Depth Acquisition from Density Modulated Binary Patterns

Zhe Yang *
Northeastern Univ.
yangzhe.research@gmail.com

Zhiwei Xiong
Microsoft Research Asia
zhxiong@microsoft.com

Yueyi Zhang
Univ. of Sci. & Tech. of China
zhyuey@mail.ustc.edu.cn

Jiao Wang
Northeastern Univ.
wangjiao@ise.neu.edu.cn

Feng Wu
Microsoft Research Asia
fengwu@microsoft.com

Abstract

This paper proposes novel density modulated binary patterns for depth acquisition. Similar to Kinect, the illumination patterns do not need a projector for generation and can be emitted by infrared lasers and diffraction gratings. Our key idea is to use the density of light spots in the patterns to carry phase information. Two technical problems are addressed here. First, we propose an algorithm to design the patterns to carry more phase information without compromising the depth reconstruction from a single captured image as with Kinect. Second, since the carried phase is not strictly sinusoidal, the depth reconstructed from the phase contains a systematic error. We further propose a pixel-based phase matching algorithm to reduce the error. Experimental results show that the depth quality can be greatly improved using the phase carried by the density of light spots. Furthermore, our scheme can achieve 20 fps depth reconstruction with GPU assistance.

1. Introduction

Microsoft Kinect is a structured-light-based depth camera [8, 14]. By replacing the projector with an infrared laser and a diffraction grating, Kinect becomes cheap and convenient for millions of users to use in many applications [12, 24, 29, 3, 22]. As shown in Figure 1(a), the simulated binary pattern used in Kinect consists of random light spots. The position of light spots in every small region is unique. So depth can be reconstructed by establishing the correspondences between the reference and captured images.

However, depth reconstructed in this way suffers from holes and severe noise caused by the pattern to some extent. First, the position of light spots has to be identified by a block of pixels. When a block is projected onto the boundary of an object, it will be so deformed that the correspondences in the region are hard to find. Second, the pattern in Figure 1(a) is binary and thus is difficult to be interpolated to a higher resolution. The found correspondences have a limited accuracy, causing the Kinect depth to have random errors from a few millimeters to 4 cm [14]. It would be desirable to improve the quality of the Kinect depth.

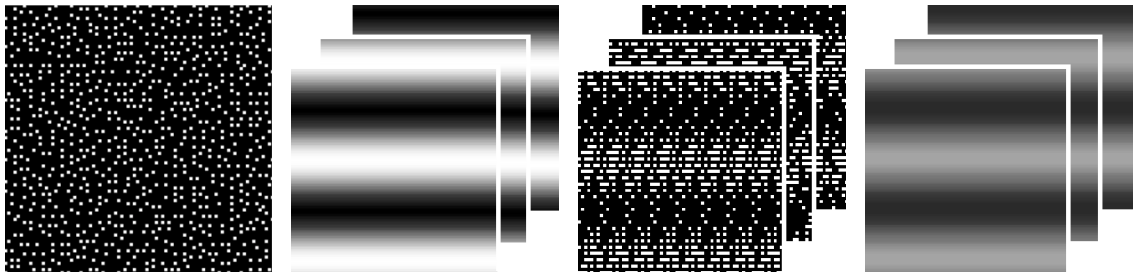
Phase shifting, which projects a series of phase-shifted sinusoidal patterns [23], achieves better quality. Depth can be reconstructed at every camera pixel with a set of captured images. Thus the depth has higher spatial resolution. Furthermore, the depth is calculated from sinusoidal phase differences. So noise is suppressed and the depth is more accurate. However, as shown in Figure 1(b), the phase shifting patterns are grayscale and hard to generate using infrared lasers and diffraction gratings like Kinect.

In this paper, we propose a novel approach to embed phase information into binary patterns that can still be generated with infrared lasers and diffraction gratings. Similar to image dithering, our idea is to use the density of light spots to represent phase, as shown in Figure 1(c). Figure 1(d) shows the energy images averaged in a sliding window from the patterns in Figure 1(c), which have similar properties to the phase shifting patterns in Figure 1(b). The immediate advantage is that the depth quality can be improved with extracted phase information.

Our technical contributions are twofold in this paper. First, an algorithm is proposed to design three density modulated binary patterns. The goal is to carry more phase information without compromising the depth reconstructed from a single captured image as with Kinect. Second, since the carried phase is not strictly sinusoidal, the depth reconstructed from the phase contains a systematic error. A pixel-based phase matching algorithm is further proposed to reduce the error. Finally, the depth data reconstructed by the position of light spots in one captured image and by the phase carried in three captured images are adaptively inte-

*This work was done during Z. Yang's internship at MSRA.

1063-6919/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPR.2013.11



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

Figure 1. (a) Simulated binary pattern used in Kinect. (b) Sinusoidal fringe patterns used in phase shifting. (c) The proposed density modulated binary patterns. (d) The energy images generated from (c) using a sliding window.

grated toward better quality.

