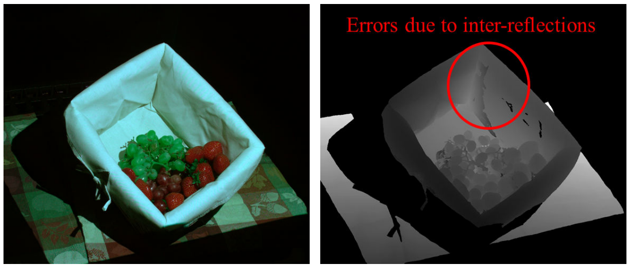
(a) Shower curtain (b) Conventional Gray codes (c) Phase shifting (d) The XOR-04 codes

Fig. 15 Shower-curtain: (a) Light diffuses through the curtain and is reflected from the background, creating long-range optical interactions. Consequently, (b) conventional Gray codes and (c) phase-shifting result in large errors and holes in the estimated shape. The correct shape of the curtain is nearly planar, with small ripples. (d) Reconstruction using our logical XOR-04 codes is nearly error free, with the same number of input images as the conventional Gray codes.

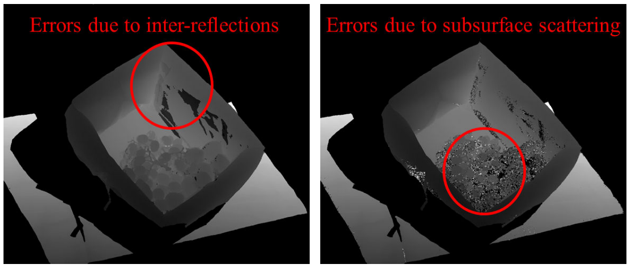


(a) Ikea lamp (b) Conventional Gray codes (c) Our code ensemble (d) Visualization of (c)

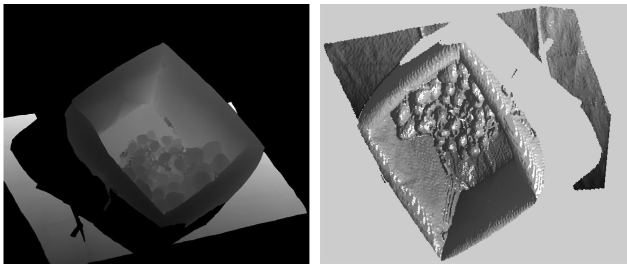
Fig. 16 Reconstructing an Ikea lamp: The lamp is made of thin translucent paper. Light diffuses inside the lamp, bounces inside and comes back out. (b) Conventional Gray codes result in errors near the periphery of the lamp. (c) Depth map using our code ensemble. (d) 3D visualization of (c).



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



(c) Phase shifting (18 images) (d) Modulated phase shifting [7] (162 images)

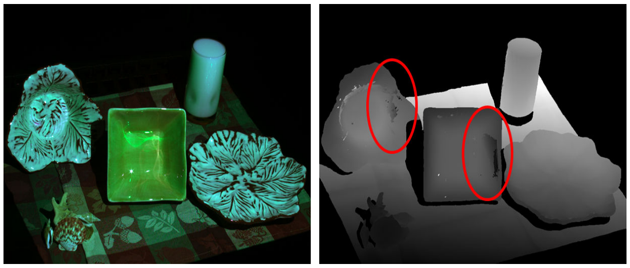


(e) Code ensemble (42 images) (f) 3D Visualization for (e)

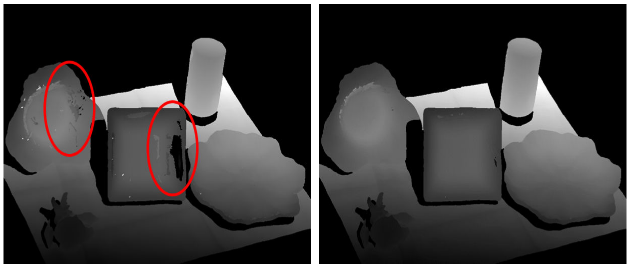
Fig. 17 Measuring 3D shape of a fruit-basket (interreflections+subsurface scattering). (a) This scene has both interreflections (corner of the fruit-basket) and subsurface scattering on the fruits. (b-c) Conventional Gray codes and phase-shifting result in errors due to interreflections. (d) On the other hand, modulated phase shifting produces errors on the translucent fruits due to low direct component. (e) Our technique using an ensemble of codes results in significantly fewer errors. Parentheses contain the number of input images.

it is sufficient to use only one of our codes, instead of the full ensemble.

