This paper proposes a novel phase-shifting method for fast, accurate, and unambiguous 3D shape measurement. The basic idea is embedding a speckle-like signal in three sinusoidal fringe patterns to eliminate the phase ambiguity, but without reducing the fringe amplitude or frequency. The absolute depth is then recovered through a robust region-wise voting strategy relying on the embedded signal. Using the theoretical minimum of only three images, the proposed method greatly facilitates the application of phase shifting in time-critical conditions. Moreover, the proposed method is resistant to the global illumination effects, as the fringe patterns used are with a single high frequency. Based on the proposed method, we further demonstrate a real-time, high-precision 3D scanning system with an off-the-shelf projector and a commodity camera. © 2013 Optical Society of America

**1. Introduction**

Phase shifting is a widely adopted structured light technique for high-precision and low-cost 3D shape measurement [1,2]. Theoretically, phase shifting requires at least three sinusoidal fringe patterns with an incrementally shifted phase to be projected and captured. However, due to the periodic nature of the fringe signal, a major problem of this technique is the phase ambiguity. That is, only depth within a range equivalent to one fringe period can be measured directly, and depth exceeding this range will be "wrapped." To recover the absolute depth, a prevalent solution is temporal phase unwrapping using a large number of additional patterns [3,4], which greatly limits the application of phase shifting in time-critical conditions. Recently, the advancement of camera and projector technology has enabled fast and accurate 3D scanning with phase shifting [5]. Hence, how to reduce the image-acquisition number for absolute depth recovery becomes critically important and has induced a great deal of research [6–10].

This paper presents a novel phase-shifting method using the theoretical minimum of only three images for absolute depth recovery. The basic idea is embedding a speckle-like signal in three sinusoidal fringe patterns. While the fringe components provide accurate yet ambiguous phase, the locally unique distribution of the speckle-like signal eliminates the phase ambiguity. To achieve competitive precision with conventional phase shifting methods using additional patterns, our method has two key contributions.

First, we embed the speckle-like signal by fully exploiting the limited redundancies in three phase-shifted fringe patterns. Specifically, the following conditions are satisfied simultaneously: (1) the phase calculation is independent of the embedded signal; (2) the fringe amplitude need not be reduced to
accommodate the embedded signal, so the signal-to-noise ratio (SNR) of phase shifting is retained; and (3) high-frequency fringe can be used together with the embedded signal for high-precision phase measurement.

Second, we propose an efficient and robust voting strategy to recover the absolute phase region by region, instead of pixel by pixel. Specifically, we take advantage of the fact that the disparities from the absolute phase on a continuous surface should be the same after direct spatial phase unwrapping [11]. Although the embedded signal may not be identified correctly at every location, the common disparity voted from a number of different locations on a continuous surface should be reliable.

Besides reducing the image-acquisition number to the theoretical minimum, our method has another distinct advantage. That is, global illumination effects, such as inter-reflections and subsurface scattering, can be alleviated to a large extent. In conventional phase-shifting methods using additional patterns, the low-frequency patterns required for phase unwrapping are vulnerable to global illumination and thus induce measurement errors [12]. Contrastively, our method is resistant to global illumination, as the fringe patterns used are with a single high frequency.

The proposed method aims to facilitate fast, accurate, and unambiguous 3D shape measurement with low-cost devices. We further develop a 3D scanning system with an off-the-shelf projector and a commodity camera. Experiments demonstrate real-time, high-precision 3D reconstruction for complex scenes with multiple isolated objects and abrupt surface changes, as well as challenging scenes with global illumination.

2. Related Work

Over the past decades, considerable progress has been made in phase-shifting profilometry, yet accurate and unambiguous 3D measurement still poses a major challenge and has attracted extensive attention in this field [2,13]. Several phase-unwrapping strategies have been developed consequently, which can be divided into the following categories.

Spatial phase unwrapping [11,14,15] can be directly applied to recover a continuous phase map from a wrapped one, but it cannot solve the ambiguity when there are multiple isolated objects or abrupt surface changes. Given these scene restrictions, Zhang and Huang developed a real-time, high-precision 3D scanner [5].

