Supplementary information for:

Dynamic 3D shape reconstruction under complex reflection and transmission conditions using multi-scale parallel single-pixel imaging

Zhoujie Wu, Haoran Wang, Feifei Chen, Xunren Li, Zhengdong Chen, and Qican Zhang[*]

School of Electronics and Information Engineering, Sichuan University, Chengdu, Sichuan Province 610065, China

*zqc@scu.edu.cn

Abstract

This document provides supplementary information for "Dynamic 3D shape reconstruction under complex reflection and transmission conditions using multi-scale parallel single-pixel imaging".

Contents

Supplementary Note S1: Theoretical basis of PSI algorithm

The emerging SI technique has the property to retrieve a 2D image only using a single-pixel detector. And the convergent lens in this technique to focus and collect the varying projected structured light is equivalent to the generated mixed illumination. Therefore, it is introduced into the 2D sensor to separate the direct and global illumination in the traditional triangulation system. Every pixel of the used camera is treated as a single-pixel detector and SI imaging algorithm is performed in parallel. And then the position of different illumination components for this pixel can be separated and located in the projecting plane according to the Helmholtz reciprocity. So, this technique is named parallel single-pixel imaging (PSI).

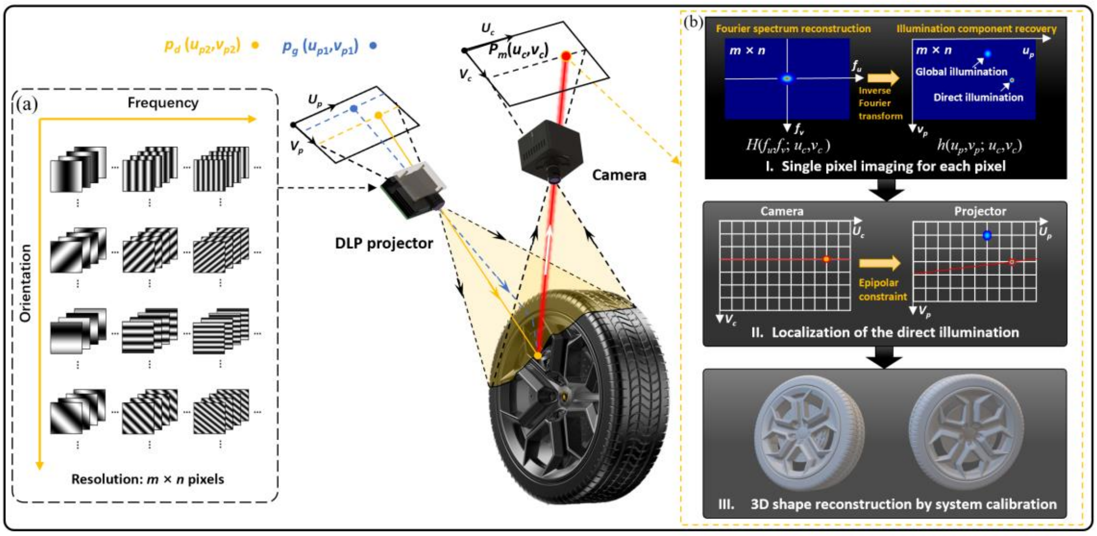

Figure S1. Reconstruction process of existing PSI methods for 3D shape measurement under global illumination. a, Projected Fourier basis pattern in PSI. **b**, 3D reconstruction process of PSI algorithm.

Light transport equation is used to describe the imaging process in PSI:

$$I(u_c,v_c) = A(u_c,v_c) + \sum_{v_p=0}^{n-1}\sum_{u_p=0}^{m-1} h(u_p,v_p;u_c,v_c)P(u_p,v_p), \tag{S1}$$

in which, $I(u_c,v_c)$ is the detected intensity of the mixed illumination for any camera pixel (u_c,v_c); $A(u_c,v_c)$ denotes the ambient illumination into the same pixel; $P(u_p,v_p)$ is projected intensity for pixel (u_p,v_p) in projecting plane with the resolution of $m \times n$ pixels; and $h(u_p,v_p;u_c,v_c)$ denotes the light transport coefficient between pixels (u_c,v_c) and (u_p,v_p), which describes the light intensity transferring ratio between any two pixels in projecting and imaging planes. In PSI, a series of Fourier basis patterns are projected as shown in Fig. S1(a), and every camera pixel is treated as a single-pixel detector to retrieve a 2D image. In this condition, 2D light transport coefficient $h(u_p,v_p;u_c,v_c)$ for pixel (u_c,v_c) exactly equals to the SI imaging result of this pixel, which can be solved by Fourier single-pixel imaging algorithm [1]. The required projected sinusoidal patterns with average a and contrast b are described as:

$$P_i(u_p,v_p;f_u,f_v)=a+b\cos[2\pi(f_u u_p+f_v v_p)+\delta_i],i=1,2,3...N, \tag{S2}$$

where f_u and f_v are the frequency component of the projected patterns in u and v directions, respectively, which can be further calculated by $f_u=k/m$, $k=0,1\ldots m-1$ and $f_v=l/n$, $l=0,1\ldots n-1$. N is the phase-shifting step and $\delta_i= 2(i-1)/N$. And extending Eq. (S1) into Eq. (S2), we obtain:

$$I_i(f_u,f_v;u_c,v_c)=A(u_c,v_c)+\sum_{v_p=0}^{n-1}\sum_{u_p=0}^{m-1}h(u_p,v_p;u_c,v_c)P_i(u_p,v_p;f_u,f_v). \tag{S3}$$