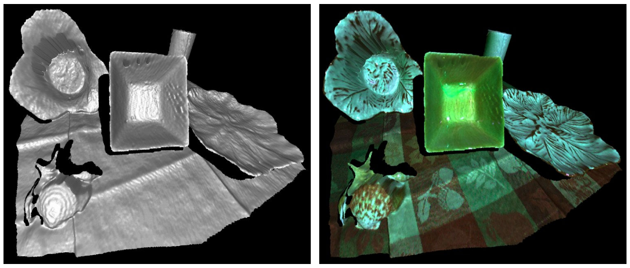
Scenes with multiple global illumination effects: Now, we show scenes which have multiple global illumination effects (but each scene point receives either long or short range effects). Figures 1 and 8 show a scene consisting of a bowl on a marble slab. Depth estimates using individual codes (Figures 8(a-d)) have errors due to various global illumination effects. The depth estimate using our code ensemble has significantly fewer errors. Corresponding 3D reconstructions are shown in Figure 1. Interestingly, by analyzing the errors made by the individual codes, we can infer qualitative information about light-transport, as shown in Figure 8 (f). Points marked in green correspond to translucent materials. Points marked in light-blue receive strong interreflections.



(a) Bowls and milk (b) Conventional Gray (11 images)



(c) Phase-shifting (18 images) (d) Code ensemble (42 images)

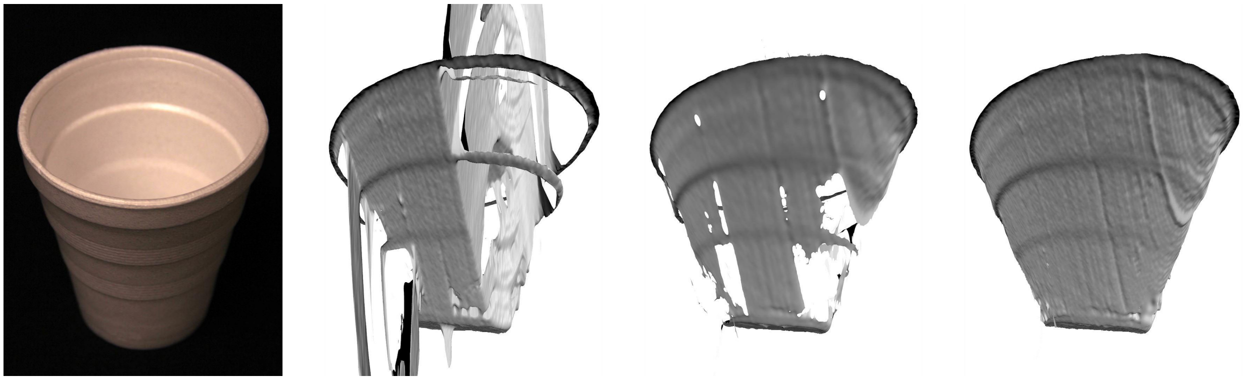


3D Visualizations for (d)

Fig. 18 Depth map computation for the bowls and milk scene (interreflections+subsurface scattering). (b) Conventional Gray codes and (c) phase-shifting result in errors at points receiving strong interreflections. (d) Result of our code ensemble.

The scenes in Figures 17 and 18 have both interreflections and subsurface scattering. Modulated phase-shifting performs poorly on translucent materials, whereas conventional Gray codes and phase-shifting produce errors in the presence of interreflections. In contrast, reconstruction produced using our ensemble of codes has significantly reduced errors.

