Parallel Single-Pixel Imaging: A General Method for Direct–Global Separation and 3D Shape Reconstruction Under Strong Global Illumination

Hongzhi Jiang¹ · Yuxi Li¹ · Huijie Zhao¹ · Xudong Li¹ · Yang Xu¹

Received: 23 April 2020 / Accepted: 2 December 2020 / Published online: 6 January 2021
© The Author(s) 2021, corrected publication 2021

Abstract

We present parallel single-pixel imaging (PSI), a photography technique that captures light transport coefficients and enables the separation of direct and global illumination, to achieve 3D shape reconstruction under strong global illumination. PSI is achieved by extending single-pixel imaging (SI) to modern digital cameras. Each pixel on an imaging sensor is considered an independent unit that can obtain an image using the SI technique. The obtained images characterize the light transport behavior between pixels on the projector and the camera. However, the required number of SI illumination patterns generally becomes unacceptably large in practical situations. We introduce local region extension (LRE) method to accelerate the data acquisition of PSI. LRE perceives that the visible region of each camera pixel accounts for a local region. Thus, the number of detected unknowns is determined by local region area, which is extremely beneficial in terms of data acquisition efficiency. PSI possesses several properties and advantages. For instance, PSI captures the complete light transport coefficients between the projector–camera pair, without making specific assumptions on measured objects and without requiring special hardware and restrictions on the arrangement of the projector–camera pair. The perfect reconstruction property of LRE can be proven mathematically. The acquisition and reconstruction stages are straightforward and easy to implement in the existing projector–camera systems. These properties and advantages make PSI a general and sound theoretical model to decompose direct and global illuminations and perform 3D shape reconstruction under global illumination.

Keywords Light transport · Computational photography · Single-pixel imaging · Direct illumination · Global illumination · 3D shape reconstruction

Communicated by Yasutaka Furukawa.

Hongzhi Jiang and Yuxi Li contributed equally to this work.

✉ Huijie Zhao
buaa_zhaohj@163.com

Hongzhi Jiang
jhzi1862@buaa.edu.cn

Yuxi Li
yuxili@buaa.edu.cn

Xudong Li
xdli@buaa.edu.cn

Yang Xu
xuyang17@buaa.edu.cn

¹ Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, School of Instrumentation and Optoelectronic Engineering, Beihang University (BUAA), No. 37 Xueyuan Road, Haidian District, Beijing 100191, China

1 Introduction

The appearance of a scene is determined by the 3D geometry structure, the material properties, and the illumination conditions. Light transport equation is an effective way to describe the image formation process in computer vision and graphics. The radiance captured by a camera pixel is calculated by the weighted sum of intensities of every possible position on the light source. These weights, which are termed as *light transport coefficients*, contain all camera pixel and light source position combinations. Given the huge data volume represented by all camera pixel and light source position combinations, capturing light transport coefficients generally requires a long time. Light transport equation plays an important role in computer vision and graphics, and a large body of work, that is, from image-based relighting (Debevec et al. 2000; Masselus et al. 2003; Peers et al. 2009; Ren et al. 2015) to 3D reconstruction (Gupta et al. 2012; O’Toole et al.

2014; Chiba and Hashimoto 2017) and computational photography (Sen et al. 2005; Garg et al. 2006; Sen and Darabi 2009; O'Toole et al. 2012), is correlated with it.

The light received by a camera pixel in the light transport equation is a mixture component and can be further divided into two components, namely, *direct* and *global* (also termed as indirect). The former is due to the illumination that is directly reflected from the light sources, whereas the latter is due to the illumination caused by other points in the scene (Nayar et al. 2006). Figure 1 depicts these two components. The blue ray is the direct component that is precisely reflected from the light source and bounces only once before arriving at the camera pixel. The red ray is the global component or interreflection light, which is reflected from another point in the scene. Together, the direct and global components contribute to the final response of the camera pixel. Thus, the contribution of each component is mixed, and their decomposition by using modern cameras is challenging. However, separating these components is desirable because each component conveys different information about a scene. For instance, direct component provides a 3D geometrical structure of the scene, and the purest measurement of the material property of the scene point. Global component conveys complex optical interactions and is vital for photorealistic rendering (Nayar et al. 2006). Interreflections and subsurface scattering are two typical illustrations of global illumination, on which this study focuses.

In this study, we develop parallel single-pixel imaging (PSI) to capture the light transport coefficients and

decompose the direct and global components with off-the-shelf cameras and projectors. We achieve this goal by extending single-pixel imaging (SI) methods to modern cameras. In PSI, each pixel on a spatially resolved imaging sensor is considered an independent imaging unit that can obtain an image using the SI technique. Figure 2 shows the comparison of SI, modern digital camera, and PSI. The obtained image for each camera pixel (u, v) precisely corresponds to the light transport coefficients $h(u', v'; u, v)$, where (u', v') is a pixel on the controllable illumination source. The direct and global illuminations are completely decomposed by the SI theory, as proven in our previous work (Jiang et al. 2019). This study is based on the previous work, where Fourier-based SI (Zhang et al. 2015) is performed on each camera pixels to achieve 3D shape reconstruction under strong subsurface scattering. However, the data acquisition efficiency should be improved, because the number of projected patterns linearly depends on the resolution of the projector. Thus, the acquisition time can become unacceptable as the resolution of the projector increases. In this work, we refer to the method in this previous work as the naive SI.

Local region extension (LRE) method is proposed to accelerate the data acquisition of PSI. LRE method assumes that the visible region of each pixel in PSI is confined in a local region (Fig. 1). In “Appendix 1”, we confirm this assumption by providing an additional experiment. Thus, the number of projected patterns can be significantly reduced according to the size of the largest observable region among all pixels in the camera. LRE method is implemented by

Fig. 1 Schematic of the critical components of PSI. The blue ray corresponds to direct illumination, which is directly reflected from the light source and only bounce once before it is received by the camera pixel. The red ray refers to global illumination or interreflection, which occurs due to the illumination from another point in the scene. The intensities received by a given camera pixel from different projector pixels are presented. In this study, we propose PSI, where each

camera pixel is considered an independent unit. PSI can reconstruct the transport intensities of all camera and projector pixel combinations, which we refer to as light transport coefficients. In contrast to our previous work (Jiang et al. 2019), PSI significantly improves data acquisition efficiency through LRE method, in which the visible region (represented by the yellow rectangle) only accounts for a small region, thereby resulting in lesser number of unknowns to be detected

Fig. 2 SI, modern digital camera, and PSI. **(a)** SI. A series of patterns is projected onto the scene, and the scene can be reconstructed using the SI reconstruction algorithm. **(b)** Modern digital camera. An unstructured illumination source is adopted, and a camera is used to record the scene. **(c)** PSI. The projector and camera are adopted. The scene modulated by the patterns are recorded by the camera

three stages. In the first stage, the visible region of each pixel in the camera is localized by adaptive regional SI based on Fourier slice theorem (Jiang et al. 2017b). In the second stage, a series of periodic extension patterns is projected according to the localization information. In the last stage, the image corresponding to each pixel is obtained by using the proposed LRE reconstruction algorithm. The perfect reconstruction property of LRE for acceleration is proven mathematically.

From the perspective of mathematics and information theory, our overall analysis of LRE involves straightforward extensions of Nyquist–Shannon sampling theorem (Pharr et al. 2017) to frequency domain because PSI captures samples on the frequency domain. The LRE reconstruction theorem in Sect. 3.3 specifies the condition that has to be satisfied such that no information is lost when reducing the sampling rate in frequency domain. Thus, LRE reconstruction theorem can be understood as a dual form of Nyquist–Shannon sampling theorem in frequency domain. Projecting periodic extension patterns is precisely reducing the sampling rate in frequency domain. Provided that the visible regions are confined in local regions, down-sampling the frequency samples does not degrade the reconstructed images. The original signals can be perfectly reconstructed in theory when the projection of the periodic extension patterns is precisely performing down-sampling on the frequency domain. To our knowledge, PSI is the first attempt to capture signals from the perspective of this dual Nyquist–Shannon sampling theorem.

PSI is most closely related to primal–dual coding (O’Toole et al. 2012), dual photography (Sen et al. 2005), and symmetric photography (Garg et al. 2006). However, these methods either require special hardware and special projector–camera arrangement or a complex adaptive projection mode that makes the capturing process extremely time-consuming. On the contrary, PSI is a general and sound theoretical model for analyzing and capturing light transport. The “general” here means that PSI can

be implemented by any commercial projectors and cameras without any additional requirement. Furthermore, under the premise of sharing a common field of view, the arrangement of the projector–camera pair is arbitrary. Thus, a coaxial arrangement is unnecessary. The “sound” here indicates the perfect reconstruction property of LRE.

To illustrate the application of PSI, we consider the separation of direct and global illumination and 3D shape reconstruction under global illumination. If the rays bounce only once in the scene, triangulation (Hartley and Peter 1996) can be adopted to reconstruct 3D shape, but considerable error would be incurred if the rays bounce two times or more in the scene (Nayar et al. 1991). Several attempts have been made to solve this problem (Chen et al. 2008; Gupta and Nayar 2012; Gupta et al. 2013; O’Toole et al. 2014); however, each attempt makes specific assumptions and has particular limitations on real-world applications. In this work, PSI captures relatively complete data between projector–camera pair and makes no specific assumption on measured objects, thereby enabling PSI to work under several general situations in real-world applications. Meanwhile, PSI becomes a sound method because of its perfect reconstruction property. Thus, PSI is a general and sound theoretical model for decomposing direct and global illuminations, thereby achieving 3D shape reconstruction under strong global illumination. The global illumination in this study mainly refers to interreflections and subsurface scattering.

The contributions of this paper are presented as follows:

1. PSI is introduced to capture light transport coefficients without special hardwares or arrangement constraints for the existing projector–camera systems.
2. LRE method is proposed such that data acquisition efficiency is significantly improved in comparison with the naive SI method.
3. The perfect reconstruction property is proven mathematically for LRE method. The underlying principle is a

straightforward extension of Nyquist–Shannon sampling theorem to frequency domain.