And the Fourier spectrum can be reconstructed using the collected information under different illumination:

$$H(f_u,f_v;u_c,v_c)=[\sum\nolimits_{i=1}^{N}I_i(f_u,f_v;u_c,v_c)\sin\delta_i]+j[\sum\nolimits_{i=1}^{N}I_i(f_u,f_v;u_c,v_c)\cos\delta_i]. \tag{S4}$$

After doing fast inverse Fourier transform, the required 2D light transport coefficient $h(u_p,v_p)$ for each camera pixel (u_c,v_c) can be obtained:

$$h(u_p,v_p;u_c,v_c)=F^{-1}[H(f_u,f_v;u_c,v_c)]/2b. \tag{S5}$$

As shown in Fig. S1(b), the obtained direct and global illumination components are separated in the projecting plane for specific camera pixel (u_c,v_c). According to precondition in triangulation that for any given camera pixel, the direct illumination component from the light emitter will locate on the epipolar determined by this pixel [2]. Therefore, epipolar constraint and subpixel searching are applied to further locate the accurate position of the direct illumination component. By this way, the unique corresponding relationship between projecting and imaging planes can be established and 3D shape measurement can be achieved through stereo vision algorithm[3] in mixed illumination conditions. However, for the PSI algorithm, the number of required Fourier basis patterns for one reconstruction, which is inefficient for fast measurement.

Supplementary Note S2: Comparative discussion between existing local region extension (LRE) method and presented MS-PSI method.

In Jiang's work[4], they wisely proposed local region extension (LRE) method to significantly decrease the required number of the proposed patterns in the naive single-pixel imaging (SI) technique. In the LRE method, it is assumed that any pixel on the camera can only receive light from a local region on the projector, which makes the visible region for each pixel confined in a local region. Based on this point, the core idea of LRE method is dividing the full resolution of the projector into several local regions. And after that, the Fourier basis patterns are generated with the size of local region and duplicated to fill other local regions in the projecting plane as shown in Fig. S2(a). In the reconstruction stage, the local region is treated as a unit to perform SI algorithm. In such way, the number of projected patterns will be largely reduced with the decrease in the size of the encoding region. But, the precondition of LRE reconstruction theorem is that, if the period of the periodic extension patterns covers the visible region, the light transport coefficients can be perfectly reconstructed. To guarantee this precondition, "coarse to fine" localization procedure is performed. Two groups of full-resolution illumination patterns with vertical and horizontal orientations are firstly projected to detect and obtain the size of the projected reception region using the 1D Fourier slice theorem as shown in Fig. S2(a). The maximum external rectangle of the reconstructed light transport coefficient among all pixels will be determined as the size of the local region. And then projected patterns can be generated and duplicated in each local regions for further fine localization. As a result, this strategy needs extra projected patterns for determining the size of local region, decreasing the measuring efficiency. More importantly, coarse localization will be redone and the projected patterns will be updated and reprojected when the tested scene changes, which makes it difficult and even impossible to apply PSI in dynamic measurement.

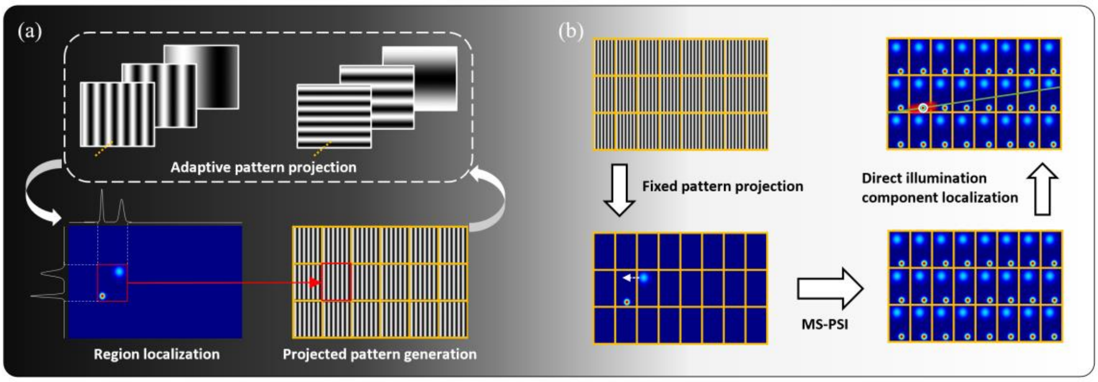

Figure S2. Illustration of principles of the a, LRE method and b, MS-PSI method.

As a contrast, the proposed MS-PSI method makes no assumption and requirements on light reflection condition of the measured scene as shown in Fig. S2(b). The projected plane is equally dividedly into local regions with any scales which only depends on the detecting accuracy we want. Even the location of the illumination component exceeds one local region, the MS-PSI method can accurately locate the position of the direct illumination component owing to the position invariant theorem. Therefore, the fixed generated illumination patterns with a small number are used to be projected for dynamic measurement, which breaks the assumption and rules in existing methods.