Finally, we consider scenes which have points that receive both short and long range effects. Figure 19 shows results for a cup made of styrofoam. Since styrofoam is weakly translucent, points inside the cup receive both subsurface scattering and strong interreflections. Conventional Gray codes produce large errors in the recovered shape. The spatial frequencies of our max min-SW Gray codes are not sufficiently high to prevent errors. However, accurate shape is recovered us-



(a) Styrofoam cup (b) Conventional Gray codes (c) Max min-SW Gray codes (d) Code ensemble

Fig. 19 Styrofoam cup (interreflections+subsurface scattering): Styrofoam is weakly translucent. Points inside the cup receive both subsurface scattering and strong interreflections. (b) Conventional Gray codes produce large errors in the recovered shape. (c) The spatial frequencies of our max min-SW Gray codes are not sufficiently high to prevent errors. (d) Accurate shape is recovered using our code ensemble because of the presence of high-frequency XOR-02 and XOR-04 codes.

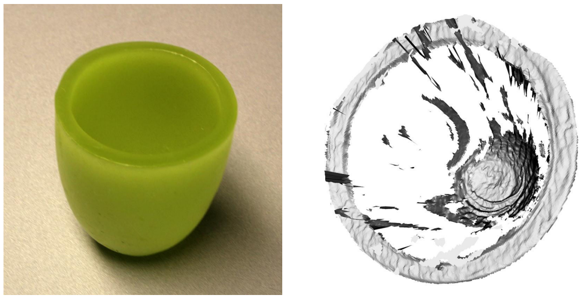
ing our code ensemble because of the presence of high-frequency XOR-02 and XOR-04 codes. Figure 20 shows depth recovery results for a wax bowl. Points inside the bowl receive strong subsurface scattering. Since the interreflections are weak (the bowl is shallow), the code ensemble produces an accurate shape. For more results and high-resolution images, **see the project webpage** [1].

Figure 21 shows a failure case - a deep container made of highly translucent wax. In this case, points inside the container receive both strong interreflections and strong subsurface scattering. Since none of the four codes compute the correct shape, the code ensemble fails to reconstruct the object.

Comparisons with Couture *et al.* [8]: Recently, Couture *et al.* [8] proposed an approach to deal with interreflections by projecting a large number (200) of random high-frequency patterns on the scene. Figure 22 shows comparisons of their approach with our XOR-04 codes, which are designed to deal with interreflections, and have similar spatial frequencies as used in [8]. Since all three scenes have strong interreflections, the conventional Gray codes result in large errors. The random high-frequency codes successfully remove the errors. The XOR-04 codes produce results of the same accuracy, while requiring an order of magnitude fewer images (12 versus 200).

(a) Wax bowl (b) Code ensemble

Fig. 20 Wax bowl (interreflections+subsurface scattering): (a) Points inside the bowl receive weak interreflections (the bowl is shallow) and strong subsurface scattering. (b) Shape computed using the code ensemble algorithm.

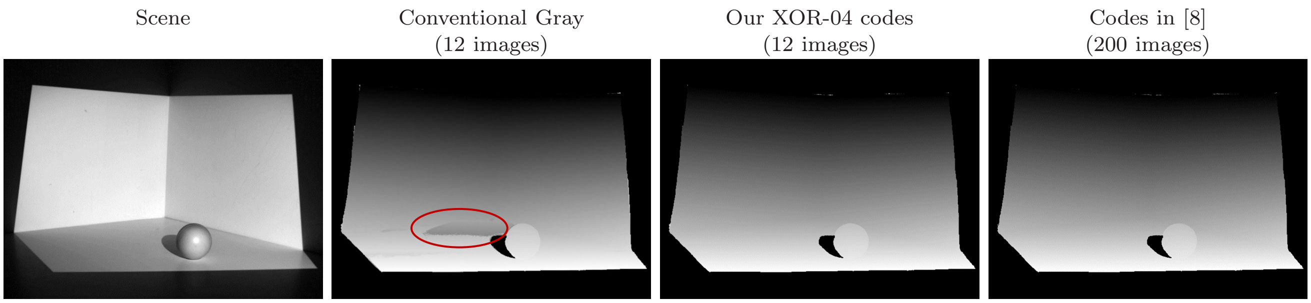


(a) Deep wax container (b) Code ensemble

Fig. 21 Deep wax container (failure case): Points inside the container receive both strong interreflections and strong subsurface scattering. Since none of the four codes compute the correct shape, the code ensemble fails to reconstruct the object.