For unambiguous 3D measurement of complex scenes, temporal phase unwrapping using additional patterns is a prevalent solution. As one common choice, at least \( \lceil \log_2 N \rceil \) Gray-code patterns are required to unwrap a phase map with \( N \) periods [4]. Another popular method is multifrequency phase shifting that requires at least two phase maps with coprime frequencies (equivalent to at least six fringe patterns) [5]. In practice, three or more frequencies are used to guarantee the robustness of this method [16]. Temporal unwrapping is well-suited for static scenes but limits the measurement speed in time-critical conditions.

To further reduce the image-acquisition number in phase shifting, Wang et al. proposed a coding method in the phase domain that requires three additional stair patterns [10]. However, to ensure reliable decoding, the number of unique codewords is limited, which restricts the frequency of fringe and thus the accuracy of phase measurement. Liu et al. combined dual-frequency fringe in five patterns [9], where the low-frequency fringe disambiguates the high-frequency one. Still, the amplitude of the high-frequency fringe needs to be reduced, so the SNR of phase shifting is lowered [17].

Another solution for the ambiguity problem is adding a second camera for stereo matching [6,18], yet at the cost of increased hardware. Besides, calibration and synchronization between multiple cameras and the projector also increase the difficulty of system setup.

Recently, embedded period codes have been used in phase shifting for absolute depth recovery. Wissmann et al. embedded a 1D binary De Bruijn sequence into four fringe patterns as period codes [8], yet the fringe amplitude is reduced to accommodate the embedded signal. In order to retain the SNR of phase shifting, Wang et al. proposed a period coding strategy without reducing the amplitude of fringe [7]. To ensure reliable decoding, however, the frequency of fringe is restricted by the number of patterns, which limits the measurement precision with fewer patterns. Compared with these two methods, our method requires only three images while retaining both the full amplitude and the high frequency of fringe.

Global illumination effects, such as inter-reflections and surface scattering, pose obstacles for structured-light-based 3D measurement. Since Nayar and Krishnan showed that high-frequency patterns can be used to separate global illumination from the direct component [12], quite a few methods have been proposed to eliminate the global illumination effects [19–21]. Within this scope, reducing the image-acquisition number has also been well investigated, either using multiplexed illumination [20] or fringe patterns within a narrow high-frequency band [21]. Still, more than three images are needed for absolute depth recovery. While further reducing the required images to three, our method is resistant to global illumination in nature, as the fringe patterns used are with a single high frequency.

3. Speckle-Embedded Fringe

A. Pattern Design

In order to recover the accurate and absolute depth, we design three speckle-embedded fringe patterns in which the locally unique distribution of speckle eliminates the ambiguity of fringe. These patterns,
as shown in Fig. 1(a), are then used for three-step phase shifting. Mathematically, the \( k \)-th \((k = 1, 2, 3)\) pattern can be described as

\[
P_k(u, v) = B(u, v)Z(u, v) + C + A \cos\left(\frac{v}{V}2\pi N + \frac{2\pi}{3}k\right),
\]

where \(A\), \(C\), and \(N\) represent the amplitude, the DC offset, and the period number of a sinusoidal fringe signal, \(Z(u, v) \in \{0, 1\}\) denotes the distribution of a random speckle signal, \(B(u, v)\) controls the intensity of speckle, and \(V\) is the vertical resolution of the pattern (suppose the fringe is horizontal).

Speckle is widely used in digital image correlation (DIC) [22] and interferometry techniques [23] due to its high spatial distinguishability. Originally, speckle refers to the interference pattern caused by coherent radiation. For DIC techniques, artificial speckle is often simulated by randomly scattering white dots on a black background. In this work, we simulate a 2D binary speckle pattern \(Z\), according to the following rules:

**Speckle size:** The white dot has a regular size of \(K\times K\) pixels, where \(K\) is determined by the resolution of projector and camera used.

**Speckle density:** In each equivalent area to \(3 \times 3\) dots, only one dot is white.

**Speckle distribution:** The white dots are randomly scattered, and no two dots are adjacent in an 8-neighborhood.

The above rules aim to guarantee the distinguishability of speckle, not only in the projected pattern but also in the captured image. Part of the simulated speckle pattern \(Z\) is shown in Fig. 1(b).