Before this paper, Zhang proposed generating phase-shifting patterns from binary patterns by projector defocusing [31]. However, this method still requires a projector for defocusing and cannot use a single captured image to reconstruct depth. In our scheme, every captured image consists of light spots and thus depth can be reconstructed just like Kinect. Meanwhile, experimental results show that the depth quality can be greatly improved using the phase carried by the density of light spots.

This paper is organized as follows. Section 2 gives an overview of structured light and phase shifting. Section 3 proposes the pattern generation algorithm. Section 4 proposes the depth reconstruction algorithm. Experimental results and comparisons are presented in Section 5. Finally, Section 6 concludes the paper.

2. Related Work

2.1. Structured Light

Triangulation-based structured light (SL) schemes are similar to passive stereo, except that one of the cameras is replaced with a projector [2]. In the early stages, SL schemes mainly use binary coded or Gray coded patterns [11, 19]. They require multiple patterns and the scene cannot have motion when the patterns are projected. SL schemes have made solid progress in terms of both pattern design and pattern emitting in recent decades.

Currently, most of the SL schemes target using one pattern. Salvi et al. proposed a cross-point pattern [20]. Albitar et al. proposed a disc, circle, and stripe pattern [1]. Maurice et al. further proposed using the Hamming distance to drive the pattern generation [17]. By taking geometric deformation into account, Kawasaki et al. proposed a simple grid pattern formed by a number of straight lines [13]. Koninckx et al. proposed a coded line pattern adaptively generated according to epipolar geometry [15]. In addition, color information was also considered in the pattern design. Fong et al. proposed combining a set of parallel color stripes and a perpendicular set of sinusoidal intensity stripes [7]. Chen et al. proposed a tessellated pattern with seven colors [4].

For pattern emitting, although some patterns can be emitted by a laser, all the above schemes use a projector. Kinect features an infrared laser that can generate and emit a constant pattern with light spots [8, 14], which makes a depth camera available as a consumer-grade device. However, the Kinect depth still suffers from noise and holes. By designing new patterns with the phase information embedded, our scheme greatly improves the depth quality.

2.2. Phase Shifting

Phase shifting is a special SL scheme that emits a series of phase shifted sinusoidal patterns. Increasing the number of stripes in the patterns can improve the measurement accuracy, but high frequency results in ambiguities that requires phase unwrapping. Weise et al. used stereo to overcome the problem but an additional camera is required [27]. Wang et al. proposed embedding a period cue into the phase shifting patterns without reducing the signal-to-noise ratio [26]. As a result, each period of the sinusoidal patterns can be identified. Wissmann et al. used a similar idea but embedded the signal during pattern emitting [28].

Global illumination and defocusing are practical factors that often introduce errors into depth measurement. Nayar et al. showed that high-frequency sinusoidal patterns can be used to separate global illumination [18]. Chen et al. proposed multiplying a high-frequency signal with other signals [6]. Gu et al. derived a lower bound for the number of required patterns and proposed a multiplexed illumination scheme to achieve it [9]. Gupta et al. constrained the frequencies of sinusoidal patterns to a narrow high-frequency band, which greatly reduces global illumination and defocusing [10]. Chen et al. proposed using polarization-difference imaging to filter out subsurface scattering [5].

For phase shifting, when a set of patterns is emitted, the scene is assumed to be static. This assumption is not true in many practical scenarios. Koninckx et al. proposed estimating lines from the projected sinusoidal patterns and

calculating motion as line translation [16]. However, this method is not suitable for high-speed acquisition. Zhang et al. reduced the motion effects by setting the last pattern as flat [32]. Weise et al. compensated for the motion by introducing the first-order Taylor term in phase shifting [27]. Recently, several approaches are proposed to combine single-shot and multi-shot SL together for scalable depth sensing. Taguchi et al. designed spatio-temporal decodable patterns [25] and Zhang et al. used hybrid patterns of random speckle and sinusoidal fringe [33].

Our key contribution in this paper is designing the phase shifting patterns that can be emitted by infrared laser, where the phase information is approximated by the density of light spots in a local region. In our scheme, the phase ambiguity problem is solved naturally by the random position of light spots. When the scene contains moving objects, depth in the corresponding regions can be still reconstructed from a single captured image as with Kinect.

3. Density Modulated Binary Patterns

As shown in Figure 1(c), the three proposed patterns are binary and can thus be generated using infrared lasers and diffraction gratings. In this section, we discuss how to modulate the density of light spots to represent phase.

3.1. Pattern Generation

Let us define a pattern as $P(r, c)$, row $r = 1, ..., R$ and column $c = 1, ..., C$. In Kinect, the light spots are randomly and uniformly distributed in $P(r, c)$. Contrastively, in the proposed patterns, the number of light spots in different rows is defined as a sinusoidal function

$$k(r) = Round\{[\sin(2\pi\frac{r}{T} + \theta) + 1] \times \alpha + 1\}, \quad (1)$$

where $Round()$ is a function to round a floating number to an integer, r is the row index, T is the number of rows in a sinusoidal period, and α is a scaling factor to control $k(r)$ as an integer from 1 to K that actually determines the number of different densities in a period. The three patterns are generated by setting θ as $-2\pi/3, 0$, and $2\pi/3$, respectively.