7 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. An example is shown in Figure 23. This object is a concave lamp made of brushed metal. This is a challenging object for active shape recovery systems due to strong, high-frequency interreflections. Figure 24 (e) shows the reconstruction results using our code ensemble. While the reconstruction is better as compared to individual codes, a significant amount of errors remain.

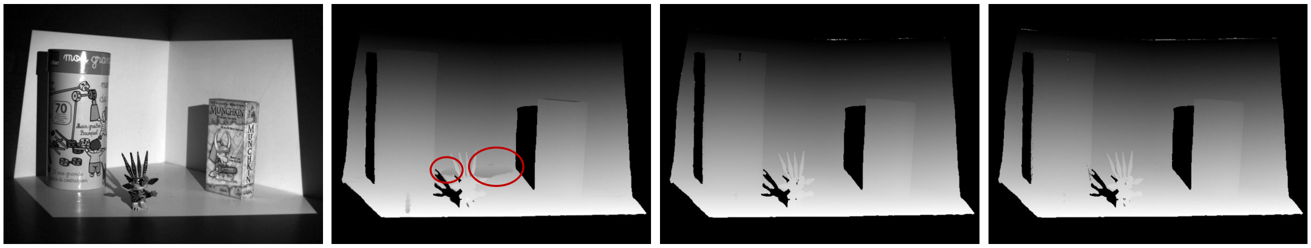

Scene

Conventional Gray (12 images)

Our XOR-04 codes (12 images)

Codes in [8] (200 images)

‘Ball’ Scene


‘Games’ Scene

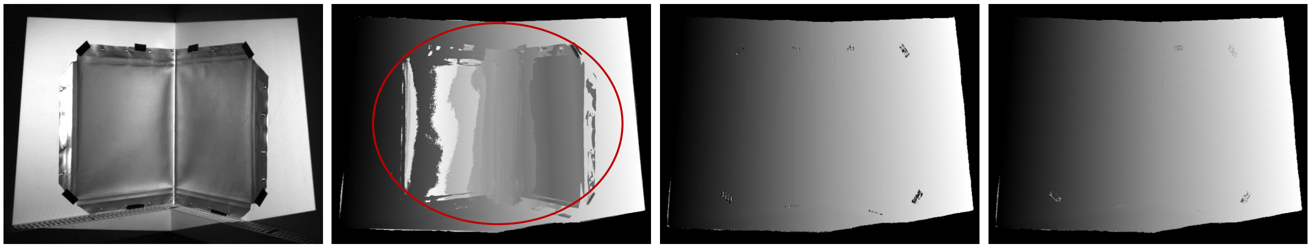

‘Corner’ Scene

Fig. 22 Comparisons with Couture *et al.* [8]: First column shows scenes with interreflections. Second column shows shape recovered with conventional Gray codes. Since all the scenes have strong interreflections, conventional Gray codes result in large errors (marked in red). The random high-frequency codes [8] successfully remove the errors, as shown in the third column. The XOR-04 codes produce results of the same accuracy (fourth column), while requiring an order of magnitude fewer images (12 versus 200).

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.* [24]. 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.

7.1 Error detection

Our error detection algorithm is based on a simple observation. 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, as illustrated in Figure 24 (f). It is possible that one of the four values might be the correct value. However, as there is an error correction stage to follow, 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.

Park *et al.* [31, 30] use similar consistency checks across range scans acquired from different view points. By registering different scans and comparing the values from different scans, they remove spurious measurements due to specular interreflections. In contrast, our technique does not require moving the acquisition setup or the object.