Suppose the projected patterns \(P_k(k = 1, 2, 3)\) are \(L\)-bit gray scale. Generally, the amplitude and the DC offset of the fringe signal is set as \(A = C = 2^{L-1}\) to maximize the SNR of phase shifting [7]. To embed the speckle signal without reducing the amplitude of fringe, a straightforward solution is to assign \(B(u, v)\) as

\[
B(u, v) = 2^L - \max_{k = 1, 2, 3} \{F_k(u, v)\}. \tag{2}
\]

where

\[
F_k(u, v) = C + A \cos\left(\frac{v}{V}2\pi N + \frac{2\pi}{3}k\right). \tag{3}
\]

However, little speckle signal can be embedded when \(F_k(u, v)\) equals or gets very close to \(2^L\). Actually, the intensity of speckle can be either positive or negative. So a better strategy is to assign \(B(u, v)\) as

\[
B(u, v) = \begin{cases} 
2^L - B_{\max} & \text{if } 2^L - B_{\max} \geq B_{\min} \\
-B_{\min} & \text{otherwise}
\end{cases} \tag{4}
\]

where

\[
B_{\max} = \max_{k = 1, 2, 3} \{F_k(u, v)\}, B_{\min} = \min_{k = 1, 2, 3} \{F_k(u, v)\}. \tag{5}
\]

As illustrated in Fig. 1(c), the limited redundancies in three phase-shifted fringe patterns are fully exploited in this way.

**B. Phase Calculation**

In correspondence to Eq. (1), the \(k\)-th \((k = 1, 2, 3)\) captured image can be described as

\[
I_k(x, y) = I_C(x, y) + I_A(x, y) \cos\left[\phi(x, y) + \frac{2\pi}{3}k\right], \tag{6}
\]

where

\[
I_A(x, y) = a(x, y)A, \tag{7}
\]

\[
I_C(x, y) = a(x, y)[B'(x, y) + C + \beta(x, y)]. \tag{8}
\]

\[ \text{10 November 2013 / Vol. 52, No. 32 / APPLIED OPTICS 7799} \]
Here \((x, y)\) denote the camera coordinates to differentiate from \((u, v)\) denoting the projector coordinates. \(a(x, y), b(x, y)\), and \(B(x, y)\) represent the albedo, the ambient illumination, and the speckle intensity at each scene point.

There are three unknowns in Eq. (6): the amplitude \(I_4(x, y)\), the DC offset \(I_C(x, y)\), and the phase \(\phi(x, y)\). Theoretically, at least three images are needed to solve the three unknowns. The premise here is \(I_4(x, y)\) and \(I_C(x, y)\) are the same in \(I_k (k = 1, 2, 3)\). Since \(I_C(x, y)\) is now related to \(B(x, y)\), it means the embedded signal must be the same in \(P_k (k = 1, 2, 3)\) as formulated in Eq. (1). The phase can be then calculated as

\[
\phi(x, y) = \arg\max_{d} \gamma\{p, q|x_0, y_0 + \delta(d, p)|\}. \quad (12)
\]

where \(\delta(d, p)\) denotes the shift derived from \(p\) and \(d\), and \(q(x, y)\) denotes a small patch centered at \((x, y)\) in the reference image. Here we assume the projector and camera optics are aligned in the vertical direction so the pattern shift only occurs vertically.

Figure 1(d) shows the NCC statistics of a captured image of \(P_k\) at a reference plane, in terms of each \(5 \times 5\) patch with respect to pixel-by-pixel vertical shift in the image. We can see there is a primary peak at zero shift in the NCC mean curve, which indicates the embedded speckle signal is still distinguishable. However, periodic secondary peaks exist as the fringe components take up most of the energy. (In comparison, there is a distinctive peak in Fig. 1(e) showing the NCC statistics of a captured image of \(Z\) at a reference plane.) In view of that, and also considering the deformation caused by different surface changes in practice, the embedded speckle signal may not give the correct disparity at every location, especially when high-frequency fringe patterns are used. Therefore, we further propose an efficient and robust voting strategy to recover the absolute phase region by region, instead of pixel by pixel.