In Figure 1(b), an important property is that pixels in the same row have the same intensity. To achieve this, we propose a pattern generation algorithm (Algorithm 1). Let us define $1 \times N$ pixels as a basic unit for pattern generation, where N is larger than the maximum $k(r)$. The position of light spots is random in a basic unit but the same for all basic units in the same row. This makes sure that every sliding window located in the same row has the same average intensity. At the same time, since the number of light spots and their positions are different in different rows, the position of light spots in every block is still unique.

In the proposed algorithm, there are two parameters that must be determined, namely, the scaling factor α and the period T. The former is decided by N and the range of $k(r)$. N is an empirical parameter. The larger N is, the greater number of different values $k(r)$ can have. If N is too large, when $k(r)$ is a small integer, the density $k(r)/N$ will be too small to establish the correspondences. Thus we set N as 8 in this paper. When N is decided, the maximum $k(r)$ can be selected as large as 5 in order to have more densities in a period.

Algorithm 1 Pattern Generation

Require: The number of rows in one period T, the scaling factor α, and the initial phase θ

```
for r = 1, ..., R do
    Calculate k(r) according to Eq. (1)
    Divide a row into M non-overlapping basic units
    for m = 1, ..., M do
        Randomly select k(r) positions from 1 to N
        Let pixels at selected positions be light spots
    end for
end for
```

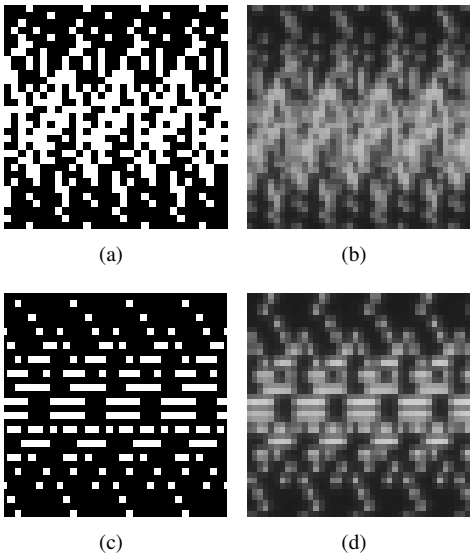


Figure 2. Two patterns with $N = 8$ and $k(r) \in \{1, ..., 5\}$. (a) The pattern with light spots in every row. (b) The captured image corresponding to (a). (c) The pattern with light spots in every other row. (d) The captured image corresponding to (c).

In Figure 2, we test two different approaches to generate patterns with $N = 8$ and $k(r) \in \{1, ..., 5\}$. In (a), every row in the pattern contains a certain number of light spots as calculated by Eq. (1). In (c), every other row contains the light spots while the remaining rows are black. From the captured images in (b) and (d), it is difficult to keep the position of light spots clear if using the pattern in (a), since the captured image may contain too many light spots. Therefore, the pattern in (c) is adopted in this paper.

As mentioned above, the smaller the period T is, the more accurate the depth measurement will be. However, in the proposed scheme, T also has a minimum value. For $N = 8$ and $k(r) \in \{1, ..., 5\}$, if T is smaller than 32, not every density $k(r)$ appears because of the rounding. In other words, the range of $k(r)$ is not fully used. Therefore, in this paper we set $T = 32$.

Finally, we briefly explain how to produce the diffraction gratings with the proposed patterns. They are produced by photolithography, a process similar to that used in producing integrated circuits (ICs). The grating consists of a transparent substrate and an opaque film. The film is first covered by the chemical photoresist and then the designed pattern is projected to the photoresist. Exposed photoresist and uncovered opaque materials can be removed. The rest is the transparent substrate.

3.2. Implicit Phase

Next we evaluate whether the densities of generated patterns can approximate the sinusoidal fringe well. The approximation is conducted by calculating the average energy in a $N \times N$ sliding window. For simplicity, we consider the ideal case, namely, the signal $k(r)$ is continuous. The energy can be calculated as

$$E = \frac{1}{N^2} \int_{r-N/2}^{r+N/2} k(r) dr. \tag{2}$$

Substituting Eq. (1) with ignoring the rounding and the constant term, we get

$$\begin{aligned} E &= \frac{1}{N^2} \int_{r-N/2}^{r+N/2} \alpha \sin(2\pi \frac{r}{T} + \theta) dr \\ &= \beta \sin(\pi \frac{N}{T}) \sin(2\pi \frac{r}{T} + \theta), \end{aligned} \tag{3}$$

where $\beta = \frac{\alpha T}{\pi N^2}$ and $sin(\pi \frac{N}{T})$ are constant. Thus the energy E in a sliding window in the proposed patterns is a sinusoidal function mathematically.