7.2 Error correction

To correct the errors, we iteratively collect additional images while illuminating only the scene points corre-

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 for the two Gray codes using conventional decoding and the two logical codes using the logical decoding (Section 4.1).
3. Apply a median filter (e.g., 3×3 or 5×5) to the individual depth values to prevent propagation of aliasing errors.
4. Compare the depth values. If any two codes are consistent, return that value as the correct depth. If the two Gray codes are consistent, return the value computed by the maximum min-SW Gray codes (Section 5).
5. **Error detection:** Mark the camera pixels where no two codes agree as error pixels (Section 7). An example is shows in Figure 24.
6. **Error correction:** Mask the patterns so that only the scene points corresponding to the error pixels are lit [37]. Repeat steps $1-5$ to progressively reduce the residual errors (Section 7, Figure 24).

(a) Concave Metal Lamp

Fig. 23 Concave metal lamp: A highly challenging object due to strong, high-frequency interreflections.

sponding to the error pixels. This technique, based on the work of Xu *et al.* [37], progressively reduces the amount of global illumination, resulting in reduction of the error pixels. In the subsequent iterations, the scene points which are already decoded correctly are not illuminated. This is achieved using illumination masks, as shown in Figures 24 (g,h). By progressively reducing the number of points getting illuminated (and hence, interreflections), the residual errors are reduced. By acquiring images in 2 extra iterations [11], we achieve a nearly perfect reconstruction.

Conventional Gray codes can not reconstruct a large portion of the object. Separation based modulated phase-shifting [7] can not remove the high-frequency interreflections, resulting in large errors. 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 $= 250mm$), respectively. 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 first iteration itself, we require only a small number of extra iterations (typically 1-2) even for challenging scenes. In comparison, the approach presented in [37] 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 subsurface scattering), this technique may not converge.

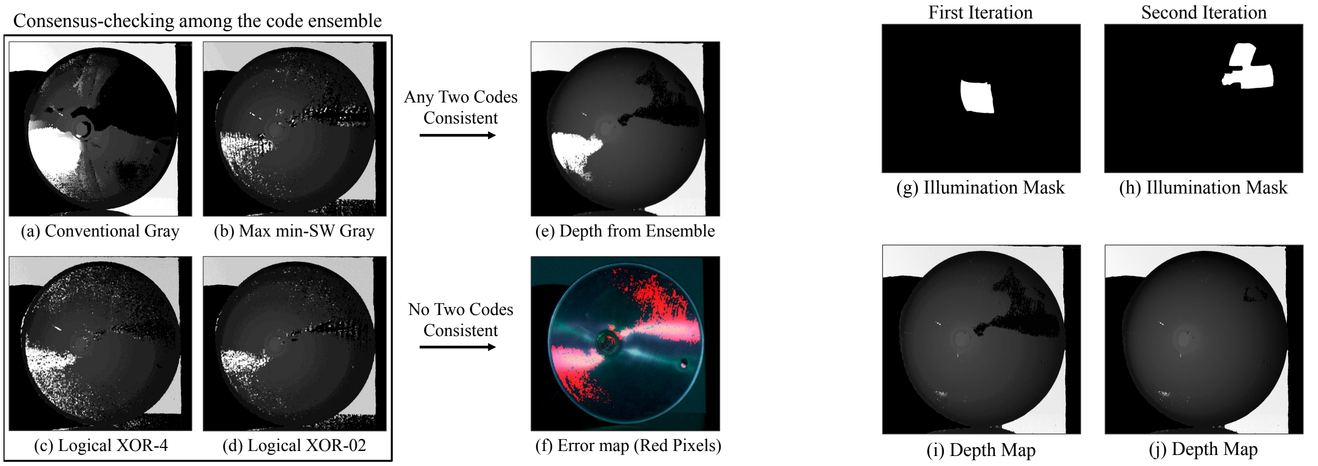
8 Discussion and Limitations

Frequencies of the projected patterns: Our methods make the following assumptions on the light transport in the scene. The high-frequency codes assume that light transport is locally smooth. The low-frequency codes assume that the light transport is local. If both these conditions are violated simultaneously, our techniques will produce incorrect results. For example, if a scene has mirror interreflections, or if the extent of subsurface scattering is significantly larger than the minimum stripe width of max min-SW codes, our techniques might fail to reconstruct the scene accurately.