4. PSI is a general and sound theoretical model for decomposing the direct and global illuminations and for performing 3D reconstruction under global illumination.

This work is organized as follows. Section 2 describes related work. Section 3 provides the theoretical foundations of PSI, the perfect reconstruction property of LRE method, and the implementation building blocks of LRE method. Section 4 describes the identification of direct and global illuminations. Section 5 shows the experimental results. Section 6 gives the conclusions and future work.

2 Related Work

2.1 SI Method

SI techniques take samples from the side of the light field by a single-pixel detector without spatial resolution (Ferri et al. 2010; Sun et al. 2012; Phillips et al. 2017; Zhang et al. 2017; Edgar et al. 2019). On the contrary, a 2D spatially resolved structured light illumination source is utilized by exploiting Helmholtz reciprocity (Sen et al. 2005; Zhang et al. 2015). Thus, lights from different positions on the light source can be separated in the nature of SI. This feature provides insights into PSI.

SI methods have been demonstrated to capture data in various domains, including studies on single-pixel cameras for multispectral (Bian et al. 2016) or hyperspectral (Hahn et al. 2014; Wang et al. 2016), infrared (Radwell et al. 2014), terahertz (Chan et al. 2008; Watts et al. 2014), 3D (Sun et al. 2013; Sun et al. 2016), and time domain imaging (Chen et al. 2014; Ryczkowski et al. 2016; Devaux et al. 2016). Although SI methods are originally designed for the cases when multipixel sensors are not preferable because of their cost or technological constraints, extending SI methods into pixel arrays have also been proposed. Gungor et al. (2018) introduced a compressive focal plane array imaging method that uses an augmented Lagrangian-based method to employ sparsity in Fourier and gradient spaces. This method yields high PSNR in short convergence time. Chen et al. (2015) used a focal plane array to achieve high spatial and temporal resolutions for short-wave infrared signals. However, these methods purely aim to improve the imaging efficiency for imaging a scene and are not suitable for analyzing light transport. Chen's prototype also required a calibration process to acquire the mapping from DMD mirror to sensor pixels, thereby limiting its applicability. The most outstanding feature of PSI is that it is designed to capture light transport coefficients, and the mapping from projector pixels to sensor

pixels is acquired by a localization stage, thereby simplifying its implementation.

2.2 Light Transport Capture

Generally, light transport behavior can be described by an 8D reflection function, which abstracts light path in terms of incident and outgoing directions in a bounding volume that surrounds the scene. Incident and outgoing fields are 4D (Peers et al. 2009). In turn, researchers attempt to capture the low dimensional subspaces of the 8D reflection field which is difficult to be directly captured. The 4D light transport matrix is such a reduced subspace captured by most studies. The light transport matrix in the reference paper and the light transport coefficients in this study are in fact different expressions of the same thing. Light transport coefficients can be understood as a 4D tensor. Given a camera pixel, the 2D projector coordinates form an image. If we linearize this image into a row vector, then it is precisely a row in the light transport matrix that corresponds to the given camera pixel. The expression of light transport coefficients is suitable for our derivation in PSI.

Debevec et al. (2000) used a light stage to sample a simplified 4D version of the reflection function. A light source was moved to a finite number of 2D positions, and a photograph was captured. Masselus et al. (2003) later used a projector–camera pair to capture a 6D slice of the full reflection field. The use of projector as light source provides flexibility in capturing light transport in different dimensions, and higher resolution on the light source side also becomes available. To improve the data acquisition efficiency, each bundle of light was assumed to be limited and only has local influence. Thus, distant regions do not influence one another and can be projected simultaneously. However, their method in determining which regions have no influence with each other was totally empirical. In PSI, we find that any pixel on the camera can only receive light from a local region on the projector. This local region is termed as the visible region for the corresponding camera pixel. The visible region information is obtained by the localization stage.

Sen et al. (2005) and Garg et al. (2006) introduced dual and symmetric photography to exploit physical properties of light transport, namely reciprocity and symmetry, to speedup reflectance field acquisition. However, these methods are totally adaptive, thereby placing most of the complexity on the acquisition system. Compressive sensing (CS) method was introduced to shift the complexity to the postprocessing stage, thereby resulting in a straightforward acquisition stage. Compressive light transport sensing method (Peers et al. 2009) that took advantage of interpixel coherency relations was proposed. However, their method required to choose a suitable compression basis before capture, and the available compression basis was limited because of the

dynamic range of the projector. Thus, the reconstructed light transport coefficients may contain errors because of the unsuitable basis function assumed for sparsity. Compressive dual photography (Sen and Darabi 2009) was introduced to overcome these drawbacks. By projecting Bernoulli patterns, the compression basis was chosen in the postprocessing stage, thus the form of basis is not limited by the hardware. However, the computational cost of CS methods is huge, thus only a coarse resolution on the projector side can be obtained. For example, compressive dual photography achieved a resolution of only a few hundred on the projector side with several hours of computation on a cluster. In contrast, PSI differs from CS fundamentally in principle. PSI reduces the number of patterns by locating the nonzero regions using LRE method and only the corresponding patches have to be imaged, while CS reduces the number of patterns by the sparsity assumption and the whole image has to be imaged. Thus, the scale of the light transport to be solved reduced dramatically by using LRE method, and light transport coefficients with much finer resolution can be easily obtained. Furthermore, one significant contribution of PSI is that the LRE reconstruction theorem provides the theoretical limit that has to be satisfied for perfect light transport reconstruction. However, CS methods have no such guarantee. Refer to “Appendix 3” for a comparison between PSI and CS method in terms of accuracy and reconstruction time on synthetic data.

Primal–dual coding (O’Toole et al. 2012) probed light transport by simultaneously controlling over the illumination and camera pixels by coaxial projector–camera pair. A follow-up work presents structured light transport (O’Toole et al. 2014) to probe light transport that works for arbitrary projector-camera set up. However, full light transport coefficients cannot be captured in their work. The main goal was to manipulate different components in light transport such that the ultimate captured image appeared as expected. In contrast, PSI is capable of efficiently capturing both the full light transport coefficients and 3D structure of the scene by using LRE method.

2.3 3D Reconstruction Under Global Illumination

Shape reconstruction plays an important role in many fields, such as industry, art, and medicine. However, global illumination, such as interreflections and subsurface scattering, incurs errors for most 3D reconstruction methods.

Nayar et al. (2006) found that high-frequency patterns are resistant to interreflections because the low-frequency interreflections can be nearly constant when high-frequency patterns are projected. Gupta et al. (2012) introduced micro phase shifting, which projected sinusoidal patterns with frequencies limited in a narrow and high frequency band. However, high-frequency illuminations reduce the contrast of the projected

patterns, thereby leading to higher errors especially when subsurface scattering dominates. Therefore, Gupta et al. (2013) took advantages of logical codes and combinatorial mathematics to handle multiple kinds of global illumination. However, the large width growths in their method also increases geometric errors for translucent objects. Moreover, all these methods are based on an important assumption that only low-frequency global illumination exists. The assumption will not hold when the interreflections occurs at highly specular or refractive surfaces.

Chen et al. (2007) developed polarization-difference imaging to reduce measurement error for translucent objects. They placed linear polarizers in front of the projector and the camera. Subsequently, they captured the images twice, once when the polarization state of the camera polarizer is parallel to that of the projector polarizer, and a second time orthogonal to the polarization state of the projector polarizer. The subsurface scattered light can be eliminated by differentiating the two captured images. However, this method assumed that direct illumination is not depolarized, which is not the case for diffuse objects. Regional projection (Jiang et al. 2017a), epipolar imaging (Zhao et al. 2018) and error compensation methods (Lutzke et al. 2011; Xu et al. 2019) were also developed to solve 3D reconstruction under global illumination. However, these methods either requires prior information for the scene and the materials or only function under some specific assumptions, thereby limiting their real-world applications.

PSI does not require prior information of the scene and materials. Although structured light transport (O’Toole et al. 2014) was introduced for 3D reconstruction under global illumination without prior information, PSI is more flexible since it captures the full light transport coefficients, and is capable of distinguishing direct illumination and global illumination when both components are very close to the epipolar line by determining the smallest speckle inside the threshold as direct illumination. However, the work of O’Toole et al. (2014) may fail to reconstruct high accuracy 3D data points in this case. Refer to “Appendix 4” for an additional experiment when *non-epipolar dominance assumption* is not met. In addition, surface reflection functions such as bidirectional surface scattering reflectance distribution function (BSSRDF) can also be obtained simultaneously in the future. Thus, PSI is a general and sound theoretical model for decomposing the direct and global illuminations and for performing 3D reconstruction under global illumination.

3 PSI: Capturing Light Transport Coefficients in High Efficiency

In this section, we introduce the foundations of PSI. First, we introduce light transport equation. Second, we consider the naive SI to show how SI method can be extended to modern

cameras for light transport coefficient capture. Third, we introduce the theoretical aspect of LRE method to accelerate data acquisition of PSI method. Last, LRE method is introduced from an implementation perspective.

3.1 Light Transport Equation

Light transport equation is an effective way to describe the image formation process in computer vision and graphics. The radiance $I(u, v)$ captured by a camera pixel (u, v) is expressed by light transport equation as

$$I(u, v) = O(u, v) + \sum_{v'=0}^{N-1} \sum_{u'=0}^{M-1} h(u', v'; u, v) P(u', v'), \quad (1)$$

where $P(u', v')$ expresses the outgoing radiance of a pixel (u', v') on the controllable illumination source, such as digital projector; $O(u, v)$ is the ambient illumination; and $h(u', v'; u, v)$ denotes the light transport coefficient between pixels (u, v) and (u', v') , which is the ratio of energy that can be received by (u, v) from (u', v') . M and N are horizontal and vertical resolution of the projector, respectively.