Supplementary Note S3: Implementation of the depth-constrained localization

In the proposed method, only vertical and horizontal Fourier basis patterns are projected as shown in Fig. S3(a). But, if only two Fourier slices are reconstructed, extra ambiguous points will occur in the back projection image as shown in Fig. S3(b). And in traditional method, Fourier slices with extra direction is required to reject ambiguous points which decreases measuring efficiency. In addition, the whole epipolar is used to find the direct component, so the wrong discrimination and location of the direct illumination component easily happens when serval candidate points disturbed by noise fall around the epipolar. However, in actual measurement, measuring depth range corresponds the length-limited epipolar and limited measuring depth is often calibrated to guarantee stable measuring accuracy. Hence, standard planes at the closest (plane 1) and farthest (plane 2) position in measuring range are used to generate length-limited epipolar and eliminate wrong points out of the defined range as shown in Fig. S3(b). It can be found that p_1 in plane 1 and p_2 in plane 2 corresponds to the constrained epipolar lined by $P_{p1}(u_{p1},v_{p1})$ and $P_{p2}(u_{p2},v_{p2})$, which can eliminate the ambiguous candidate. To calculate the length-limited epipolar for each pixel, plane 1 and plane 2 are measured using the traditional Gray code plus phase shifting method [5] by projecting vertical and horizontal fringe patterns in the calibration process as shown in Fig. S3(c). The absolute phase Φ_v and Φ_h can be solved and the accurate pixel position of the direct illumination in projecting plane for each camera pixel can be determined and recalibrated using:

$$u_p = \Phi_v(u_c, v_c) / 2\pi \times \frac{m}{f_u} \tag{S6}$$

$$v_p = \Phi_h(u_c, v_c) / 2\pi \times \frac{n}{f_v}, \tag{S7}$$

in which, m/f_u and n/f_v are the fringe period of sinusoidal patterns in two directions. Hence, for any camera pixel, two points $P_{p1}(u_{p1},v_{p1})$ and $P_{p2}(u_{p2},v_{p2})$ can be located in the projecting plane, and the direct illumination will lie on the connecting line between these two points in the measuring volume. In this way, searching area for direct illumination can be constrained into a short line, which can avoid ambiguous points in dual-slice localization and speedup the searching process.