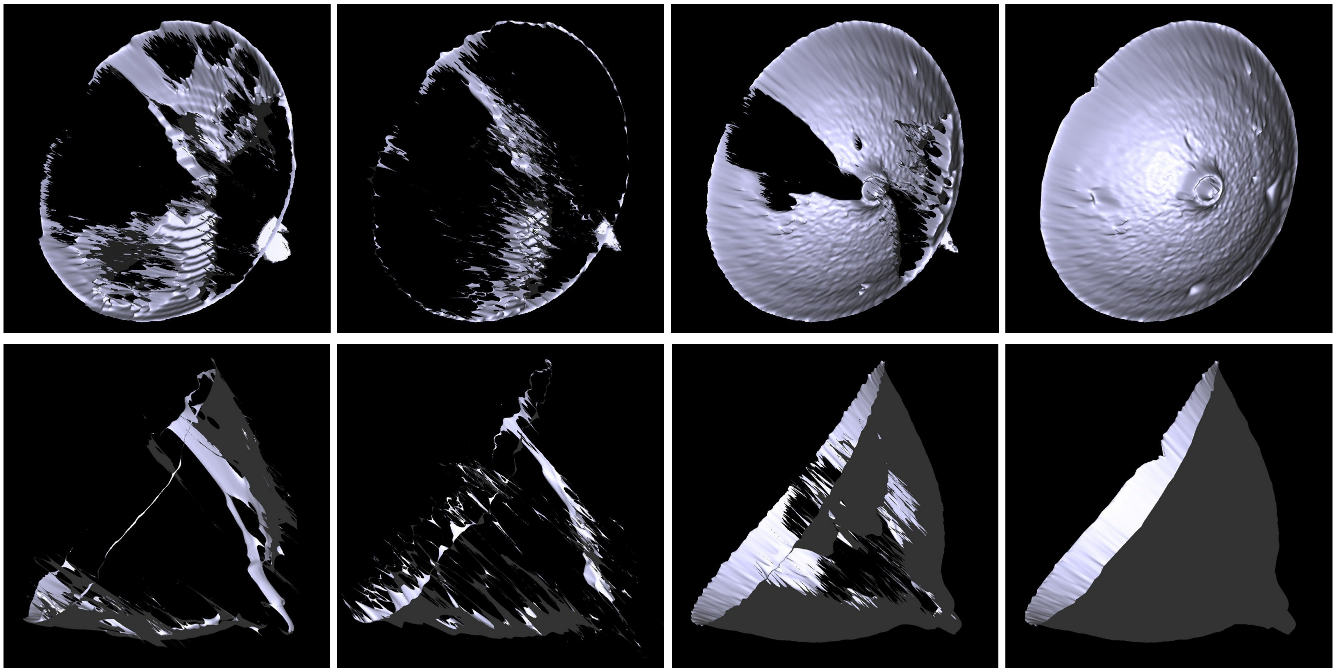
This limitation is because we have classified the light transport effects two way into long and short range. Patterns with a continuous set of frequencies can be used for handling a larger class of scenes. For example, it is possible to construct different band-pass codes by doing the XOR operations. Instead of only XOR-02 and XOR-04 codes, depending on the scene, XOR-08, XOR-16, XOR-32 codes can be used. Alternatively, sinusoidal patterns can be used as they provide more flexibility in controlling spatial frequencies. Ultimately, there is a trade-off between acquisition speed and the range of scenes that can be handled. Four sets of patterns with extreme frequencies can be considered to be the *minimal set*.

---

[11] We projected only the logical codes in subsequent iterations, thus requiring 122 images in total.



Error detection using consistency check **Error correction**



(k) Conventional Gray codes (11 images) (l) Modulated phase shifting (162 images) (m) Our code ensemble (42 images) (n) After error detection and correction (122 images)

Fig. 24 Error Detection and Correction: (a) We use the same consistency check as in the code ensemble algorithm for detecting errors. (a-d) Four depth maps using the individual codes. (e) Depth using the code ensemble algorithm has a significant amount of residual errors. (f) For a pixel, if no two codes agree on a depth value, it is marked as an error pixel (red). Since no extra patterns are projected, the error detection stage places no overhead in terms of acquisition time. In the subsequent iterations, the scene points which are already decoded correctly are not illuminated. This is achieved using an illumination masks (g,h). By progressively reducing the number of points getting illuminated (and hence, interreflections), the residual errors are reduced [37] (i,j). This object is very hard to reconstruct with existing schemes (k,l). Using our techniques, we achieve a high quality reconstruction (n). The mean errors for our result (n), conventional Gray codes (k) and modulated PS (l) are $1.2mm$, $29.8mm$ and $43.9mm$ respectively (height of lamp = 250mm). The parentheses contain number of input images.