3.2 Naive SI: SI for Light Transport Coefficients Capture

If we consider each camera pixel as an independent unit for SI and project the required patterns, the reconstructed image of each camera pixel precisely represents the light transport coefficients. We show this conclusion by Fourier-based SI, where the four-step sinusoidal patterns required are

$$P_{\phi}(u', v'; k, l) = a + b \cos \left[2\pi \left(\frac{k \cdot u'}{M} + \frac{l \cdot v'}{N} \right) + \phi \right], \quad (2)$$

where (u', v') denotes a pixel on the projector; k and l denote discrete frequency samples, and take the values of $k = 0, 1 \dots M-1$ and $l = 0, 1 \dots N-1$. M and N are horizontal and vertical resolution of the projector, respectively. ϕ is the initial phase, which can take any values of $0, \pi/2, \pi$, and $3\pi/2$. a and b are the average intensity and contrast of the image, respectively.

According to Eq. (1), the captured intensity of a pixel (u, v) on the camera can be expressed as

$$I_{\phi}(u, v; k, l) = O(u, v) + \sum_{v'=0}^{N-1} \sum_{u'=0}^{M-1} h(u', v'; u, v) P_{\phi}(u', v'; k, l). \quad (3)$$

Fourier-based SI method captures samples in frequency domain. Each sample in the frequency domain is obtained by phase-shifting. Provided the captured intensities of camera pixels, the frequency samples are given by

$$H(k, l; u, v) = [I_0(u, v; k, l) - I_{\pi}(u, v; k, l)] + j[I_{\frac{\pi}{2}}(u, v; k, l) - I_{\frac{3\pi}{2}}(u, v; k, l)]. \quad (4)$$

The light transport coefficients can be calculated by Fourier-based SI reconstruction algorithm, which exactly corresponds to applying inverse discrete Fourier transform (IDFT) to the frequency samples, as given by

$$\begin{aligned} F^{-1}[H(k, l; u, v)] \\ = F^{-1} \left\{ \sum_{v'=0}^{N-1} \sum_{u'=0}^{M-1} 2b \cdot h(u', v'; u, v) \exp[-j \cdot 2\pi (\frac{k \cdot u'}{M} + \frac{l \cdot v'}{N})] \right\} \\ = 2b \cdot h(u', v'; u, v), \end{aligned} \quad (5)$$

where $F^{-1}[\cdot]$ is the 2D IDFT.

From Eq. (5), we conclude that by extending Fourier-based SI method to modern cameras, the light transport coefficients can be obtained by projecting sinusoidal patterns and applying Fourier-based SI reconstruction algorithm.

3.3 LRE Method with Perfect Reconstruction Property

The required number of patterns for the naive SI linearly increases with the resolution of the projector. For a typical projector with a resolution of 1920×1080 , millions of patterns are required, thereby making this approach unacceptable (Refer to “Appendix 5” for detailed information on the number of patterns required). However, any pixel on the camera can only receive light from a local region on the projector, making the visible region of each pixel confined in a local region. This property implies that the detected unknowns are determined by local region area, and LRE method takes advantage of this property to significantly improve acquisition efficiency. LRE method does not require the visible region can evenly divide the resolution of the projector. PSI is specially referred to the LRE-improved SI based light transport capture method.

In this subsection, we introduce the concept of the LRE method from theoretical aspect. We project periodic extension patterns to improve acquisition efficiency. LRE reconstruction theorem is introduced, which states that light transport coefficients can be perfectly reconstructed if the period of the periodic extension patterns covers the visible region. The fundamental idea is shown in Fig. 3.

3.3.1 Periodic Extension Patterns

For a certain camera pixel (u, v) , the observation that visible region is confined in local region can be mathematically expressed by that $h(u', v'; u, v)$, which is defined in region $\Omega = \{(u', v') | u' \in (0, M-1), v' \in (0, N-1)\}$, only has nonzero values confined in subregion

Fig. 3 Theoretical basics of PSI. For light transport coefficients where visible region is confined in a local region, periodic extension patterns with period that covers the visible region can be projected such that no aliasing occurs

$\Omega_s = \{(u', v') | u' \in (u_0, u_0 + M_s - 1), v' \in (v_0, v_0 + N_s - 1)\} \in \Omega$. Thus, the unknowns of the light transport coefficients significantly reduced from $M \times N$, which is determined by the resolution of the projector, to $M_s \times N_s$, which is determined by the visible region (LRE method does not require M and N can be exactly divided by M_s and N_s). From the mathematical perspective, $h(u', v'; u, v)$ can be obtained by far fewer acquisitions, implying that instead of projecting patterns with $M \times N$ freedom of degree, we only need to project patterns with $M_s \times N_s$ freedom of degrees. The required patterns, which we refer to as *basic patterns*, are expressed as

$$P_\phi^B(u', v'; k_s, l_s) = a + b \cos \left[2\pi \left(\frac{k_s \cdot u'}{M_s} + \frac{l_s \cdot v'}{N_s} \right) + \phi \right], \quad (6)$$

where (u', v') is a pixel on the basic pattern with size of $M_s \times N_s$, and take values of $u' = 0, 1 \dots M_s - 1$ and $v' = 0, 1 \dots N_s - 1$, respectively. k_s and l_s denote the discrete frequency samples, and take values of $k_s = 0, 1 \dots M_s - 1$ and $l_s = 0, 1 \dots N_s - 1$, respectively. However, given that the location of the visible region is arbitrary and varies for different camera pixels, we cannot decide which block of the projector is required to be projected. Periodic extension patterns that have full resolution of the projector should be projected

$$\begin{aligned} \tilde{P}_\phi(u', v'; k_s, l_s) &= \sum_{r_1=0}^{\lceil \frac{M}{M_s} \rceil} \sum_{r_2=0}^{\lceil \frac{N}{N_s} \rceil} P_\phi^B(u' - r_1 M_s, v' - r_2 N_s; k_s, l_s) \\ &= a + b \cos \left[2\pi \cdot \left(\frac{k_s \cdot u'}{M_s} + \frac{l_s \cdot v'}{N_s} \right) + \phi \right], \end{aligned} \quad (7)$$

where $\lceil \cdot \rceil$ denotes the ceiling function. u' and v' are the pixels on the projector, and take values of $u' = 0, 1 \dots M - 1$ and $v' = 0, 1 \dots N - 1$, respectively. r_1 and r_2 are integers. In the first equation, for $u' - r_1 M_s$ and $v' - r_2 N_s$ that beyond the domain of definition of the basic patterns, the corresponding pattern values are assumed to be zeros. The second equation holds because the value of pixel (u', v') in the periodic extension pattern is found in the basic pattern by rounding the pixel index into the region of $\{(u', v') | u' \in (0, M_s - 1), v' \in (0, N_s - 1)\}$ and by taking advantage of 2π period of cosine functions. We refer to the period of the periodic extension patterns generated by Eq. (7) as $M_s \times N_s$.

3.3.2 Perfect Reconstruction of Light Transport Coefficients

In this subsection, we introduce the perfect reconstruction property of LRE. First, we provide Lemma 1 which states that the reconstructed image by Fourier-based SI reconstruction algorithm corresponds to a periodic extension version of the light transport coefficients $h(u', v'; u, v)$ when the periodic extension patterns generated in previous subsection are projected.

Lemma 1 Assume $h(u', v'; u, v)$ is the light transport coefficient between camera pixel (u, v) and projector pixel (u', v') , by projecting periodic extension patterns in the form of Eq. (7), the reconstructed image of camera pixel (u, v) by Fourier-based SI reconstruction algorithm becomes a periodic extension version of the original light transport coefficients

$$\tilde{h}_r(u'_r, v'_r; u, v) = 2b \cdot \sum_{r_1=-\infty}^{+\infty} \sum_{r_2=-\infty}^{+\infty} h(u'_r - r_1 M_s, v'_r - r_2 N_s; u, v), \quad (8)$$

where (u'_r, v'_r) is a pixel on the reconstructed image, and r_1 and r_2 are integers.

Proof Proof of Lemma 1 can be found in “Appendix 2”.

LRE Reconstruction Theorem *If the period $M_s \times N_s$ of the projected periodic extension patterns covers the visible region, light transport coefficients can be perfectly reconstructed by adopting Fourier-based SI reconstruction algorithm, that is, the light transport coefficients obtained by reconstruction is exactly equal to the light transport coefficients reconstructed by the naive SI.*

Proof Proof of LRE Reconstruction Theorem can be found in “Appendix 2”.

LRE reconstruction theorem specifies the condition that has to be satisfied such that no information is lost (aliasing does not occur, refer to Fig. 3) when reducing the sampling rate in frequency domain. Thus, LRE reconstruction theorem can be understood as a dual form of Nyquist–Shannon sampling theorem in frequency domain. Projecting periodic extension patterns is precisely reducing the sampling rate in frequency domain. From Eqs. (2) and (7), when M and N can be exactly divided by M_s and N_s respectively, periodic extension patterns are precisely down-sampling in the frequency domain of the naive SI, because the patterns generated by Eq. (7) are precisely the patterns generated by Eq. (2) at fixed frequency intervals, with step size of M/M_s and N/N_s for k and l , respectively. The proposed LRE method for acceleration in this study takes advantage of the LRE reconstruction theorem, thus the light transport coefficients can be perfectly reconstructed.

From previous deduction, three building blocks are necessary for LRE method. First, the visible region Ω_s of each pixel is necessary such that the minimum information required is available. Second, instead of projecting patterns with $M \times N$ freedom of degree, we project periodic extension patterns with $M_s \times N_s$ freedom of degrees. Lastly, a reconstruction algorithm is necessary such that the periodic extension version of the original function $\tilde{h}_r(u', v'; u, v)$ is appropriately processed to reconstruct the original function $h(u', v'; u, v)$. In the next subsection, we introduce each of these building blocks from the implementation perspective.

3.4 Implementation of LRE Method

In this subsection, we introduce the LRE method from the implementation perspective. LRE is implemented by three stages:

Stage 1: Adaptive regional SI (Jiang et al. 2017b) based on Fourier slice theorem is utilized to obtain the visible region location of each pixel.

Stage 2: Periodic extension patterns are projected, and the scene illuminated by these patterns is simultaneously photographed by a camera.

Stage 3: LRE reconstruction algorithm is applied to obtain the light transport coefficients in the projector–camera pair.