However, since the proposed patterns are binary, $k(r)$ in Eq. (1) has to be rounded to an integer. What is the effect of this rounding when approximating sinusoidal fringes? Figure 3 shows the actual energy calculated from the three proposed patterns. We can observe that the overall shape of the energy is still a good approximation. However, there are obvious stair-wise errors. If these errors are not handled carefully during depth reconstruction, a systematic error will be introduced to the reconstructed data.

4. Depth Reconstruction

In our scheme, depth can be reconstructed in two ways. First, it can be reconstructed from every captured image by block matching. Second, it can be reconstructed by phase shifting from three energy images. These two kinds of depth data can be further integrated toward better quality.

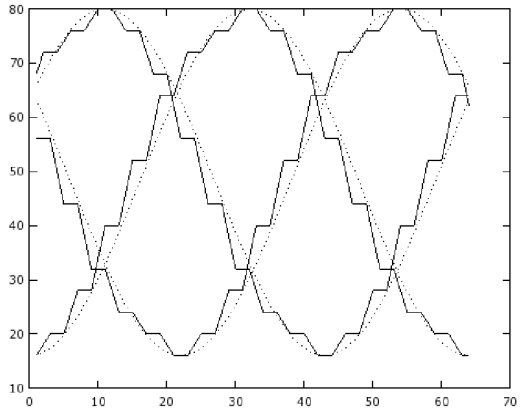


Figure 3. Approximation to sinusoidal fringes with $N = 8$, $T = 32$, and $k \in \{1, ..., 5\}$. Solid curves denote the energy in a sliding window, and dotted curves denote ideal sinusoidal fringes.

4.1. Block Matching

Figure 4(a) shows one captured image with the emitted pattern. Since the position of light spots in every small region is still unique, we can use a block matching method to get the disparity of every pixel between the reference image $\bar{I}(r, c)$ and the captured image $I(r, c)$. The reference image $\bar{I}(r, c)$ is captured at a reference plane with a known, constant depth. Similar to [21], we adopt normalized cross-correlation (NCC) between two blocks as the matching criterion. The disparity at (r, c) is calculated as

$$D_1(r,c) = \arg\max_{r',c'} \frac{\sum_{i,j} A(r,c,i,j) B(r',c',i,j)}{\sqrt{\sum_{i,j} A^2(r,c,i,j) \sum_{i,j} B^2(r',c',i,j)}}, \tag{4}$$

where $A(r, c, i, j) = I(r+i, c+j) - I_0$ and $B(r', c', i, j) = \bar{I}(r' + i, c' + j) - \bar{I}_0$. I_0 and $\bar{I}_0$ are the average intensity of I and $\bar{I}$ in a block, respectively. $D_1(r, c)$ is the row and column disparity with the largest NCC.

Every pixel belongs to multiple overlapping blocks for calculating NCC and thus has multiple disparities calculated from Eq. (4). Similar to the method used in Kinect [30], we adopt a winner-takes-all strategy. The disparity of each pixel is finally decided by the block with the largest NCC. Figure 4(b) shows the reconstructed depth from Figure 4(a). We can clearly observe the discontinuities on the smooth surface.

4.2. Pixel-based Phase Matching

For every captured image, we can calculate an energy image from the discrete version of Eq. (2). Figure 4(c) is the generated energy image from Figure 4(a). Let us define the energy images of the reference and captured images as

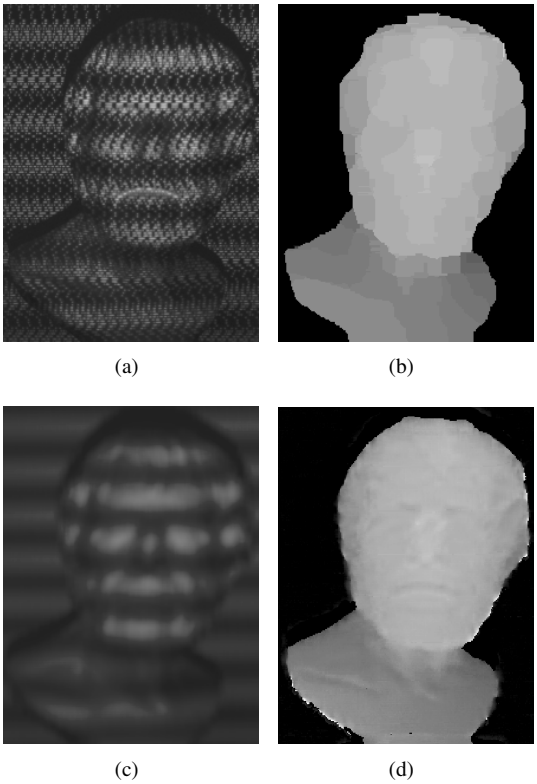


Figure 4. (a) One captured image with the emitted pattern. (b) Depth reconstructed from (a) by block matching. (c) The energy image generated from (a). (d) Depth reconstructed from three energy images by pixel-based phase matching.