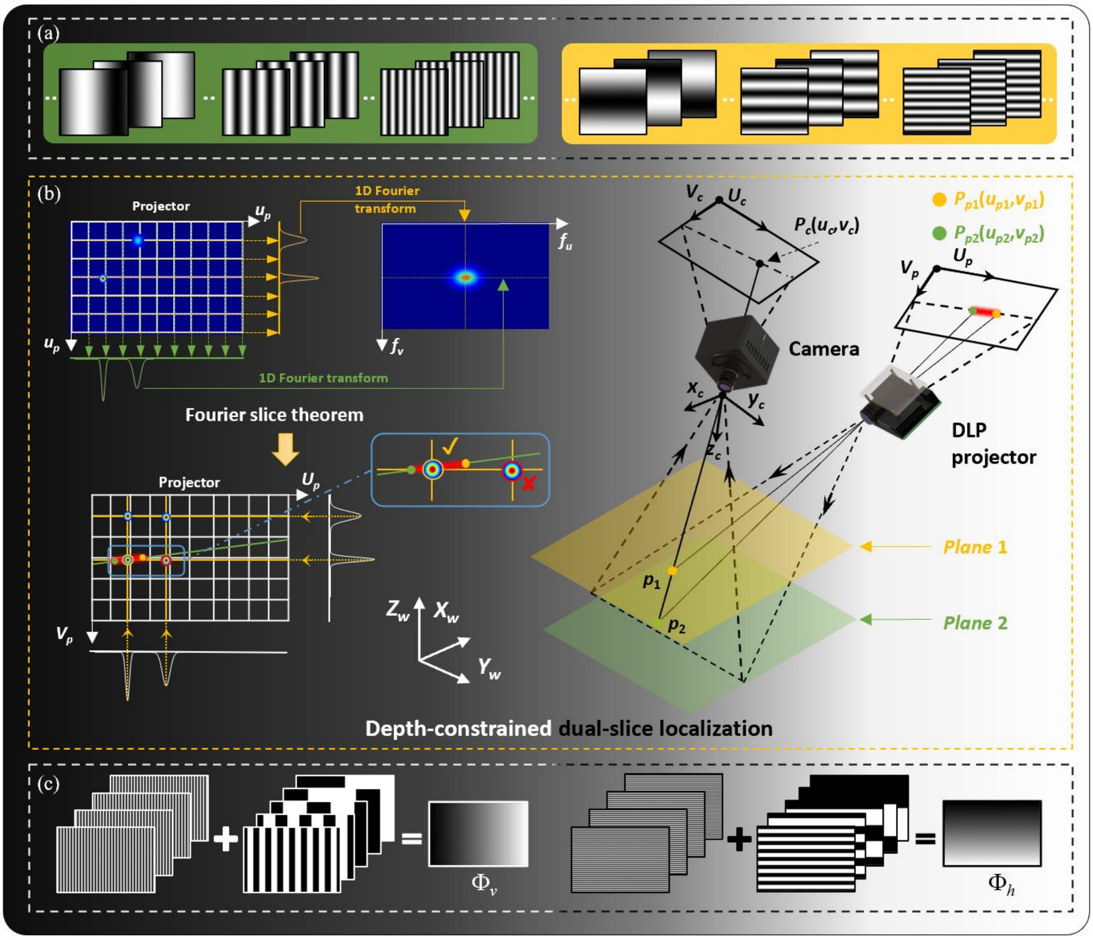

Figure S3. Implementation of depth-constrained localization. a, Projected two-direction Fourier basis patterns for measurement. **b,** Illustration of depth-constrained dual-slice localization. **c,** Projected structured light patterns for mapping corresponding points between projecting and imaging planes in calibration process.

Supplementary Note S4: Proof of the position invariant theorem

Position invariant theorem: *If the projecting patterns is equally divided and regenerated into subregions with any scale in PSI, the location of the detected direct illumination component remains unchanged in the reconstructed image using MS-PSI.*

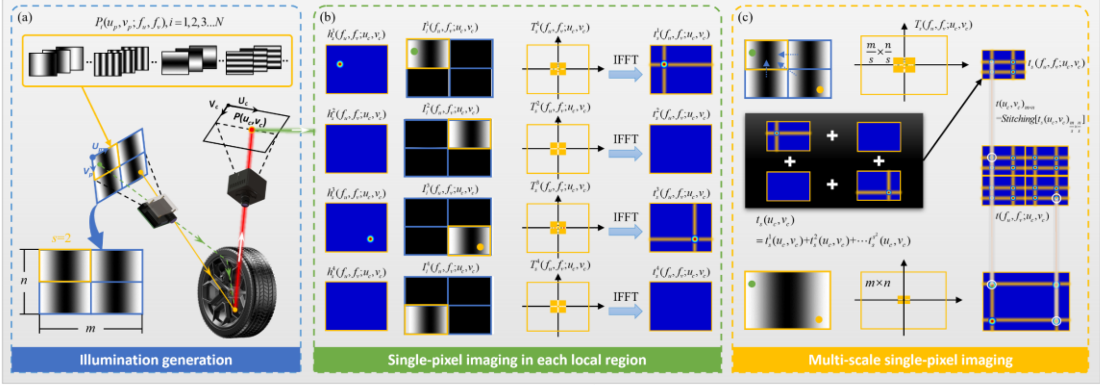

Figure S4. Illustration of position invariant theorem. a, Illumination pattern generation when *s*=2. **b**, Single-pixel imaging for each illumination subregion, respectively. **c**, Multi-scale single-pixel imaging for the whole illumination region.

To prove the position invariant theorem, we summarize the whole process of the multi-scale single-pixel imaging algorithm as shown in Fig. S4. Firstly, the projecting vertical and horizontal phase-shifting sinusoidal patterns with different frequencies are generated and duplicated into $s\times s$ subregions ($s=2$ in this case), and the projected patterns can be described as:

$$P_i(u_p,v_p;f_u,f_v)=a+b\cos[2\pi(f_u u_p+f_v v_p)+\delta_i], i=1,2,3...N. \tag{S8}$$

N is chosen as 3 for dynamic measurement. And then all generated patterns are projected onto the tested surface in sequence. The camera collects the reflected light from another angle and each pixel in camera is treated as a single-pixel detector to retrieve the collected light field. The corresponding light transport coefficient $h_s^r(u_p,v_p;u_c,v_c)$ in each subregion is used to describe the transporting path and intensity of the light and detected intensity $I_i^r(f_u,f_v;u_c,v_c)$ in a specific camera pixel (u_c,v_c) from different subregion can be represented as:

$$I_i^r(f_u,f_v;u_c,v_c)=A(u_c,v_c)+\sum_{v_p=0}^{n/s-1}\sum_{u_p=0}^{m/s-1}h_s^r(u_p,v_p;u_c,v_c)P_i(u_p,v_p;f_u,f_v), r=1,2,3...s^2. \tag{S9}$$

It should be noted that although projected patterns are the same in each subregion, the light transport coefficient h_s^r is different, which means the emitted light from different position in each subregion will be detected by the pixel in camera. To illustrate this phenomenon, it is assumed that each subregion can illuminate and be collected independently, and then the generalized light transport coefficient $t_s^r(u_p, v_p; u_c, v_c)$ for each subregion can be reconstructed as shown in Fig. S4(b). However, in actual measurement, all subregions illuminate simultaneously and the fact is that the detected intensity is the superposition of illumination from all subregions. Therefore, the reconstructed spectrum of actual generalized light transport coefficient $t_s(u_p, v_p; u_c, v_c)$ is described as:

$$T_s(f_u, f_v; u_c, v_c) = \left\langle \{\sum_{i=1}^{N}[\sum_{r=1}^{s^2} I_i^r(f_u, f_v; u_c, v_c)]\sin\delta_i\} + j\{\sum_{i=1}^{N}[\sum_{r=1}^{s^2} I_i^r(f_u, f_v; u_c, v_c)]\cos\delta_i\} \right\rangle \cdot W_s(f_u, f_v). \tag{S10}$$