4. Absolute Phase Recovery

Figure 2(a) shows the flowchart of our method for absolute phase recovery, and Figs. 2(b)–2(f) give corresponding examples of intermediate results. After three images are captured, we first get a wrapped phase map, \(\Phi_W\), as in Fig. 2(c). Based on the wrapped phase map, two procedures are performed in parallel. One is spatial phase unwrapping [15], which produces a relatively unwrapped phase map \(\Phi_R\) as in Fig. 2(d). The other is continuous region detection, which produces a region mask \(\Psi\) indicating the boundaries of continuous surfaces as in Fig. 2(e). Since the \(2\pi\) discontinuities are removed in \(\Phi_R\), the phase value of each pixel in a continuous region should have a common disparity from the absolute phase. The region-wise disparities are then obtained through a voting procedure relying on the embedded speckle signal. Adding these disparities back to \(\Phi_R\), the absolute phase map is recovered and so is the absolute depth.

A. Continuous Region Detection

In the wrapped phase map, \(\Phi_W\), there are two kinds of discontinuities. One is the \(2\pi\) discontinuities caused by the periodicity of fringe, which only occurs in the vertical direction (for horizontal fringe); the other is caused by isolated objects or abrupt surface changes, which may occur in any direction. The latter actually indicates the boundaries of continuous surfaces. Therefore, we are able to perform a simple region segmentation on \(\Phi_W\). As shown in Fig. 3(a), considering a pixel in \(\Phi_W\) with phase value \(\phi_0\), it is determined to be connected with its four nearest neighbors if

\[
\hat{d}(x_0, y_0) = \arg\max_{d=0, \ldots, N-1} \gamma\{p, q|x_0, y_0 + \delta(d, p)|\}.
\]
up: \( \max\{d(\phi_0, \phi_1), d(\phi_0, \phi_5), d(\phi_0, \phi_6)\} < T_V \). \hspace{1cm} (13)

down: \( \max\{d(\phi_0, \phi_3), d(\phi_0, \phi_7), d(\phi_0, \phi_8)\} < T_V \). \hspace{1cm} (14)

left: \( \max\{d(\phi_0, \phi_4), d(\phi_0, \phi_9)\} < T_H \). \hspace{1cm} (15)

right: \( \max\{d(\phi_0, \phi_2), d(\phi_0, \phi_{10})\} < T_H \). \hspace{1cm} (16)

where

\[
d(\phi_i, \phi_j) = \min\{|\phi_i - \phi_j|, |\phi_i - \phi_j + 2\pi|, |\phi_i - \phi_j - 2\pi|\}.
\]

\hspace{1cm} (17)

\( T_V \) and \( T_H \) are two predefined thresholds.

The region segmentation is based on a two-pass floodfill algorithm. In the first pass, the neighboring pixels are labeled as the same region if a certain condition among Eqs. (13)–(16) is satisfied. In the second pass, tiny regions that are smaller than a threshold are either removed as noise or merged into neighboring regions, depending on whether they are isolated or not. Then we have the continuous region mask \( \Psi \). Figure 3(b)–3(d) give several typical examples where two continuous surfaces are detected.

B. Disparity Voting

Suppose there are \( M \) continuous regions in the region mask \( \Psi \). To recover the absolute phase, we only need to find the region-wise disparities \( \{D_m\}_{m=1}^M \), where \( D_m \) is an integer within \([0, N - 1]\). As mentioned above, the embedded speckle signal may not give the correct disparity at every location. However, if we look at a number of different locations in a continuous region, the common disparity voted by the majority will be much more reliable. Figure 4 shows the disparity voting results for the scene in Fig. 2, where high-frequency fringe components are used. The five histograms represent five continuous regions, and 3% points are randomly sampled in each region from each of the three captured images. As can be seen, on the one hand, there is a distinct peak in each histogram in accordance with the correct disparity, which demonstrates the reliability of voting. On the other hand, not all locations are correctly identified, which suggests the necessity of voting.

5. Resistance to Global Illumination

As pointed out in [12], high-frequency fringe patterns are resistant to the global illumination effects, as global illumination at each scene point will remain constant and can be counteracted in Eq. (9). Using three fringe patterns with a single high frequency, the proposed method solves the phase ambiguity by the embedded speckle signal and is resistant to global illumination in nature. We further design an experiment to validate this property.