$\bar{E}_i(r,c)$ and $E_i(r,c)$, where $i = 1, 2, 3$ denote different $\theta's$. Similar to original phase shifting [23], the three energy images can be represented as

$$\begin{aligned} E_1(r,c) &= E'(r,c) + E''(r,c)\cos[\phi(r,c) - 2\pi/3], \\ E_2(r,c) &= E'(r,c) + E''(r,c)\cos[\phi(r,c)], \\ E_3(r,c) &= E'(r,c) + E''(r,c)\cos[\phi(r,c) + 2\pi/3]. \end{aligned} \quad (5)$$

$E'(r,c)$ is the background intensity, $E''(r,c)$ is the sinusoidal amplitude, and $\phi(r,c)$ is the phase to be solved. From the three equations in Eq. (5), $\phi(r,c)$ can be obtained as

$$\phi(r,c) = \arctan\{\frac{\sqrt{3}[E_1(r,c) - E_3(r,c)]}{2E_2(r,c) - E_1(r,c) - E_3(r,c)}\}. \quad (6)$$

Since the proposed patterns contain a stair-wise error for approximating sinusoidal fringes as shown in Figure 3, we cannot calculate the depth directly from $\phi(r,c)$. Instead, we propose a pixel-based phase matching method. First, we calculate $\bar{\phi}(r,c)$ from the reference energy images $\bar{E}_i(r,c)$. For every phase $\phi(r,c)$, we then search for the most matched $\bar{\phi}(r - \Delta r, c)$ within one period. The disparity in one period is calculated as

$$d_2(r,c) = \begin{cases} \Delta r + \frac{\bar{\phi}(r-\Delta r,c) - \phi(r,c)}{\bar{\phi}(r-\Delta r,c) - \bar{\phi}(r-\Delta r-1,c)}, & \phi(r,c) \le \bar{\phi}(r-\Delta r,c), \\ \Delta r - \frac{\phi(r,c) - \bar{\phi}(r-\Delta r,c)}{\bar{\phi}(r-\Delta r+1,c) - \bar{\phi}(r-\Delta r,c)}, & \phi(r,c) > \bar{\phi}(r-\Delta r,c). \end{cases} \quad (7)$$

Δr is the integer row disparity, and the other term in Eq. (7) is the fractional row disparity obtained by linear interpolation. Here we assume the phase varies linearly at the sub-pixel level. Also, for simplicity, the optics are vertically aligned so $d_2(r,c)$ only contains the row disparity.

The phase ambiguity problem still needs to be solved. If there are W periods in the captured images, we need to identify in which period $d_2(r,c)$ is. It is easy to be solved for the proposed patterns because the position of light spots is unique in every period. We can check the NCC values for $w = 1, ..., W$. Then w with the largest NCC is selected. Finally, the disparity can be obtained by

$$D_2(r,c) = d_2(r,c) + w \times L, \quad (8)$$

where L is the number of rows in one period in the captured images.

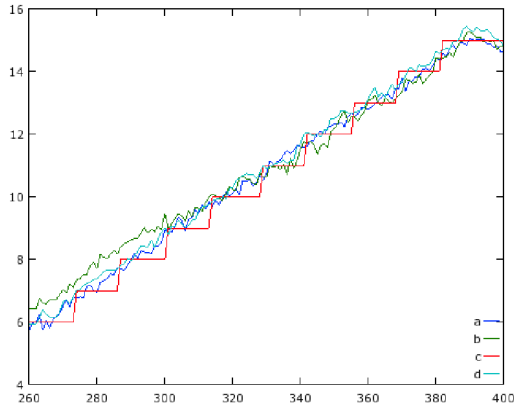


Figure 5. Accuracy of depth reconstruction. Curve "a" : depth from original phase shifting. Curve "b" : depth directly from $\phi(r,c)$. Curve "c" : depth by block matching. Curve "d" : depth by pixel-based phase matching. The horizontal axis denotes pixel index and the vertical axis denotes disparity in pixel.

Figure 5 evaluates the accuracy of depth reconstruction by different methods. The scene is a simple inclined plane. The blue curve "a" is the reconstruction from original phase shifting. It should have the best quality and is the anchor for evaluation. The red curve "c" is the reconstruction by block matching. It is stair-wise and consistent with the visual result in Figure 4(b). The green curve "b" is the reconstruction directly from $\phi(r,c)$. It has a considerable error on the left side. The light-blue curve "d" is the reconstruction by the proposed pixel-based phase matching algorithm. It is most similar to original phase shifting. The visual result is

shown in Figure 4(d), which looks much better than the one in Figure 4(b).