After expanding this formula and we can obtain:

$$\begin{aligned} T_s(f_u, f_v; u_c, v_c) &= \left\langle \begin{aligned} &\{\sum_{i=1}^{N}[I_i^1(f_u, f_v; u_c, v_c) + I_i^2(f_u, f_v; u_c, v_c) + \cdots I_i^{s^2}(f_u, f_v; u_c, v_c)]\sin\delta_i\} \\ &+ j\{\sum_{i=1}^{N}[I_i^1(f_u, f_v; u_c, v_c) + I_i^2(f_u, f_v; u_c, v_c) + \cdots I_i^{s^2}(f_u, f_v; u_c, v_c)]\cos\delta_i\} \end{aligned} \right\rangle \cdot W_s(f_u, f_v) \\ &= \left\{[\sum_{i=1}^{N} I_i^1(f_u, f_v; u_c, v_c)\sin\delta_i] + j[\sum_{i=1}^{N} I_i^1(f_u, f_v; u_c, v_c)\cos\delta_i]\right\} \cdot W_s(f_u, f_v) + \\ &\left\{[\sum_{i=1}^{N} I_i^2(f_u, f_v; u_c, v_c)\sin\delta_i] + j[\sum_{i=1}^{N} I_i^2(f_u, f_v; u_c, v_c)\cos\delta_i]\right\} \cdot W_s(f_u, f_v) + \cdots \\ &\left\{[\sum_{i=1}^{N} I_i^{s^2}(f_u, f_v; u_c, v_c)\sin\delta_i] + j[\sum_{i=1}^{N} I_i^{s^2}(f_u, f_v; u_c, v_c)\cos\delta_i]\right\} \cdot W_s(f_u, f_v) \\ &= T_s^1(f_u, f_v; u_c, v_c) + T_s^2(f_u, f_v; u_c, v_c) + \cdots T_s^{s^2}(f_u, f_v; u_c, v_c). \end{aligned} \tag{S11}$$

It indicates the reconstructed spectrum of the actual generalized light transport coefficient is the sum of the spectrum of each subregion illumination independently. And after inverse Fourier transform, we can get:

$$t_s(u_c, v_c) = F^{-1}[T_s(u_c, v_c)]/2b = F^{-1}[T_s^1(u_c, v_c) + T_s^2(u_c, v_c) + \cdots T_s^{s^2}(u_c, v_c)]/2b = t_s^1(u_c, v_c) + t_s^2(u_c, v_c) + \cdots t_s^{s^2}(u_c, v_c). \tag{S12}$$

At this point, we can draw an important lemma: the generalized light transport coefficient of full-field illumination is the sum of that with each subregion illumination independently.

With this lemma, it can be explained why the location of the detected direct illumination component remained unchanged in the reconstructed image using MS-PSI as shown in Fig. S4(c). In MS-PSI, one subregion rather than whole region is our reconstructed target, but illumination from all subregions makes contributions to detected intensity. Due to the same projecting patterns in each subregion, it can be equivalent that all the subregions are periodically moved and merged in a unified coordinate system and the retrieved image is a superposition of the generalized light transport coefficient for each subregion as proved in Eq. (S12). After obtaining the actual generalized light transport coefficient $t_s(u_p, v_p; u_c, v_c)$ in subregion, the full-field generalized light transport coefficient $t(u_p,v_p;u_c,v_c)$ can be obtained by duplicating and stitching $t_s(u_p,v_p;u_c,v_c)$:

$$t(u_p,v_p;u_c,v_c)_{m\times n}=Stitching[t_s(u_p,v_p;u_c,v_c)_{\frac{m}{s}\times\frac{n}{s}}]. \tag{S13}$$

Whether the direct illumination component locates in the chosen reconstructed subregion or not, the periodically duplication will move it back to the original position which can be aligned with the full-field SI reconstructing result as shown in Fig. S4(c). Therefore, the position invariant theorem has been proved.

Supplementary Note S5: Simulation on traditional spectrum component optimization and MS-PSI optimization

To demonstrate the sampling and locating optimization ability, simulation on traditional spectrum component optimization and MS-PSI optimization is performed. As shown in Fig. S5(a), the light transport coefficient with discrete energy distribution is generated and it contains three illumination regions as shown in Figs. S5(b) and S5(c), which indicates the detected energy for this pixel comes from three positions in projecting plane.