In Fig. 5(a), a slant planar surface \( S_1 \) with uniform albedos is placed in the scene, where there is only

---

**Fig. 2.** (a) Flowchart of our method for absolute phase recovery. (b) One of the three captured images. (c) Wrapped phase map. (d) Spatially unwrapped phase map. (e) Continuous region mask (each color denotes a continuous surface). (f) Absolute depth map.

**Fig. 3.** (a) Connectivity decision in a four-neighborhood. (b)–(d) Typical examples of continuous region detection.
direct illumination from the projected light. In Fig. 5(b), another planar surface $S_2$ is placed against $S_1$ to form a V-groove, and now there is global illumination caused by inter-reflections in the scene. Under these two cases, we perform the proposed method with two different fringe frequencies and conduct the following comparisons.

First, we investigate the influence of global illumination on speckle disambiguation. Specifically, the disparity voting results from all locations on $S_1$ are presented. As can be seen in Fig. 5(c), the results with and without global illumination are well coincident with each other, and both give the correct disparity. Figure 5(d) shows similar results when the fringe components are with a higher frequency. This comparison demonstrates that the distinguishability of speckle is not degraded in the presence of global illumination, either with low- or high-frequency fringe.

Second, we investigate the influence of global illumination on phase calculation. Given that the correct disparity of $S_1$ has been solved, this influence is reflected by the difference of the recovered absolute depth on $S_1$, with and without global illumination. The statistics of averaged depth difference along each column on $S_1$ are shown in Fig. 5(e). It can be observed that, for the low-frequency fringe, the depth difference increases considerably on the right side of $S_1$ where the global illumination tends to be more severe. In comparison, the high-frequency fringe drastically reduces the global illumination effects, which is in accordance with the insights revealed in [12].

This experiment demonstrates that the proposed method with high-frequency fringe is resistant to global illumination in terms of both speckle disambiguation and phase calculation.

6. Experimental Results

Based on the proposed method, we establish a 3D scanning system with an off-the-shelf DLP projector and a commodity monochrome CCD camera. The projector emits 8 bit gray-scale patterns with 1024 × 768 resolution, while the camera captures 8 bit gray-scale images with 640 × 480 resolution. The speckle size is set to 3 × 3 pixels, so the speckle can be distinctly recorded by the camera. The NCC is calculated on 5 × 5 patches in the captured images. The default period number of fringe is set as 30. The two thresholds for continuous region detection are empirically determined as $T_V = 0.16\pi$ and $T_H = 0.08\pi$. During disparity voting, 3% points are randomly sampled in each continuous region from each of the three captured images. Our current setup allows 3D scanning in a frustum of 60 cm depth with a 64 cm × 48 cm front end and a 96 cm × 72 cm back end.

A. Quantitative Evaluation

To evaluate the accuracy of the proposed method, we compare it with three different phase-shifting methods. The first method is Gray-code+phase shifting (GC+PS) using $\lceil \log_2 N \rceil$ Gray-code patterns besides three fringe patterns, where $N$ is the period number of fringe. The second method is three-frequency phase shifting (3F-PS) using nine fringe patterns. The last method, named composite phase shifting, simply reduces the fringe amplitude to 80% and embeds a random speckle signal in the remaining intensity range. Composite phase shifting can also recover the absolute depth using three images, following the same flowchart of the proposed method.

Figure 6(a) shows the root-mean-square (RMS) reconstruction error for a planar surface at different depths, averaged over 10 times of measurements for
The RMS error is calculated at each point and then averaged on the whole surface. We can see the reconstruction error decreases at nearer depth due to improved SNR (about 0.5 mm at 1300 mm). The performance of our method is quite close to both GC+PS and 3F-PS, but composite phase shifting lowers the measurement precision to a certain extent, since the amplitude of fringe is reduced.

Similarly, Fig. 6(b) shows the evaluation result in terms of the period number of fringe. As can be observed, the reconstruction error decreases as the period number increases (about 0.5 mm for 30 periods). Again, our method achieves competitive precision with GC+PS and 3F-PS especially at high frequency, but uses much fewer patterns. In contrast, composite phase shifting degrades the measurement precision considerably.