4.3. Depth Integration

Although using the embedded phase to reconstruct depth produces better quality, this method requires three captured images and is not good at handling moving objects in the scene. However, the depth reconstructed by block matching only needs one captured image, which is more suitable for moving objects. Therefore, the depth data $D_1(r, c)$ and $D_2(r, c)$ should be integrated according to the motion detected in the scene. We consider simple intensity change to detect if there is any motion in the scene, by comparing the current captured image with the previous third image because their patterns are the same. If the average intensity change in a region is larger than a threshold, we mark the region as moving. $D_1(r, c)$ is adopted in the moving regions and $D_2(r, c)$ in the remaining regions.

Finally, the implementation of our scheme is briefly discussed here. The block matching is carried out on a GeForce GTX 580 GPU. The phase calculation and pixel-based phase matching are carried out on an Intel Xeon E5440 CPU with 2.83 GHz working frequency. The depth integration is carried out on the same CPU too. The speed of depth reconstruction is 20 fps currently.

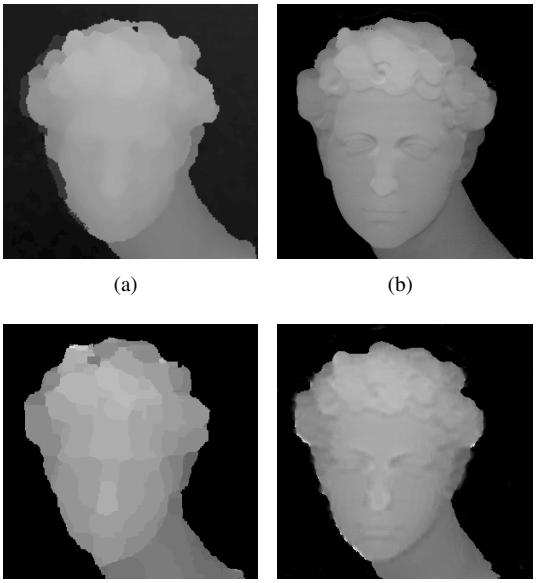
5. Experimental Results

In this section, we evaluate the depth reconstructed by different schemes. The first is Kinect, which we use as a black box. The second is original phase shifting, which uses three grayscale sinusoidal patterns emitted by a DLP projector and reconstructs depth directly from phase. The third is the proposed scheme, and the binary patterns are also emitted by the DLP projector for simulation.

5.1. Static Face

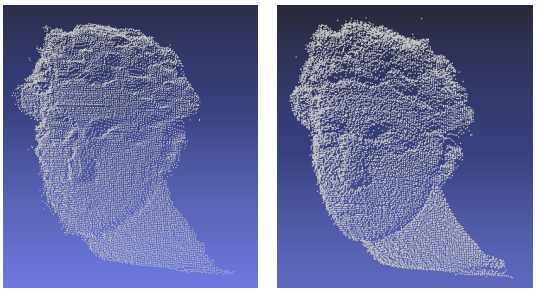
Figure 6(a) shows the depth from Kinect. It looks too smooth and some surface details are lost. In addition, part of the left cheek on the boundary is lost too. Figure 6(b) is the depth reconstructed from original phase shifting. The quality looks very good in terms of both surface details and object boundaries, which is the target of our scheme.

Figure 6(c) is the depth reconstructed by the proposed scheme but using block matching only. The result should be similar to that from Kinect but it actually looks worse because of the obvious stair-wise discontinuities. The main reason is that the camera used in Kinect is 1280×1024 and 11-bit [22] while our camera is only 640×480 and 8-bit. Figure 6(d) is the depth reconstructed by the proposed scheme using the embedded phase and pixel-based phase matching. It is much better than that from Kinect and also close to that from original phase shifting. So it is verified



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

Figure 6. Depth reconstruction of David. (a) Kinect. (b) Original phase shifting. (c) The proposed scheme using block matching. (d) The proposed scheme using pixel-based phase matching.



(a) (b)

Figure 7. Point cloud of David. (a) The proposed scheme. (b) Original phase shifting.

that carrying phase information in the binary patterns is a feasible way to improve the depth quality over Kinect.

We further compare the point cloud from the proposed scheme and original phase shifting, as shown in Figure 7. They look similar but we can still observe some unsatisfactory regions (e.g., the mouth) in our result. Continuously improving the quality deserves future efforts, yet there should be an inherent compromise as binary patterns cannot represent phase as perfectly as grayscale patterns.