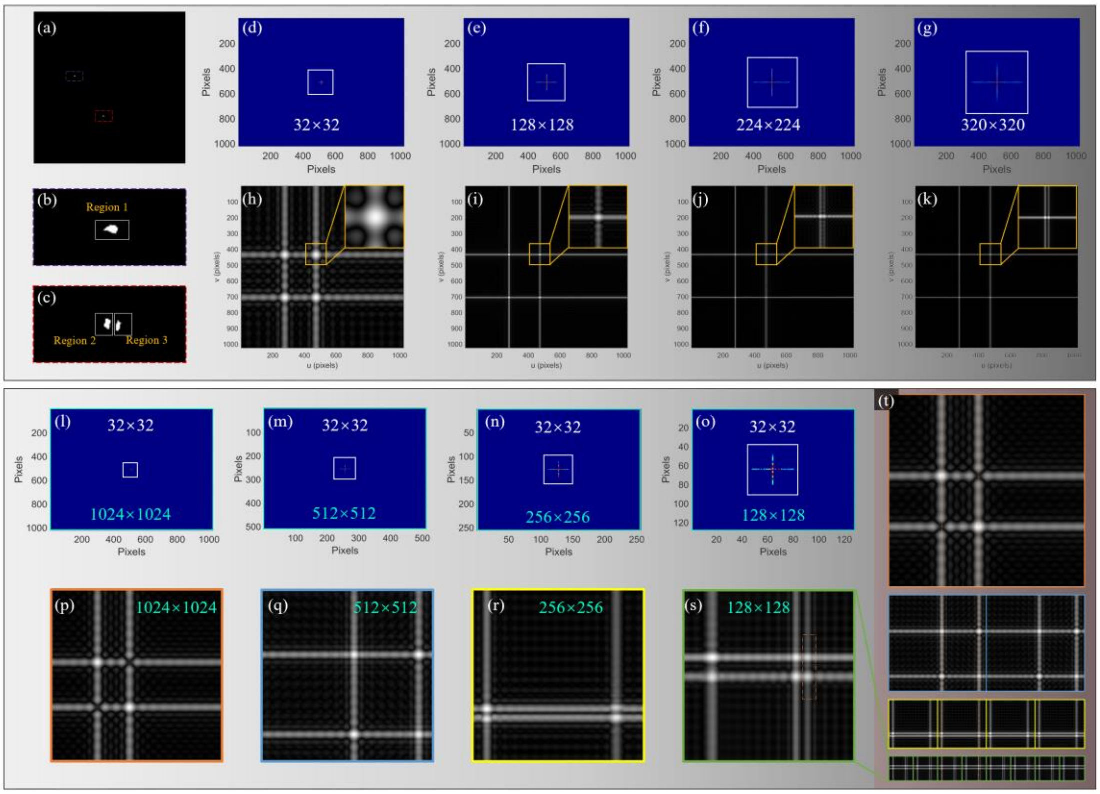

Figure S5. Simulation on traditional spectrum component optimization and MS-PSI method. **a**, Simulated discrete light transport coefficient. **b-c**, Amplified images of the subregions in **a**. **d-g**, Increasement of the sampling coefficients for image optimization using traditional method. **h-k**, Corresponding optimization result using traditional method. **l-o**, Decrease of the imaging scale for image optimization using MS-PSI method. **p-s**, Corresponding optimization result using MS-PSI method. **t**, Optimization result comparison using different scale factor *s*.

After applying Fourier transform, two Fourier slices with different sampling rates are used to reconstruct the generalized light transport coefficient $t(u_p,v_p;u_c,v_c)$. It can be shown that when the number of sampling coefficients is small (32×32 and 128×128) as shown in Figs. S5(d) and S5(e), the locations between region 2 and region 3 cannot be distinguished as shown in Figs. S5(h) and S5(i). With the increasement of the sampling rate as shown in Figs. S5(f) and S5(g), the gap between two regions become clear and can be distinguished as shown in Figs. S5(j) and S5(k). Simulation results show traditional spectrum coefficients optimization method improve the detection resolution by increasing the number of the spectrum coefficients, which is not suitable for efficient 3D measurement.

As a contrast, the proposed MS-PSI method improves the detection resolution by reducing the scale of the imaging region. By this way, the sampling rate of the reconstructed spectrum can be equivalently increased with the fixed spectrum coefficients (32×32) as shown in Figs. S5(l)- S5(o). And it results in higher imaging resolution and aids in distinguishing the region 2 and region 3 as shown in Figs. S5(p)-S5(s). Furthermore, the location of the detected direct illumination component remained unchanged in the reconstructed image using MS-PSI as shown in Fig. S5(t), which has been proved in position invariant theorem as seen in Supplementary Information Note S4. Therefore, the desired direct illumination component with higher resolution can be located using the MS-PSI method with limited spectrum coefficients, which is superior in dynamic measurement.

Supplementary Note S6: Performance comparison between FPP and MS-PSI on surfaces with different materials

To compare measuring performance between FPP and MS-PSI on surfaces with different materials, a scene contains a white paper, ice and candle was measured. Two-frequency 3-step phase-shifting patterns (ω_l=4, ω_h=16) are projected in FPP as shown in Fig. S6(a), and the geometric-constrained two-frequency phase-shifting method [5] and calibration are used to obtain absolute phase and depth map as shown in Figs. S6(b) and S6(c).