What are the good spatial frequencies to use? Answering this requires a more thorough theoretical analysis of the frequencies of light transport. While such an analysis is hard for general scenes, we believe that studying the statistics for light transport for natural scenes will provide useful insights. This forms a promising direction of future research.

Single dominant mode of light transport: Our techniques assume a single dominant mode of light transport for every scene point. If a scene point receives both strong short-range and long-range effects, for example, inside of a strongly translucent and deep bowl, none of the codes will produce the correct result. An example is shown in Figure 20. In this case, the code ensemble algorithm and the error correction step will not be able to retrieve the correct result. Our techniques can not handle scenes in the presence of participating media as volumetric scattering also results in both short-range and long-range interactions.

Qualitative classification of light transport: The qualitative classification of light transport shown in Figure 8 is specific to the projector camera configuration. So far, we haven't reached a stage where this classification can provide reliable quantitative information about the scene. For example, most of the points inside the bowl receive interreflections. But since this classification is based on the errors that the code ensemble algorithm makes, only a few points are classified as receiving interreflections.

Conventional Gray codes as short-range codes: In our code ensemble, we have considered conventional Gray codes as being resistant to short-range effects. This is an approximation. Due to local effects, the higher frequency images in the conventional Gray codes will get blurred, and might not be decoded correctly. However, since the high-frequency patterns correspond to the lower significance bits, the resulting errors are small (e.g., < 4 pixels if the last two patterns are lost). Hence, in the consistency check step, the result of conventional Gray codes will still agree with that of the max min-SW Gray codes. In this case, the value computed by the minimum min-SW codes is returned. A future research direction is to design more codes with large minimum stripe widths.

Acquisition speed: Our techniques are currently limited to binary codes (monochrome and color) and thus require capturing several tens of images, making them unsuitable for dynamic scenes. The number of input images can be decreased by having more than two intensity levels in the projected images. An interesting direction of future work is to extend our techniques to N-ary ($N > 2$) codes and continuous schemes, such as phase shifting, which require fewer images as compared to discrete binary patterns. The number of images can also be reduced using a priori knowledge about the scene. For example, if the scene is known to have only interreflections, then it is sufficient to use only the logical codes, e.g., XOR-04. If, however, no a priori knowledge about the scene is available, then the code ensemble should be used.

Acknowledgements. This research was conducted while Mohit Gupta was an intern at MERL. We thank Jay Thornton, Joseph Katz, John Barnwell and Haruhisa Okuda (Mitsubishi Electric Japan) for their help and support. Mohit Gupta was partially supported by ONR grant N00014-11-1-0295. Prof. Srinivasa Narasimhan was partially supported by NSF grants IIS-0964562 and CAREER IIS-0643628 and a Samsung SAIT GRO grant. The authors thank Vincent Chapdelaine-Couture for sharing their data-sets.