The overall pipeline of LRE method is shown in Fig. 4.

3.4.1 Localization by Adaptive Regional SI

Fourier slice theorem, which states that the 1D scan of an image in the frequency domain along any orientation equals to the Fourier transform of the 1D projection function along that orientation in the spatial domain, is applied for visible region localization of any pixel in the detector array. Refer to (Jiang et al. 2017b) for additional details about the use of Fourier slice theorem in SI. This process is illustrated in Fig. 4a.

The patterns of

$$\begin{aligned} P_\phi^V(u', v'; k) &= P_\phi(u', v'; k, 0) \\ &= a + b \cos \left(2\pi \cdot \frac{k \cdot u'}{M} + \phi \right), \end{aligned} \quad (9)$$

and

$$\begin{aligned} P_\phi^H(u', v'; l) &= P_\phi(u', v'; 0, l) \\ &= a + b \cos \left(2\pi \cdot \frac{l \cdot v'}{N} + \phi \right), \end{aligned} \quad (10)$$

which are vertical and horizontal patterns, are projected to scan two slices along the u' , v' axes. u' and v' take values of $u' = 0, 1 \dots M-1$ and $v' = 0, 1 \dots N-1$, respectively. ϕ take values of $0, \pi/2, \pi$, and $3\pi/2$. The camera is photographing when the patterns are projected. Similar to Eq. (4), two slices of frequency samples $H^V(k; u, v)$ and $H^H(l; u, v)$ can be obtained. Then, we can calculate $h^H(u'; u, v)$ and $h^V(v'; u, v)$, which represents the projection function along the horizontal and vertical axes respectively, by applying 1D IDFT on $H^V(k; u, v)$ and $H^H(l; u, v)$.

The visible region of each pixel in the projector coordinate system can be determined by finding the areas where

For camera pixel (u, v)

Fig. 4 LRE implementation pipeline. (a) In stage 1, the visible region is located by adaptive regional SI based on Fourier slice theorem. Vertical and horizontal patterns are projected for the horizontal and vertical slices of the light transport coefficients, respectively. (b) In stage 2, the period of the periodic extension patterns is determined by

the localization information, and the periodic extension patterns are generated. (c) In stage 3, the light transport coefficients are obtained by periodically extending the reconstructed patch and preserving the correct region according to the localization information

the values of $h^H(u'; u, v)$ and $h^V(v'; u, v)$ are greater than the noise threshold. A rectangle area can be located by using two orthogonal slices.

3.4.2 Projecting the Periodic Extension Patterns and Simultaneously Photographing the Scene Illuminated by These Patterns with Camera

In this stage, we project periodic extension patterns with periods of M_s and N_s on the horizontal and vertical orientations. Aliasing will not occur if the size of $M_s \times N_s$ covers all nonzero values of the light transport coefficients of any pixel in the camera. Thus, M_s and N_s can be determined by the largest horizontal and vertical visible lengths among all pixels in the camera. Suppose $M_l(u, v)$ and $N_l(u, v)$ denote the horizontal and vertical visible lengths of the pixel whose position is (u, v) in the detector, M_s and N_s are given by

$$M_s = (1 + \eta_M) \cdot \max_{(u,v)} [M_l(u, v)], \quad (11)$$

$$N_s = (1 + \eta_N) \cdot \max_{(u,v)} [N_l(u, v)], \quad (12)$$

where M_s and N_s are the horizontal and vertical periods of the projected patterns, $\eta_M \geq 0$ and $\eta_N \geq 0$ are set for appropriate margins, and 0.1 to 0.2 are typical values. In this study, we set η_M and η_N to 0.1.

$M_l(u, v)$ and $N_l(u, v)$ can be determined from the projection function in the first stage. A noise threshold is set, and $M_l(u, v)$ or $N_l(u, v)$ can be determined by the range greater than the noise threshold [Fig. 4b]. The center point $B(u, v)$ of this range can be obtained for each pixel in the camera for later use. $B_M(u, v)$ is denoted as the u' axis component of $B(u, v)$, and $B_N(u, v)$ is denoted as the v' axis component of $B(u, v)$.

The projected patterns $\tilde{P}_\phi(u', v'; k_s, l_s)$ are then generated by Eq. (7). The camera is used to simultaneously receive the radiance from the scene when the scene is illuminated.

3.4.3 LRE Reconstruction Algorithm

LRE reconstruction algorithm is separately processed for each pixel on the camera and consists two substeps, namely, (1) application of Fourier-based SI reconstruction algorithm and (2) preservation of the actual visible region according to the localization information in stage 1 by setting nonvisible regions to zero. These two steps should be performed on each pixel in the camera.

- (1) Application of Fourier-based SI reconstruction algorithm

Application of Fourier-based SI reconstruction algorithm is exactly the application of 2D IDFT. Before applying 2D IDFT, the intensities captured by the camera are rearranged upon completion of the previous two stages, as expressed in Eq. (4). Samples in the frequency domain $H(k_s, l_s; u, v)$ for each pixel in the camera can be obtained. After the acquisition of the required samples in the frequency domain, the reconstructed images patches $h_r^B(u', v'; u, v)$ are calculated by 2D IDFT. The calculation can be efficiently completed by applying inverse fast Fourier transform (IFFT).

- (2) Preservation of the actual visible region

Since IFFT returns the first period of the periodic extension function $\tilde{h}_r(u', v'; u, v)$, we must obtain the periodic extension function $\tilde{h}_r(u', v'; u, v)$ from $h_r^B(u', v'; u, v)$ using

$$\tilde{h}_r(u', v'; u, v) = \sum_{r_1=0}^{\lceil \frac{M}{M_s} \rceil} \sum_{r_2=0}^{\lceil \frac{N}{N_s} \rceil} h_r^B(u' - r_1 M_s, v' - r_2 N_s; u, v), \quad (13)$$

where $\lceil \cdot \rceil$ denotes the ceiling function. u' and v' take values of $u' = 0, 1 \dots M - 1$ and $v' = 0, 1 \dots N - 1$, respectively. r_1 and r_2 are integers. This process is shown in Fig. 4c. For $u' - r_1 M_s$ and $v' - r_2 N_s$ that beyond the region $\{(u', v') | u' \in (0, M_s - 1), v' \in (0, N_s - 1)\}$, the corresponding values of $h_r^B(u', v'; u, v)$ are assumed to be zeros.

Only the actual region of $\tilde{h}_r(u', v'; u, v)$ should be preserved to obtain the light transport coefficients $h_r(u', v'; u, v)$. The preservation process is accomplished by using the localization information obtained from the previous stages. A rectangle region can be identified by the horizontal and vertical slices obtained in stage 1. The center of the rectangle region is determined by the center point $B(u, v)$ in the second stage. The rectangle region can then be found by setting its side lengths as M_s and N_s with respect to the center point.

Let $\Omega_r(u, v)$ denote the rectangle region of the pixel on the position of (u, v) in the camera, $\Omega_r(u, v)$ is given by

$$\Omega_r(u, v) = \{(u', v') | B_M(u, v) - \lfloor M_s/2 \rfloor \leq u' < B_M(u, v) + \lceil M_s/2 \rceil, \\ B_N(u, v) - \lfloor N_s/2 \rfloor \leq v' < B_N(u, v) + \lceil N_s/2 \rceil\}, \quad (14)$$

where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ denote the flooring and ceiling function, respectively.

The light transport coefficients can be calculated by the element-wise product between $\tilde{h}_r(u', v'; u, v)$ and a mask $M(u', v'; u, v)$, which is 1 within $\Omega_r(u, v)$ and 0 otherwise.

$$h_r(u', v'; u, v) = \tilde{h}_r(u', v'; u, v) \cdot M(u', v'; u, v), \quad (15)$$

where $\cdot$ represents the element-wise product, and $M(u', v'; u, v)$ has the form of

$$M(u', v'; u, v) = \begin{cases} 1, & (u', v') \in \Omega_r \\ 0, & (u', v') \notin \Omega_r. \end{cases} \quad (16)$$

This process is depicted in Fig. 4c.

4 Identification of Direct and Global Illumination

PSI reconstructs the light transport coefficients $h(u', v'; u, v)$ for each pixel (u, v) in the camera. For a specific location (u, v) in the camera, only one or several discrete regions

called speckles on the light transport coefficients have high intensities [Fig. 5a]. These speckles are the results of PSI and located on projector coordinate. Each speckle records the amount of light that can be received by the camera pixel from the corresponding projector locations. Because direct and global components generally come from different pixels on the projector, they are already separated by PSI. The only work remained is to identify each speckle to be direct or global. Before identification, we used depth-first search algorithm (Tarjan 1972) to determine the 8-connected projector pixels inside each speckle.

In this section, we use a method with three stages to identify whether a speckle corresponds to direct or global illumination. We assume that the extrinsic and intrinsic parameters and distortion coefficients have been estimated by calibration. Refer to Zhang (1999) for additional information about

Fig. 5 Pipeline for identification of direct and global illumination. The light transport coefficients of a camera pixel (u, v) is shown. (a) In stage 1, the representing point for each speckle is determined, and the epipolar line is computed. (b) In stage 2, the direct illumination

point is determined. (c) In stage 3, the direct and global components are determined. The intensity value of the direct and global components of each camera pixel are calculated by summing the direct and global components in the light transport coefficients, respectively

calibration. The direct and global illumination can be distinguished in the following stages:

Stage 1: Obtaining the epipolar line. In this stage, we determine a representing point for each speckle and compute the epipolar line.

Stage 2: Determination of the direct illumination point. The direct illumination point is determined by finding the nearest representing point to the epipolar line.

Stage 3: Calculation of the direct and global illumination intensities. We separate the direct and global illumination components in the light transport coefficients and sum up these two components separately to produce intensities of the corresponding camera pixel.

The overall pipeline is shown in Fig. 5.