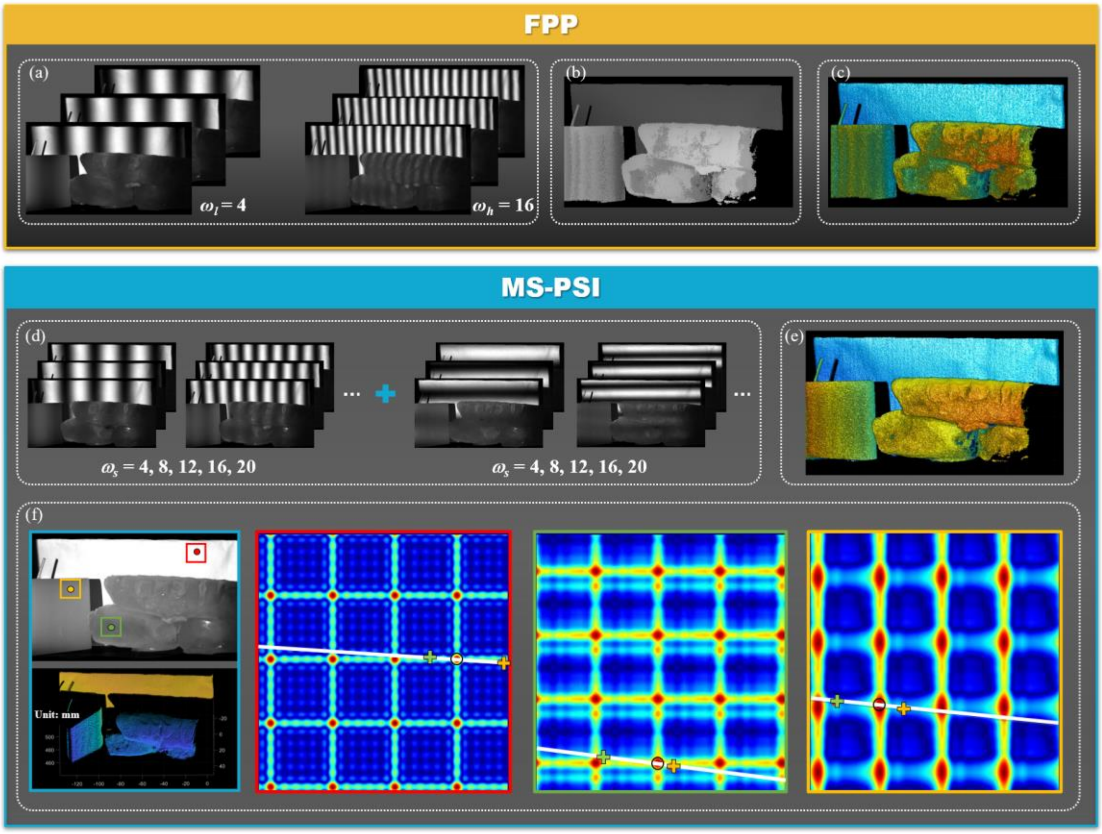

Figure S6. Comparative experiment on scene with subsurface scattering. a, Captured two groups of sinusoidal patterns using FPP. **b,** Reconstructed phase by FPP. **c,** Reconstructed depth map by FPP. **d,** Captured two groups of Fourier basis patterns in MS-PSI. **e,** Reconstructed depth map by MS-PSI. **f,** Reconstructed generalized light transport coefficient for points on different types of surfaces.

It can be found that reconstructed errors gathered in the strong scattering regions with low fringe contrast. This is because FPP method highly depends on the fringe quality and contrast for high-accuracy shape reconstruction. As a contrast, five groups of two-direction Fourier basis patterns (c=5, ω=8, 16, 24, 32, 40) are projected in MS-PSI method as shown in Fig. S6(d), and the scale factor s=4. The reconstructed depth map is given in Fig. S6(e). And the generalized light transport coefficient for three specific points on white paper, ice and candle are shown in Fig. S6(f). It indicates that single and energy-concentrated illumination component occurs on diffuse paper surface while it becomes a diffuse spot on ice and candle surface due to the subsurface scattering. In addition, indirect illumination emerges on ice surface because of multilayer reflection. However, the defocus of the spot causes tiny influence on positioning the gravity center of the direct component [6] and depth-constrained dual-slice localization can assist to exclude ambiguous indirect component. So, the complete and accurate depth map and 3D cloud data can be reconstructed by MS-PSI. Results demonstrated the proposed method has better performance on strong scattering regions compared with traditional FPP method because measuring accuracy of MS-PSI depends on locating accuracy of illumination component rather than phase accuracy.

Supplementary Note S7: Optimal parameters determination of MS-PSI for dynamic measurement

The scale factor *s* and the number of unidirectional Fourier coefficient *c* are two key parameters for MS-PSI, which directly determines the measuring accuracy and efficiency in dynamic scenes. Therefore, we performed the comparative experiments and analysis to choose the optimal parameters for MS-PSI. As shown in Fig. S7(a), the standard ceramic plane with micro-level accuracy was placed at the back in the measured scene to evaluate the measured accuracy.