5.2. Indoor Scenes with Multiple Objects

We also evaluate the above three schemes in two indoor scenes with multiple objects. The results are shown in Figure 8 and Figure 9. The depth from the proposed scheme is

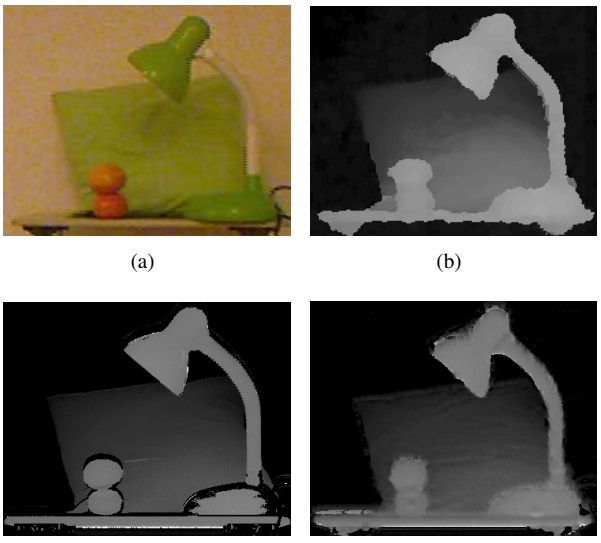


Figure 8. Scene with a pillow, a lamp and two oranges. (a) RGB image. (b) Kinect. (c) Original phase shifting. (d) Our scheme.

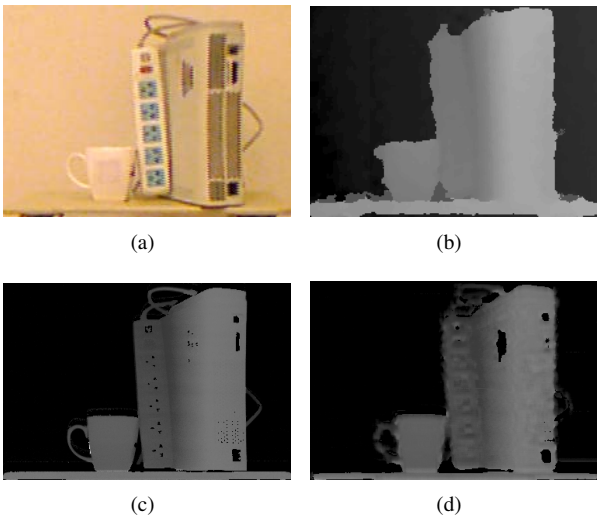


Figure 9. Scene with an Xbox, a cup and a receptacle. (a) RGB image. (b) Kinect. (c) Original phase shifting. (d) Our scheme.

always better than that from Kinect, while both of them emit binary patterns. Although our results look close to the results from original phase shifting using grayscale patterns, there are some noticeable differences. Since these scenes are static, they favor original phase shifting. However, our scheme can readily handle moving objects in the scene.

5.3. Scene with Moving and Static Objects

The test scene has two objects with the left one static and the right one moving forward. Figure 10 shows three results from the proposed scheme. (a) is the result obtained using block matching. The two objects have the same quality whether they are moving or not. (b) is the result obtained using pixel-based phase matching. The depth of the moving object is really poor because there is motion when the three patterns are emitted. Original phase shifting will suffer from the same problem. (c) is the result obtained by integrating the depth data of the above two algorithms, where both objects have the optimal depth.

5.4. Limitations

The proposed scheme focuses on indoor scenes. If more practical factors, such as global illumination and defocusing, are taken into account, the frequency of the proposed patterns should increase greatly. However, as we have discussed in Section 3.1, the period T has a minimum number of rows, which constrains us to increase the frequency. We have to find other ways to suppress the practical factors.

In our scheme, when the objects in the scene start to move, the depth reconstruction will degrade to block matching. Are there more intelligent ways to handle this problem? For example, can we use the depth from block matching to drive the previous obtained high quality depth? Or can we use the depth from block matching to compensate for the motion? This will definitely be something we will examine in the future research.

6. Conclusions

The proposed density modulated binary patterns carve out a feasible way to improve the depth quality over Kinect. By embedding phase into the binary patterns, it provides more information for depth acquisition. We propose the pattern generation algorithm and the pixel-based phase matching algorithm for reconstruction. Experimental results show that our scheme consistently outperforms Kinect for static scenes and original phase shifting for moving objects.

References

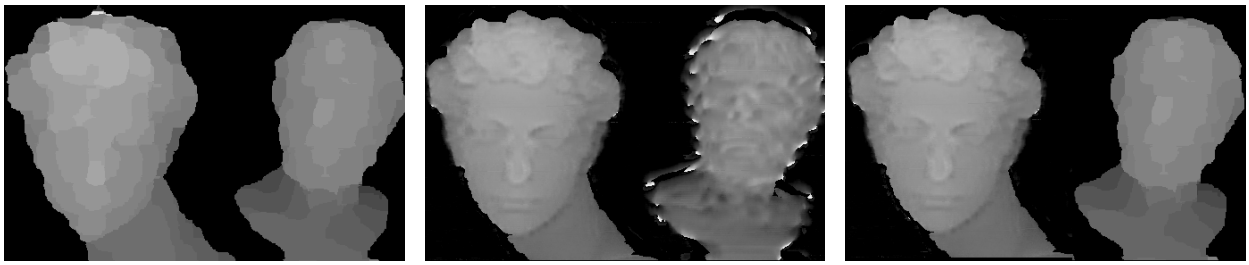
[1] C. Albitar, P. Graebling, and C. Doignon. Robust structured light coding for 3d reconstruction. In *ICCV*, 2007. 2