4.1 Obtaining the Epipolar Line

For a camera pixel (u, v) , the components of direct and global illumination are identified by first calculating the epipolar line $\mathbf{l}'(u, v)$ and then distinguishing the direct and global illumination by calculating the distance of each speckle to the epipolar line. To compute such distance, a representing point for each speckle is required. We use the position $(u'_c, v'_c)_i$ which has the highest value as the representing point for the i -th speckle. The epipolar line can be easily computed by a matrix multiplication process (Zhao et al. 2018). Distortion correction should be simultaneously applied on the camera and projector images to obtain accurate results. Refer to (Park et al. 2009) for additional information about distortion correction.

4.2 Determination of the Direct Illumination Point

After computation of the epipolar line, we can determine direct illumination by the link between light transport and stereo geometry, which states that direct illumination lies on the epipolar line and global illumination does not (O'Toole et al. 2014). This is determined by calculating the distance between each representing point $(u'_c, v'_c)_i$ and the epipolar line $\mathbf{l}'(u, v)$ and denote this distance by d'_i . The direct illumination point is determined by finding the nearest point to the epipolar line among the representing points within a predefined threshold [Fig. 5b]. This process can be expressed as

$$(u'_d, v'_d) = \arg \min_i d'_i \quad (17)$$

$$s.t. d' \leq \epsilon,$$

where (u'_d, v'_d) denotes the direct illumination point, and ϵ is the predefined threshold. We set it to 3 pixels. $\arg \min$ returns the point in the feasible region whose distance d'_i is minimum.

Equation (17) implies that we choose the representing point with the smallest distance as the direct illumination point if the distance smaller than a threshold. We skip the corresponding camera pixel which has no direct illumination point if the nearest representing point is outside the threshold.

When the non-epipolar dominance assumption (O'Toole et al. 2014), which states that epipolar element only contributes to direct illumination, is not met, Eq. (17) can be changed slightly to determine the representing point of the smallest speckle (the speckle with the minimum number of pixels) as the direct illumination point to make sure PSI works faithfully in this situation (Refer to "Appendix 4" for an additional experiment when non-epipolar dominance assumption is not met).

4.3 Calculation of the Direct and Global Illumination Intensities

Given the direct illumination point, the speckle from which the direct illumination point is chosen contribute to direct illumination. However, in practice, the direct illumination speckle may turn out to be large, consider subsurface scattering where a large speckle is incurred. Thus, we should only include a small neighboring region of the direct illumination point as the final direct illumination, the remaining part is considered as global illumination component.

The obtained direct illumination component $h_r^d(u', v'; u, v)$ is given by: [shown in Fig. 5c]

$$h_r^d(u', v'; u, v) = h_r(u', v'; u, v) \cdot M_d(u', v'; u, v), \quad (18)$$

where $\cdot$ represents the element-wise product, and $M_d(u', v'; u, v)$ is a mask matrix which has the form of

$$M_d(u', v'; u, v) = \begin{cases} 1, & (u', v') \in \Omega_d(u, v) \\ 0, & (u', v') \notin \Omega_d(u, v), \end{cases} \quad (19)$$

where $\Omega_d(u, v)$ corresponds to the region inside a radius threshold centered on the direct illumination point. We specify the radius threshold as two pixels.

The global illumination component $h_r^g(u', v'; u, v)$ in the light transport coefficients is required [shown in Fig. 5c] to obtain the global component image, and the equation is given by

$$h_r^g(u', v'; u, v) = h_r(u', v'; u, v) \cdot M_g(u', v'; u, v), \quad (20)$$

where $\cdot$ represents the element-wise product, and $M_g(u', v'; u, v)$ is given by

$$M_g(u', v'; u, v) = \mathbf{1} - M_d(u', v'; u, v), \quad (21)$$

where $\mathbf{1}$ is a matrix that has the same size as $M_d(u', v'; u, v)$ whose all elements are 1.

Now that the direct and global components in the light transport coefficients are identified, we can calculate the intensity values of the direct and global components for all camera pixels. For each camera pixel, the intensity values for the two components are calculated by summing the direct and global components of the light transport coefficients, respectively.

The intensity value of the direct component for camera pixel (u, v) is calculated by

$$I_d(u, v) = \sum_{v'=0}^{N-1} \sum_{u'=0}^{M-1} h_r^d(u', v'; u, v), \quad (22)$$

where $I_d(u, v)$ denotes the resulting direct illumination component at camera location (u, v) .

Then, we sum up the global component $h_r^g(u', v'; u, v)$ to obtain the global illumination intensity as follows

$$I_g(u, v) = \sum_{v'=0}^{N-1} \sum_{u'=0}^{M-1} h_r^g(u', v'; u, v), \quad (23)$$

where $I_g(u, v)$ denotes the resulting global illumination component at camera location (u, v) .

5 Results and Evaluations

A diagram of the proposed PSI technique is shown in Fig. 2c, which comprised a computer, a digital projector, a camera, and the scene to be imaged. Critical components of PSI are shown in Fig. 1. The computer generated the projected patterns and implemented the PSI method after all the required images were taken. In this study, four-step phase-shifting sinusoidal patterns with 1920×1080 pixels were generated by the computer. The illumination patterns had 256 grayscale levels and were successively projected by a digital projector. The projector consisted of a lens, a digital micromirror device (DMD), and a light-emitting diode (LED) light source. We used a 256-grayscale-level camera with 1920×1200 pixels to collect the resulting light. The projector and the camera were synchronized such that the projector switched the illumination patterns after the camera finished recording. The camera, which was used to record the scene illuminated by the projected pattern, was composed of a lens and an imaging sensor. The capturing frame rate was 120 frames per second (fps).

Two experiments are conducted in this section. First, a scene composed of five objects with various material properties is captured by using the PSI method. The reconstructed light transport coefficients for eight typical points are depicted, the separation results of direct and global illumination components are displayed, and the reconstructed 3D

shape of the scene is shown. Second, the accuracy of the 3D reconstruction results by using the PSI method is evaluated.

5.1 Capturing the Compound Scene by PSI

In this section, we used PSI to capture a scene composed by five objects, including an onion and a white gourd in which subsurface scattering occurs, and an impeller and a metal part in which strong interreflections occur. A gypsum toy bear in which global illumination is not significant was also included for comparison. The images of the investigated objects are shown in Fig. 6. The reconstructed light transport coefficients for eight typical points are depicted, the separation results of direct and global illumination components are displayed, and the reconstructed 3D shape of the scene is shown.

In this experiment, the resolution of the projector was 1920×1080 . Thus, a total of $1920 + 1080$ coefficients were required for the localization of vertical and horizontal orientations, respectively. After the localization stage, we found that 160×160 region on the projector was enough to cover the largest visible region, which required 25,600 Fourier coefficients to reconstruct the full light transport coefficients for each pixel on the camera. The Fourier spectrum of a real-valued signal was *conjugate symmetric*; thus, approximately half of the coefficients were redundant. A total of 14,304 Fourier coefficients were required to reconstruct the

Fig. 6 Investigated objects. (a) Gypsum toy bear, which was included for comparison. (b) Onion in which subsurface scattering occurs. (c) White gourd in which subsurface scattering occurs. (d) Impeller in which interreflections occur. (e) Metal part in which interreflections occur

2D image of light transport coefficients for each pixel on the camera when PSI is applied. However, for the naive SI projecting scheme, the number of required Fourier coefficients was 1,036,802. Thus, the proposed PSI method provided more than 70-fold of improvement on data acquisition efficiency in this experiment (Refer to “Appendix 5” for detailed information on how to calculate the number of Fourier coefficients and the number of patterns required).

5.1.1 PSI for Light Transport Coefficients

Our first goal in this experiment was to implement PSI to reconstruct light transport coefficients for each pixel on the camera. Given the large data volume of the light transport coefficients, we could not display all of them in this paper. Thus, Fig. 7b–i show eight typical points on the scene that were chosen to illustrate the reconstructed light transport coefficients. From these figures, the relationship between the light transport coefficients and material properties of objects was identified easily. For opaque and diffuse reflection objects, the reconstructed light transport coefficients had high intensities confined within a small region. Figure 7b correspond to point on the head of toy bear, where global illumination is not significant. On the other hand, the shape of the object also had effects on light transport coefficients. Figure 7c corresponds to a point on the ear of toy bear, and the concave structure resulted in larger influence area in the light transport coefficients. For translucent objects, the light transport tended to have a large influence area due to subsurface scattering. Figure 7d and e correspond to the points on the onion and the white gourd, respectively. For opaque objects with specular or glossy reflection, light transport coefficients may have influence over several distinct regions due to interreflections. Figure 7f–i show the reconstructed light transport coefficients of the positions on the industrial products made by metal. From this experiment, we conclude that different kinds of global illumination dominated in this scene.

5.1.2 Separation of Direct and Global Illuminations

Our second goal in this experiment was to distinguish direct and global illuminations using the crucial light transport–stereo geometry link, which states that direct illumination lies on the epipolar line and global illumination does not (O’Toole et al. 2014). Results are shown in Fig. 8 in which the captured and the generated images are multiplied by a factor of 3 for visual clarity. This experiment illustrates that PSI is a general model for decomposing direct and global illumination.

Figure 8b–d show that the global illumination of the toy bear accounted for a small proportion, whereas the intensity of the direct illumination of the onion and the white gourd was in a relatively low level. This result is consistent with

our knowledge that the toy bear barely presents global illumination, whereas the onion and the white gourd present strong global illumination due to subsurface scattering. One interesting observation is that the bear claw was mirrored on the impeller due to specular reflection (Row e in Fig. 8). PSI could distinguish this mirrored image from the direct component. Figure 8f shows the detailed separation result of the impeller. The highlight that occurs on the upper blade turned out to be the specular reflection from other blades. In Fig. 8g, the interreflections caused by the specular reflection of the metal part can be also distinguished. Attention should be focused on the bottom face of the metal part. Most of the light on the bottom face was indeed reflected from the back face. The mirrored image on the bottom face was correctly identified as global illumination.

5.1.3 3D Shape Reconstruction Under Strong Global Illumination

Our last goal in this experiment was to achieve shape reconstruction under strong global illumination by PSI. This experiment is a typical illustration of PSI as a general model for performing 3D reconstruction under global illumination.