References

1. Webpage. *http://graphics.cs.cmu.edu/projects/StructuredLight3DScanning/*.
2. D. G. Aliaga and Y. Xu. Photogeometric structured light: A self-calibrating and multi-viewpoint framework for accurate 3D modeling. In *CVPR*, 2008.
3. B. Atcheson, I. Ihrke, W. Heidrich, A. Tevs, D. Bradley, M. Magnor, and H. Seidel. Time-resolved 3D capture of non-stationary gas flows. *ACM Trans. Graph.*, 27(3), 2008.
4. D. Caspi, N. Kiryati, and J. Shamir. Range imaging with adaptive color structured light. *IEEE PAMI*, 20, 1998.
5. M. K. Chandraker, F. Kahl, and D. J. Kriegman. Reflections on the generalized bas-relief ambiguity. In *CVPR*, 2005.
6. 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.
7. T. Chen, H.-P. Seidel, and H. P. A. Lensch. Modulated phase-shifting for 3D scanning. In *CVPR*, 2008.
8. V. Couture, N. Martin, and S. Roy. Unstructured light scanning to overcome interreflections. In *ICCV*, 2011.
9. M. C. Er. On generating the n-ary reflected gray codes. *IEEE Transactions on Computers*, C-33(8), 1984.
10. 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.
11. L. Goddyn and P. Gvozdjak. Binary gray codes with long bit runs. *The electronic journal of combinatorics*, 2003.
12. 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.
13. J. Gu, T. Kobayashi, M. Gupta, and S. K. Nayar. Multiplexed illumination for scene recovery in the presence of global illumination. In *ICCV*, 2011.
14. J. Gu, S. K. Nayar, E. Grinspun, P. N. Belhumeur, and R. Ramamoorthi. Compressive structured light for recovering inhomogeneous participating media. In *ECCV*, 2008.
15. M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan. Structured light 3d scanning in the presence of global illumination. In *CVPR*, 2011.
16. M. Gupta, S. G. Narasimhan, and Y. Y. Schechner. On controlling light transport in poor visibility environments. In *CVPR*, 2008.
17. M. Gupta, Y. Tian, S. G. Narasimhan, and L. Zhang. (De) Focusing on global light transport for active scene recovery. In *CVPR*, 2009.

18. C. Hermans, Y. Francken, T. Cuypers, and P.Bekaert. Depth from sliding projections. In *CVPR*, 2009.
19. 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.
20. E. Horn and N. Kiryati. Toward optimal structured light patterns. *Proc. International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, 1997.
21. I. Ihrke, K. N. Kutulakos, H. P. A. Lensch, M. Magnor, and W. Heidrich. State of the art in transparent and specular object reconstruction. *STAR Proceedings of Eurographics*, 2008.
22. D. Kim, M. Ryu, and S. Lee. Antipodal gray codes for structured light. In *ICRA*, 2008.
23. S. Liu, T. T. Ng, and Y. Matsushita. Shape from second-bounce of light transport. In *ECCV*, 2010.
24. 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.
25. N. J. W. Morris and K. N. Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter trace photography. In *ICCV*, 2007.
26. S. G. Narasimhan, S. K. Nayar, B. Sun, and S. J. Koppal. Structured light in scattering media. In *ICCV*, 2005.
27. S. K. Nayar, K. Ikeuchi, and T. Kanade. Shape from Interreflections. *IJCV*, 6(3), 1991.
28. 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.
29. D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM Trans. Graph.*, 24(3), 2005.
30. J. Park and A. C. Kak. Multi-peak range imaging for accurate 3D reconstruction of specular objects. In *ACCV*, 2004.
31. J. Park and A. C. Kak. 3D modeling of optically challenging objects. *IEEE Transactions on Visualization and Computer Graphics*, 14(2), 2008.
32. J. Posdamer and M. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1), 1982.
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, 2010.
34. E. Steger and K. N. Kutulakos. A theory of refractive and specular 3D shape by light-path triangulation. *IJCV*, 76(1), 2008.
35. M. Trobina. Error model of a coded-light range sensor. *Technical Report*, 1995.
36. P. M. Will and K. S. Pennington. Grid coding: A preprocessing technique for robot and machine vision. *Artificial Intelligence*, 2(3-4), 1971.
37. Y. Xu and D. Aliaga. An adaptive correspondence algorithm for modeling scenes with strong interreflections. *IEEE TVCG*, 2009.
38. L. Zhang and S. K. Nayar. Projection Defocus Analysis for Scene Capture and Image Display. *ACM Trans. Graph.*, 25(3), 2006.
39. S. Zhang. High-resolution, three-dimensional shape measurement. *Ph.D. Thesis, Stony Brook University*, 2005.
40. S. Zhang, D. V. D. Weide, and J. Oliver. Superfast phase-shifting method for 3-D shape measurement. *Optics Express*, 18(9), 2010.