[2] F. Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1):231–240, 2004. 2

[3] J. Charles and M. Everingham. Learning shape models for monocular human pose estimation from microsoft xbox kinect. In *ICCV Workshops*, 2011. 1

[4] S. Y. Chen, Y. F. Li, and J. W. Zhang. Vision processing for realtime 3-d data acqusition based on coded structure light. *IEEE Transactions on Image Processing*, 17(2):167–176, 2008. 2

[5] T. Chen, H. Lensch, C. Fuchs, and H. P. Seidel. Polarization and phase-shifting for 3d scanning of translucent objects. In *CVPR*, 2007. 2



(a) (b) (c)

Figure 10. Scene with a static object (left) and a moving object (right). (a) The proposed scheme using block matching. (b) The proposed scheme using pixel-based phase matching. (c) The proposed scheme after depth integration.

[6] T. Chen, H. P. Seidel, and H. Lensch. Modulated phase-shifting for 3d scanning. In *CVPR*, 2008. 2

[7] P. Fong and F. Buron. Sensing deforming and moving objects with commercial off the shelf hardware. In *CVPR*, 2005. 2

[8] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli. Depth mapping using projected patterns. *US Patent Application 2010/0118123*, 2010. 1, 2

[9] J. Gu, T. Kobayashi, M. Gupta, and S. K. Nayar. Multiplexed illumination for scene recovery in the presence of global illumination. In *ICCV*, 2011. 2

[10] M. Gupta and S. K. Nayar. Micro phase shifting. In *CVPR*, 2012. 2

[11] S. Inokuchi, K. Sato, and F. Matsuda. Range-imageing system for 3-d object recognition. In *ICPR*, 1984. 2

[12] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *ACM Symposium on User Interface Software and Technology*, 2011. 1

[13] H. Kawasaki, R. Furukawa, R. Sagawa, and Y. Yagi. Dynamic scene shape reconstruction using a single structure light pattern. In *CVPR*, 2008. 2

[14] K. Khoshelham and S. O. Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, 2012. 1, 2

[15] T. P. Koninckx and L. V. Gool. Real-time range acquisition by adaptive structure light. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):432–445, 2006. 2

[16] T. P. Koninckx, P. Peers, P. Dutre, and L. V. Gool. Scene-adapted structure light. In *CVPR*, 2005. 3

[17] X. Maurice, P. Graebling, and C. Doignon. A pattern framework driven by the hamming distance for structured light-based reconstruction with a single image. In *CVPR*, 2011. 2

[18] 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. In *ACM SIGGRAPH*, 2006. 2

[19] P.Vuylsteke and A. Oosterlinck. Range image acqusition with a single binary-encoded light pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):148–164, 1990. 2

[20] J. Salvi, J. Pages, and J. Batlle. Pattern codifications strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 2004. 2

[21] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42, 2002. 4

[22] J. Smisek, M. Jancosek, and T. Pajdla. 3d with kinect. In *ICCV Workshops*, 2011. 1, 6

[23] V. Srinivasan, H. C. Liu, and M. Halioua. Automated phase-measuring profilometry of 3-d diffuse objects. *Applied Optics*, 23(18):3105–3108, 1984. 1, 5

[24] W. Susanto, M. Rohrbach, and B. Schiele. 3d object detection with multiple kinects. In *ECCV Workshops*, 2012. 1

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

[26] Y. Wang, K. Liu, Q. Hao, D. L. Lau, and L. G. Hassebrook. Period coded phase shifting strategy for real-time 3-d structured light illumination. *IEEE Transactions on Image Processing*, 20(11):3001–3013, 2011. 2

[27] T. Weise, B. Leibe, and L. van Gool. Fast 3d scanning with automatic motion compensation. In *CVPR*, 2007. 2, 3

[28] P. Wissmann, R. Schmitt, and F. Forster. Fast and accurate 3d scanning using coded phase shifting and high speed pattern projection. In *3DIMPVT*, 2011. 2

[29] L. Xia, C. C. Chen, and J. K. Aggarwal. Human detection using depth information by kinect. In *CVPR Workshops*, 2011. 1

[30] Z. Zalevsky, A. Shpunt, A. Maizels, and J. Garcia. Method and system for object reconstruction. *US Patent Application 2010/0177164*, 2010. 4

[31] S. Zhang. Flexible 3d shape measurement using projector defocusing: extended measurement range. *Optics letters*, 35(7):934–936, 2010. 2

[32] S. Zhang, D. Royer, and S. T. Yau. Gpu-assisted high-resolution, real-time 3-d shape measurement. *Optics Express*, 14(20):9120–9129, 2006. 3

[33] Y. Zhang, Z. Xiong, and F. Wu. Hybrid structured light for scalable depth sensing. In *ICIP*, 2012. 3