Figure 9a shows the entire scene illuminated by sinusoidal fringe pattern to illustrate the impact of global illumination phenomenon for 3D shape reconstruction by fringe projection profilometry. Figure 9c–h are the enlarged figures that correspond to the positions indicated in Fig. 9a. The toy bear is presented for comparison. In this experiment, we obtained the 3D shape of the five objects by triangulation (Hartley and Peter 1996). Given that the PSI obtained the light transport coefficients between each pair of camera and projector pixels at the resolution of both devices, we can easily obtain the matching point pairs required by triangulation after the determination of direct illumination (Jiang et al. 2019). Figure 9b depicts the reconstructed 3D shape. Figure 9i–n show the enlarged details of the reconstructed results; the rectangles indicate the positions.

5.2 Accuracy Evaluation of PSI

To illustrate the accuracy of PSI, we initially compare the difference of the 3D points reconstructed by the naive SI and the PSI method. Subsequently, we analyzed the effect of the period size of the periodic extension patterns on accuracy.

5.2.1 Naive SI and PSI Method

We used a jade horse [Fig. 10a] as the investigated object in this subsection. Strong subsurface scattering occurs in the jade horse. Since the required number of patterns for the naive SI becomes unacceptably large for the resolution of projector in our experiment, we grouped the resolution of

Fig. 7 Light transport coefficients acquired by PSI. **(a)** Investigated scene. This scene consists of five objects, which are indicated by corresponding Arabic numerals. The positions where we presented the reconstructed light transport coefficients are marked by red circles with a letter inside. The letter indicates which subfigure is responsible for the corresponding position. **(b–i)** Reconstructed light transport coefficients by PSI. The projector coordinates are shown in each enlarged figure. The position of the camera pixel to reconstruct the corresponding light transport is marked on the upper right corner of each subfigure. The reconstructed image is rescaled for visual clarity. The multiplied factor is also marked on the upper right corner of each subfigure

(a)

(b)

- 1: Toy bear (global illumination is not significant)
 2: Onion (subsurface scattering)
 3: White gourd (subsurface scattering)
 4: Impeller (strong interreflections)
 5: Metal part (strong interreflections)

- (b)** Subfigure b **(c)** Subfigure c
(d) Subfigure d **(e)** Subfigure e
(f) Subfigure f **(g)** Subfigure g
(h) Subfigure h **(i)** Subfigure i

(c)

(d)

(e)

(f)

(g)

(h)

(i)

Fig. 8 Separation of direct and global illumination by PSI. The first column corresponds to the original captured image. The second and third columns correspond to direct and global components separated by PSI. The fourth column corresponds to the sum of the second and third columns or also known as mixed illumination. (a) Entire scene. (b) The zoom-in region of toy bear. Most energy are concentrated in the direct component. (c) Onion presents strong global illumination

due to subsurface scattering. The glossy reflection at the center of the onion is classified as direct illumination, which is reasonable. (d) White gourd presents a strong global illumination due to subsurface scattering. (e) Bear claw is mirrored on the impeller due to specular reflection and can be distinguished by PSI. (f) Detailed result of the impeller. (g) Zoom-in region of the metal part. Interreflections caused by specular reflection can be separated

the projector to be 192×108 . We used an interpolation mode for grouping in the projector side to achieve smooth results. First, we generate the required patterns in low resolution. Second, we normalize the actual resolution of the projector into the low resolution. Finally, the values of the patterns projected are interpolated from the patterns with low resolution. For sinusoidal patterns, the interpolated values can be directly calculated from the sine functions. This experiment illustrates that PSI is a sound model for performing 3D reconstruction under global illumination

The required size to cover the visible region was 10×10 . Thus, the required number of Fourier samples for the naive SI and PSI were 10,370 and 204, respectively. PSI provides approximately 50-fold of improvement in data acquisition efficiency in this experiment. The reconstructed 3D points of each method are given in Fig. 10b and c, and the error between them is shown in Fig. 10d. We also calculated the mean absolute error and the root mean square error between naive SI and PSI method. The mean square error is 0.014 mm, and the root mean square error is 0.021 mm. The error

Fig. 9 3D shape reconstruction when global illumination is dominant. **(a)** Reconstructed scene illuminated by sinusoidal fringe pattern. **(b)** 3D shape reconstructed by PSI. **(c–h)** Enlarged figures for the original scene illuminated by sinusoidal fringe pattern. The toy bear is shown for comparison. Except for subfigure (c), the stripes on each

of subfigure show different degrees of blur, especially in subfigures (d) and (e), the fringes are nearly invisible due to subsurface scattering. Besides, in subfigures (g) and (h), fringes are unclear in specific regions due to the strong global illumination presented in the scene. **(i–n)** Enlarged details of the reconstructed 3D shape

Fig. 10 Comparison between naive SI and PSI. **(a)** Jade horse as the investigated object. **(b)** Reconstructed 3D points by the naive SI. **(c)** Reconstructed 3D points by PSI. **(d)** Error of the 3D points between naive SI and PSI

Table 1 Results of naive SI and PSI

Acquisition method	Number of Fourier samples	Acquisition time (s)	Improvement (times)
Naive SI	10,370	350	
PSI method	204	7	50

difference is insignificant, thereby illustrating the perfect reconstruction property of LRE. However, considering the huge benefit in data acquisition efficiency provided by PSI method, PSI makes it possible for SI-based methods to perform 3D reconstruction under global illumination within an acceptable time. Refer to Table 1 for detailed acquisition time of the naive SI and PSI method in this experiment.

5.2.2 Effect of the Period Size of Periodic Extension Patterns on Accuracy

In this subsection, we show the effect of the period size of the periodic extension patterns on accuracy. We used the jade horse in the previous subsection as the investigated object. The resolution of the projector was 1920×1080 . From the localization stage, a period of 80×80 was enough to cover the visible region. We also used periodic extension patterns with a period of 10×10 and 40×40 to test the effect of using patterns with smaller periods on accuracy. To evaluate the accuracy, the reference 3D shape is obtained by powdering

the jade horse statue. The error maps are depicted in Fig. 11, and the results are summarized in Table 2. The period of the periodic extension patterns in Fig. 11a was not enough to cover the visible region, thus large error occurs. If the period size is enough to cover the visible region, for instance, the situation when period size of 80×80 is used [Fig. 11c], then the reconstructed result has high accuracy. Projecting the patterns with period of 40×40 is a trade-off between acquisition time and accuracy [Fig. 11b], where approximately twofold of improvement in data acquisition efficiency can be obtained with a slightly degraded accuracy than the situation when the period size of 80×80 is used.

6 Conclusions

In this study, PSI extends the traditional SI method to a modern camera. In PSI, each pixel on the camera is considered an independent imaging unit that can simultaneously obtain an image using the SI technique. On this basis, PSI can solve complex problems that neither SI nor modern cameras could solve before. The data captured by PSI are light transport coefficients, which are important in computer vision and graphics. To improve the efficiency of data acquisition in PSI, we introduce LRE method, and the perfect reconstruction property of LRE method can be proven mathematically. We illustrate the application of PSI by considering the separation of direct and global illumination and 3D shape reconstruction under global illumination.

Fig. 11 Effect of the period size of periodic extension patterns on accuracy. In this figure, the first and second rows present the reconstructed 3D shape by the corresponding period size and the error map, respectively. Results when the period sizes are (a) 10×10 , (b) 40×40 , and (c) 80×80 . (d) Reference 3D shape is obtained when the jade horse is coated with powders. The bottom right figure is the image of the jade horse

Table 2 Summarized results of PSI method using various period sizes

Period size (pixels)	Number of Fourier samples	Acquisition time (s)	Mean absolute error (mm)	Root mean square error (mm)
10×10	1554	52	0.218	0.268
40×40	2304	78	0.087	0.117
80×80	4704	160	0.057	0.077

Several advantages and properties of PSI are summarized as follows:

- PSI is a general method for capturing light transport coefficients, without requiring special hardware or arrangement constraint for the existing projector–camera system.
- The acquisition and reconstruction stages are straightforward and easy to implement in existing projector–camera system.
- PSI data acquisition efficiency is remarkably improved by the proposed LRE method.
- The perfect reconstruction property of LRE method is proven mathematically. The principle that underlies this property is a straightforward extension of Nyquist–Shannon sampling theorem to frequency domain.
- PSI is a general and sound theoretical model to decompose the direct and global illuminations and to perform 3D reconstruction under global illumination. PSI is a general method because it makes no specific assumption on measured objects, and captures the complete information of the projector–camera pair without requiring special hardware and restrictions on arrangement of the projector–camera pair. It thus can be used to solve this complex problem in the case of more general practical applications. PSI is a sound method because of the perfect reconstruction property of the LRE method.

In the future, research can be conducted on different base patterns for PSI. PSI obtains complete projector–camera pair information; thus, future research can further use this information. For instance, high-order interreflections can be separated, or BSSRDF of objects in a scene can be acquired. Studies on new PSI applications can also be conducted. For example, PSI can be adopted for super-resolution imaging and used for measuring and modeling the spatially varying point spread function of imaging lens.

Appendix 1

The cornerstone of LRE method is the observation made in this study that any pixel on the camera can only receive light from a local region on the projector, which will be confirmed in this section.

O’Toole and Kutulakos (2010) classified light transport capture methods into four categories (Fig. 2 in their paper). PSI corresponds to the first situation, where light transport matrix is sparse and high-rank. However, the sparsity of this light transport matrix has its unique characteristic, which is the observation that any pixel on the camera can only receive light from a local region on the projector made in this study.

O’Toole et al. (2012) have given an experiment that showed the light transport matrix along a slice in the camera perspective. They obtained their result with a coaxial arrangement by using the method of Schechner et al. (2007), which is similar to Naive SI implemented by Hadamard basis. O’Toole et al. (2012) pointed out that “Note that non-zero elements in this slice mostly concentrate around the diagonal. This indicates that most light is transported between nearby camera/projector pixels”. This statement is precisely the observation made in this study. We further investigated this property in this study to accelerate capture efficiency.