B. Qualitative Evaluation

Figure 7 displays the 3D shape results of several complex scenes containing multiple objects. In Fig. 7(a), there are two isolated statues, “David” and “Venus,” and a cushion behind David. From the view of the camera, the heads of David and Venus are not continuous with their bodies due to abrupt surface changes. So there are actually five continuous regions with potentially different period disparities. In this case, spatial unwrapping would fail to recover the absolute phase. Contrastively, the proposed

Fig. 6. (a) Reconstruction error against depth (using fringe with 30 periods). (b) Reconstruction error against the period number of fringe (at 1300 mm depth).

Fig. 7. 3D reconstruction of several complex scenes in front, side, and top views. (The small color maps denote continuous region masks.)

Fig. 8. 3D reconstruction of scenes with global illumination. (a) Color scene image. (b) Proposed method using three images. (c) Three-frequency phase shifting using nine images.
method accurately reconstructs the entire scene, with only three images. In Fig. 7(b), a lamp separates a planar surface and also blocks one corner of a face from the view of the camera. Again, the 3D position of each surface is correctly recovered. In Fig. 7(c), a human places one arm on a step-height cube with the other arm far behind. While the unambiguous 3D scene is reconstructed, the drape on the clothes, the bone texture on the hand, and the uniformity of the step-height cube are well preserved, which demonstrates the high precision of our method.

Two challenging scenes—a ceramic bowl and a V-groove—are tested in Fig. 8. Here global illumination effects especially inter-reflections pose obstacles for the 3D measurement. Our method is resistant to global illumination, as can be seen from the accurate results. In comparison, three-frequency phase-shifting introduces errors as low-frequency patterns are involved for phase unwrapping.

C. Dynamic Response

Figure 9 presents the online reconstruction for two dynamic scenes at 15 fps. We can see the accurate and unambiguous 3D shapes of moving parts are consistently recovered along with the static parts. In our system, all the computations are accomplished on a single CPU core with 2.83 GHz working frequency. The current measurement speed is 30–60 ms for generating a depth map, depending on the scene content.

The detailed execution time of several main procedures, for the scenes in Fig. 9, is listed in Table 1. With parallel implementation, e.g., on GPU or FPGA, even higher measurement speed is promising. In addition, motion compensation strategies, such as the one proposed in [6], can be further adopted to alleviate the hardware delay within three images.

D. Limitations

The proposed method may not recover the absolute depth correctly in some special cases. One example is tiny surfaces with abrupt depth changes relative to the surrounding, such as a small and deep hole. These surfaces may be removed as noise or merged into neighboring surfaces during region segmentation, or there may not be sufficient voting candidates to give a reliable disparity. Another example is two overlapped surfaces with the same wrapped phase values but different disparities along the boundaries. Dealing with these special cases deserves future efforts. Still, the proposed method is quite robust for 3D scanning of general scenes.

7. Conclusion

Phase shifting is a promising approach for high-precision and low-cost 3D shape measurement, yet the intrinsic phase ambiguity limits its application in time-critical conditions. In this paper, we have presented a novel method to solve the ambiguity by using an embedded speckle signal instead of additional patterns. Moreover, the high precision of phase measurement is well retained since neither the amplitude nor the frequency of fringe is reduced. The proposed method is also resistant to global illumination, as the fringe patterns used are with a single high frequency. With the rapidly increasing demand for fast and accurate 3D scanning nowadays, the proposed method is expected to be useful in a variety of applications.

Table 1. Execution Time of Main Procedures (ms)

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Hand Gesture</th>
<th>Face Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrapped phase</td>
<td>4.97</td>
<td>4.94</td>
</tr>
<tr>
<td>Spatial unwrapping</td>
<td>9.37</td>
<td>9.13</td>
</tr>
<tr>
<td>Region detection</td>
<td>3.60</td>
<td>3.55</td>
</tr>
<tr>
<td>Disparity voting</td>
<td>15.77</td>
<td>15.62</td>
</tr>
<tr>
<td>Subtotal</td>
<td>33.71</td>
<td>33.24</td>
</tr>
</tbody>
</table>

Fig. 9. Online 3D reconstruction of dynamic scenes at 15 fps. Top: hand gesture; bottom: face expression (with black eyeglasses).
This work was done while Yueyi Zhang was an intern at Microsoft Research Asia. The authors also would like to thank Zhe Yang for assistance in system implementation.

References