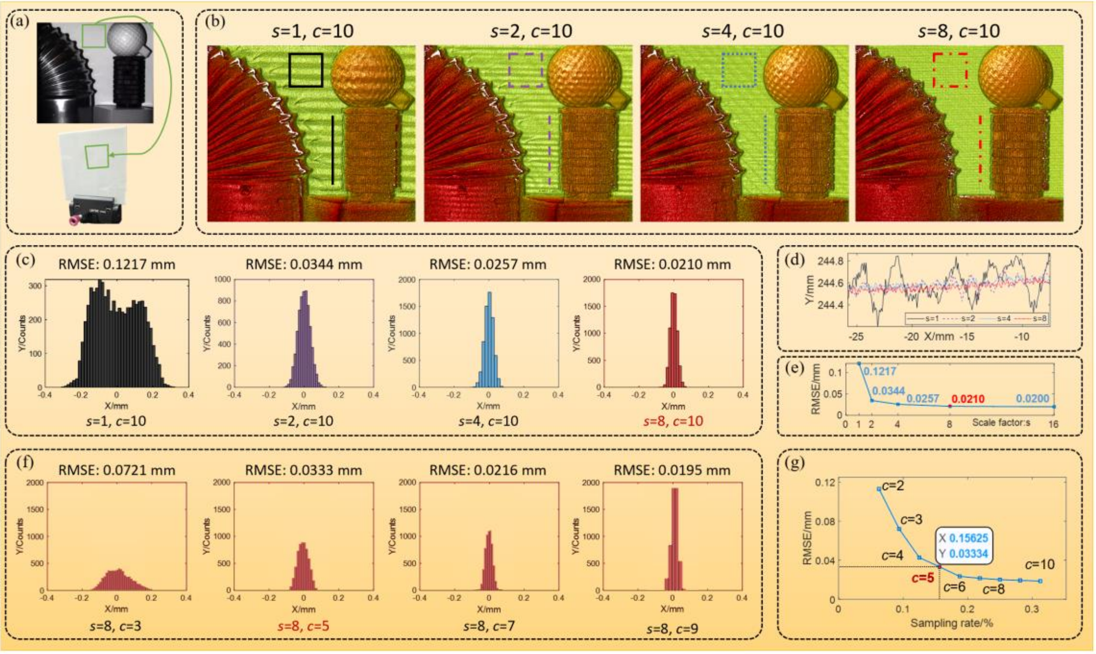

Figure S7. Optimal parameters determination of MS-PSI for dynamic measurement. a, Measured scene. **b**, Reconstructed result using different scale factor *s*. **c**, Error distribution in rectangle region of standard ceramic plane for results using different *s* when *c*=10. **d**, Error curve in the labeled line of standard ceramic plane for results using different *s* when *c*=10. **e**, Relationship between measured errors and varying scale factor *s* when *c*=10. **f**, Error distribution in rectangle region of standard ceramic plane for results using different *c* when *s*=8. **g**, Relationship between measured errors and varying Fourier coefficient number *c* when *s*=8.

In this experiment, the resolution of the projected patterns is 512×512 pixel and firstly, Fourier coefficient number *c* is fixed to 10 which means 10 groups of Fourier basis patterns for each slice direction are generated. And MS-PSI algorithm is performed with varying *s* form 1, 2, 4 to 8. The corresponding results are shown in Fig. S7(b), and it can be found that periodic error on ceramic plane is suppressed and data integrity on the edge of conduit is improved with the increasement of *s*. The root-meansquare error (RMSE) on the box area and height distribution in the labeling line of the results are further calculated as shown in Figs. S7(c) and S7(d), which quantitively shows the accuracy improvement. Results are consistent with the theoretical and simulated analysis. The RMSE is 0.0210 mm when s=8 and the measuring accuracy gets stable as shown in Figs. S7(e). So, s is determined as 8 in final dynamic experiment. To further choose optimal c to balance the measuring accuracy and speed in dynamic measurement, error distribution with varying c is shown in Fig. S7(f) and the relationship between RMSE and sampling rate (defined as the ratio between c and half width of the projected pattern's resolution) is shown in Fig. S7(g). It indicates measuring accuracy dramatically decreased when c is smaller than 5 and the RMSE is 0.0334 mm when c=5, which is satisfied for most measuring requirement. Therefore, c is determined as 5 to pursue higher measuring efficiency in dynamic scene, and the final measuring accuracy of depth measurement is 0.0334 mm in our experiments.

References

1 Zhang, Z., Ma, X. & Zhong, J. Single-pixel imaging by means of Fourier spectrum acquisition. *Nat Commun* **6**, 6225, doi:10.1038/ncomms7225 (2015).

2 O'Toole, M., Mather, J. & Kutulakos, K. N. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 3246-3253.

3 Scharstein, D. & Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int J Comput Vision* **47**, 7-42 (2002).

4 Jiang, H., Li, Y., Zhao, H., Li, X. & Xu, Y. Parallel Single-Pixel Imaging: A General Method for Direct–Global Separation and 3D Shape Reconstruction Under Strong Global Illumination. *Int J Comput Vision* **129**, 1060-1086 (2021).

5 Wu, Z., Guo, W. & Zhang, Q. Two-frequency phase-shifting method vs. Gray-coded-based method in dynamic fringe projection profilometry: A comparative review. *Opt Lasers Eng* **153**, 106995, doi:https://doi.org/10.1016/j.optlaseng.2022.106995 (2022).

6 Li, B., Karpinsky, N. & Zhang, S. Novel calibration method for structured-light system with an out-of-focus projector. *Appl Optics* **53**, 3415-3426 (2014).