To further confirm that our observation is still reasonable in more general situations where non-coaxial arrangement is adopted, we provide an additional experiment in this section. In this additional experiment, the localization slices obtained after using adaptive regional SI (the first stage of LRE method) are shown. To be more specific, the vertical slices $h^V(v'; u, v)$ along the six red lines in Fig. 12a are shown. Each column in Fig. 12b–g corresponds to one vertical slice $h^V(v'; u, v)$. Since the whole light transport matrix is so huge that we are not able to show it, this sliced version of light transport matrix can well support the observation made in the paper.

In each subfigure of Fig. 12b–g, a line through the whole figure can be found. This line may be interrupted in the middle due to occlusion. This line is caused by direct light. As can be seen clearly, non-zero regions are concentrated near this direct line, which is a strong evidence to support the observation made in this study.

In Fig. 12b, a single line can be found, and this line is narrow compared to the line in Fig. 12c and d. This is due to the fact that global illumination is not significant on the toy bear, while subsurface scattering dominated in Fig. 12c and d that cause light scatters around the direct illumination point and incurs a wider line. Figure 12e shows the slice along the bear claw. There is a line, which is not very obvious, parallel to the direct line on the left part of this subfigure. This line corresponds to the interreflection light reflected from the impeller. On the right part of this subfigure, the line becomes wider and is caused by mirrored light reflect from the bear claw to the impeller. Figure 12f and g correspond to slices along the impeller and the metal part. Interreflections can be clearly seen in these subfigures. These interreflections are also concentrate near the direct illumination line.

Appendix 2

Lemma 1 Assume $h(u', v'; u, v)$ is the light transport coefficient between camera pixel (u, v) and projector pixel (u', v') , by projecting periodic extension patterns in the form of Eq. (7), the reconstructed image of camera pixel (u, v) by

Fig. 12 Localization slices obtained after using adaptive regional SI. **(a)** Investigated scene. This scene consists of five objects, which are indicated by corresponding Arabic numerals. The positions where we presented the localization slices are marked by red lines, beside each of which exist a circle with a letter inside. The letter indicates which subfigure is responsible for the corresponding position. **(b–g)** Localization slices obtained after using adaptive regional SI for each posi-

tion indicated in **(a)**. The vertical axis of each subfigure corresponds to v' axis of projector, and the horizontal axis of each subfigure corresponds to u axis of camera (also known as camera column number). The position of v axis of camera is marked by the camera row number in each subfigure. Each image is rescaled for visual clarity. The multiplied factor is also marked on each subfigure

Fourier-based SI reconstruction algorithm becomes a periodic extension version of the original light transport coefficients

$$\tilde{h}_r(u'_r, v'_r; u, v) = 2b \cdot \sum_{r_1=-\infty}^{+\infty} \sum_{r_2=-\infty}^{+\infty} h(u'_r - r_1 M_s, v'_r - r_2 N_s; u, v), \quad (\text{A-1})$$

where (u'_r, v'_r) is a pixel on the reconstructed image, and r_1 and r_2 are integers.

Proof Similar to Eqs. (4) and (5), when each sample in the frequency domain is obtained by using four periodic extension patterns, the reconstructed image of Fourier-based SI reconstruction algorithm is given by applying IDFT on the captured intensity as

$$\begin{aligned} \tilde{h}_r(u'_r, v'_r; u, v) &= F^{-1}[H(k_s, l_s; u, v)] \\ &= F^{-1} \left[\sum_{v'=0}^{N-1} \sum_{u'=0}^{M-1} 2b \cdot h(u', v'; u, v) \cdot W_{M_s, N_s}^{u' k_s, v' l_s} \right], \end{aligned} \quad (\text{A-2})$$

where $W_{M_s, N_s}^{u' k_s, v' l_s} = \exp\{-j \cdot 2\pi[(u' \cdot k_s)/M_s + (v' \cdot l_s)/N_s]\}$. (u', v') is a pixel of the projector, and (u'_r, v'_r) is a pixel on the reconstructed image. $F^{-1}[\cdot]$ is the 2D IDFT operator.

We can explicitly express the 2D IDFT operator in Eq. (A-2), then Eq. (A-2) can be written as

$$\begin{aligned} \tilde{h}_r(u'_r, v'_r; u, v) &= \frac{2b}{M_s \cdot N_s} \sum_{k_s=0}^{M_s-1} \sum_{l_s=0}^{N_s-1} \sum_{v'=0}^{N-1} \sum_{u'=0}^{M-1} h(u', v'; u, v) \cdot W_{M_s, N_s}^{u' k_s, v' l_s} \cdot \hat{W}_{M_s, N_s}^{u'_r k_s, v'_r l_s} \\ &= \frac{2b}{M_s \cdot N_s} \sum_{u'=0}^{M-1} \sum_{v'=0}^{N-1} h(u', v'; u, v) \cdot \sum_{k_s=0}^{M_s-1} \sum_{l_s=0}^{N_s-1} W_{M_s, N_s}^{u' k_s, v' l_s} \cdot \hat{W}_{M_s, N_s}^{u'_r k_s, v'_r l_s}, \end{aligned} \quad (\text{A-3})$$

where $\hat{W}_{M_s, N_s}^{u'_r k_s, v'_r l_s} = \exp\{j \cdot 2\pi[(u'_r \cdot k_s)/M_s + (v'_r \cdot l_s)/N_s]\}$.

The term $\frac{1}{M_s \cdot N_s} \sum_{k_s=0}^{M_s-1} \sum_{l_s=0}^{N_s-1} W_{M_s, N_s}^{u' k_s, v' l_s} \cdot \hat{W}_{M_s, N_s}^{u'_r k_s, v'_r l_s}$ in Eq. (A-3) can be expressed as $\frac{1}{M_s \cdot N_s} \sum_{k_s=0}^{M_s-1} \sum_{l_s=0}^{N_s-1} \exp\{j \cdot 2\pi[(u'_r - u') \cdot k_s/M_s + (v'_r - v') \cdot l_s/N_s]\}$, which equals to the periodic impulse train $\sum_{r_1=-\infty}^{+\infty} \sum_{r_2=-\infty}^{+\infty} \delta(u'_r - u' - r_1 M_s, v'_r - v' - r_2 N_s)$. r_1 and r_2 are integers.

After simplifications, Eq. (A-3) can be written as

$$\begin{aligned} \tilde{h}_r(u'_r, v'_r; u, v) &= 2b \cdot \sum_{r_1=-\infty}^{+\infty} \sum_{r_2=-\infty}^{+\infty} \sum_{v'=0}^{N-1} \sum_{u'=0}^{M-1} h(u', v'; u, v) \cdot \delta(u'_r - r_1 M_s - u', v'_r - r_2 N_s - v') \\ &= 2b \cdot \sum_{r_1=-\infty}^{+\infty} \sum_{r_2=-\infty}^{+\infty} h(u'_r - r_1 M_s, v'_r - r_2 N_s; u, v) \\ &= 2b \cdot \sum_{r_1=-\infty}^{+\infty} \sum_{r_2=-\infty}^{+\infty} h(u'_r - r_1 M_s, v'_r - r_2 N_s; u, v). \end{aligned} \quad (\text{A-4})$$

where $*$ denotes the convolution operator. The second equation holds because the values of $h(u', v'; u, v)$ are implicitly assumed to be zeros for u' and v' that beyond the region of $\Omega = \{(u', v') | u' \in (0, M-1), v' \in (0, N-1)\}$.

LRE Reconstruction Theorem If the period $M_s \times N_s$ of the projected periodic extension patterns covers the visible region, light transport coefficients can be perfectly reconstructed by adopting Fourier-based SI reconstruction algorithm, that is, the light transport coefficients obtained by reconstruction is exactly equal to the light transport coefficients reconstructed by the naive SI.

Proof If the size of $M_s \times N_s$ covers all nonzero regions of $h(u', v'; u, v)$, then from Lemma 1, the reconstructed image can be regarded as a periodic extension version of the local visible region, with step size of M_s on the u' axis, and N_s on the v' axis. Aliasing does not occur in this situation (Fig. 3), and all information of $h(u', v'; u, v)$ is preserved in $\tilde{h}_r(u', v'; u, v)$. Thus, by adopting Fourier-based SI

reconstruction algorithm, light transport coefficients can be exactly reconstructed, which equals to light transport coefficients reconstructed by the naive SI.

Provided that the information of the visible region Ω_s is known, the light transport coefficients $h_r(u', v'; u, v)$ can be exactly obtained by taking the values inside the visible region Ω_s of $\tilde{h}_r(u', v'; u, v)$ and setting zeros outside Ω_s of $\tilde{h}_r(u', v'; u, v)$ as given by

$$\begin{aligned} h_r(u', v'; u, v) &= \tilde{h}_r(u', v'; u, v) \cdot M_r(u', v'; u, v) \\ &= 2b \cdot \sum_{r_1=-\infty}^{+\infty} \sum_{r_2=-\infty}^{+\infty} h(u'_r - r_1 M_s, v'_r - r_2 N_s; u, v) \cdot M_r(u', v'; u, v) \\ &= 2b \cdot h(u', v'; u, v), \end{aligned} \quad (\text{A-5})$$

where $\cdot$ represents the element-wise product. u' and v' take values of $u'=0, 1 \dots M-1$ and $v'=0, 1 \dots N-1$. $M_r(u', v'; u, v)$ is a mask, which has the form of

$$M_r(u', v'; u, v) = \begin{cases} 1, & (u', v') \in \Omega_s \\ 0, & (u', v') \notin \Omega_s. \end{cases} \quad (\text{A-6})$$

In Eq. (A-5), the reconstructed $h_r(u', v'; u, v)$ exactly equals to the light transport coefficients reconstructed by the naive SI, which is $2b \cdot h(u', v'; u, v)$ in Eq. (5).

Appendix 3

CS methods are introduced for efficient signal capture and can be used for light transport capture (Peers et al. 2009; Sen and Darabi 2009). In this section, we compare PSI and CS method in terms of accuracy and reconstruction time.

The most significant contribution of PSI is that the LRE reconstruction theorem provides the theoretical limit that has to be satisfied for perfect light transport reconstruction. However, in its theory, CS methods have no such guarantee. The perfect reconstruction property of LRE method comes from the fact that we further investigated the unique property exists in light transport coefficients, which is the observation made in this study that “any pixel on the camera can only receive light from a local region on the projector”. This property eliminates many unnecessary solutions and reduces the scale of the problem to be solved. To be more specific, LRE reconstruction theorem makes the reconstruction algorithm focus on the region where non-zeros values exist.

We used synthetic data to illustrate the perfect reconstruction property of PSI and compared with CS method in terms of accuracy and reconstruction time. We used two different kinds of light transport coefficients in this simulation, as shown in Fig. 13. The reference data is shown in Fig. 13a and f. There are three speckles in Fig. 13a, which simulate the effect of interreflections. In Fig. 13f, only one larger speckle exists to simulate the effect of subsurface scattering. The resolution of the light transport coefficients was 384×216 , and a size of 32×32 is enough to cover the visible region. PSI requires 816 Fourier coefficients for exact reconstruction (with the Fourier coefficients consumed in localization stage included as well). The reconstructed results are shown in Fig. 13b and g, with the calculation time at the bottom of each subfigure. The error maps between the reconstructed results and the reference are shown in Fig. 13d and i, which indicate that the errors are not visible. From the error maps, we can conclude that exact reconstruction is achieved. The PSNR is also calculated and marked. The PSNR for Fig. 13b is 370.2079 db, and the PSNR for Fig. 13g is 371.9682 db. Only the interested regions indicated in Fig. 13a and f are shown in subfigures.

CS method with the same number of captured Fourier coefficients is also applied. We randomly selected 816 Fourier coefficients to measure, and minimize the total variation (TV) of the reconstructed light transport coefficients. The CS method used is similar to Meng et al. (2019). The reconstructed results are shown in Fig. 13c and h, with the calculation time at the bottom of each subfigure. The time required for CS is much longer than PSI, this is because PSI uses inverse FFT for reconstruction (the size of 32×32 image patch also requires a small amount of computation time), while CS needs to solve an optimization problem with 384×216 unknowns and 816×2 constraints (the $\times 2$ implies for each Fourier coefficient, both of its real and imaginary component can provide a constraint). Although both the reconstructed results by PSI and CS method look the same as the reference, the error maps indicate that PSI is more accurate than CS method. The error maps between the reconstructed results and the reference are shown in Fig. 13e and j. The largest error in Fig. 13e is about +4, which is nearly 1.57 percentage of the maximum (which is 255) in the reference light transport coefficients. The PSNR value is also calculated and marked. The PSNR is 73.0570 db for this situation. The largest error in Fig. 13j is about +6, which is nearly 2.35 percentage of the maximum (which is 255) in the reference light transport coefficients. The PSNR is 71.5602 db for this situation.

Appendix 4

Structured light transport (O'Toole et al. 2014) was introduced for 3D reconstruction under global illumination. However, non-epipolar dominance assumption was made, which implies that epipolar indirect illumination is not strong, to ensure the success of 3D reconstruction in presence of global illumination. In this section, we provide an experiment where epipolar global illumination is dominant such that the non-epipolar dominance assumption is not met. Under this situation, structured light transport would fail to provide high accuracy 3D data points, while PSI is still capable of distinguishing direct illumination and global illumination by determining the smallest speckle (the speckle with the minimum number of pixels) inside the epipolar threshold as direct illumination. Refer to (Nayar et al. 2006) for an analysis of the smooth assumption of global term, based on which the smallest speckle can be identified as direct term. Currently, PSI is able to distinguish between direct and epipolar dominance global illumination that is not highly specular. For highly specular epipolar dominance global illumination caused by mirror, spatial correlation can further be considered in future works.

The investigated object is an etalon formed by two steel gauge blocks, as shown in Fig. 14a. Each of these two steel gauge blocks can be considered as standard plane. From

Fig. 13 Comparison between PSI and CS method. Reference light transport coefficients simulated for (a) interreflections, and (f) subsurface scattering. The resolution was 384×216 . Light transport coefficients simulated for interreflections reconstructed by (b) PSI, and (c) CS method. Error maps for (d) PSI reconstructed result, (e) CS

method reconstructed result. Light transport coefficients simulated for subsurface scattering reconstructed by (g) PSI, and (h) CS method. Error maps for (i) PSI reconstructed result, (j) CS method reconstructed result

localization stage, a period of 300×40 was enough to cover the visible region. A total of 6002 Fourier coefficients were required to reconstruct the 2D image of light transport coefficients for each pixel on the camera when PSI is applied. Together with the localization stage, the total acquisition time was about 250 s. The scene captured by the camera

under uniform illumination is shown in Fig. 14b, and the scene captured by the camera under sinusoidal pattern is shown in Fig. 14c. The regions inside the yellow boundaries correspond to epipolar global illumination dominated regions. Light transport coefficients with epipolar lines plotted are shown for several camera pixels inside this region

[Fig. 14d, e]. As can be seen in these subfigures, more than one speckles present near to the epipolar line, which indicates that epipolar global illumination is dominated. In this situation, Eq. (17) is changed slightly to determine the

representing point of the smallest speckle as the direct illumination point such that PSI still works faithfully. The 3D shape reconstruction result of PSI is shown in Fig. 14f. To evaluate the accuracy of PSI under situation when epipolar

Fig. 14 PSI under situation when epipolar global illumination dominates. (a) Investigated object. (b) Scene captured by camera. The positions where we presented the light transport coefficients are marked by red circles with a letter inside. The letter indicates which subfigure is responsible for the corresponding position. (c) The captured scene when sinusoidal pattern is projected. Overlapped patterns can be observed, especially in the regions bounded by the yellow lines. The regions inside the yellow boundaries correspond to epipolar global illumination dominated regions. (d–e) Reconstructed light

transport coefficients by PSI. The epipolar lines are plotted, and the projector coordinates are shown in each enlarged figure. The position of the camera pixel to reconstruct the corresponding light transport is marked on the upper right corner of each subfigure. The reconstructed image is rescaled for visual clarity. The multiplied factor is also marked on the upper right corner of each subfigure. (f) 3D shape reconstructed by PSI. (g) Fitting plane error of the right part of the etalon. (h) Fitting plane error of the left part of the etalon

global illumination dominates, a plane is fitted for both right and left part separately. The error maps between the fitted plane and the reconstructed data points are shown in Fig. 14g and h. The mean absolute errors of the right and left part of the etalon are 0.015 (mm) and 0.013 (mm), respectively, and the root mean square errors of the right and left part of the etalon are 0.018 (mm) and 0.015 (mm), respectively.

Appendix 5

The number of patterns required by Naive SI and PSI will be introduced in this section.

The number of patterns required is dependent on the number of Fourier coefficients required. Thus, we first introduce the number required for Fourier coefficients. Since the resolution used for projection is usually even number, we provide formula for the number of Fourier coefficients required for 1D projection functions and 2D images under even resolution in this section.

Firstly, for 1D projection functions [Fig. 15a], the number of Fourier coefficients required is:

$$\begin{aligned} C_1 &= \frac{M-2}{2} + 2 \\ &= \frac{M}{2} + 1, \end{aligned} \quad (\text{A-7})$$

where M is the 1D resolution, and minus two on the numerator indicates the two symmetric centers that do not share conjugate symmetric with any coefficient [squares with mesh in Fig. 15a]. The plus two indicates the capture of the two symmetric centers. The symmetric centers are in fact real numbers, thus, the number of patterns required is halved (only the real parts are required to capture, and the imagery parts are zeros). Suppose S -step (S is an even number) phase-shifting sinusoidal patterns is used, the number of patterns required is:

$$S_1 = S \cdot \frac{M-2}{2} + 2 \cdot S/2. \quad (\text{A-8})$$

If four-step phase-shifting sinusoidal patterns is used (we used four-step phase-shifting sinusoidal patterns because the reconstructed images have high signal-to-noise ratio), the number of patterns required is $2M$.

Secondly, for 2D images [Fig. 15b], the number of Fourier coefficients required is:

$$\begin{aligned} C_2 &= \left[\frac{M-2}{2} + 2 \right] + \left[\frac{N-2}{2} + 1 \right] + \left[\frac{(M-1)(N-1)-1}{2} + 1 \right] \\ &= \frac{MN}{2} + 2, \end{aligned} \quad (\text{A-9})$$

Fig. 15 Conjugate symmetric property of real valued signals with resolution of even number for (a) 1D projection functions, and (b) 2D images. The squares with the same color indicate positions that are in conjugate symmetric with each other. For visual clarity, upper case letters are marked inside squares to indicate conjugate symmetric, i.e., the squares with the same letters are conjugate symmetric with each other. The squares with mesh inside indicate the center of symmetry. The squares inside the region made up of the solid black lines corresponds to the Fourier coefficients required to capture. The horizontal and vertical resolution M and N are assumed as even numbers

where M and N are the horizontal and vertical resolution of the images, and the term inside the three square-brackets correspond to the number of Fourier coefficients in the first row, the number of Fourier coefficients in the first column, and the number of Fourier coefficients on the lower right region exclude the elements in the first row and first column. The added integers inside the square brackets are the number of symmetric centers which do not share conjugate symmetric with any element.

Suppose S -step (S is an even number) phase-shifting sinusoidal patterns is used, the number of patterns required is:

$$\begin{aligned} S_2 &= \left[S \cdot \frac{M-2}{2} + 2 \cdot S/2 \right] + \left[S \cdot \frac{N-2}{2} + S/2 \right] \\ &\quad + \left[S \cdot \frac{(M-1)(N-1)-1}{2} + S/2 \right]. \end{aligned} \quad (\text{A-10})$$

If four-step phase-shifting sinusoidal patterns is used, the number of patterns required is $2MN$.

Lastly, the number of Fourier coefficients required by LRE method is the summation of the Fourier coefficients required by the 1D localization stage (stage 1 of LRE method) and the Fourier coefficients required by the reduced 2D image patches (stage 2 of LRE method). Suppose M and N be the horizontal and vertical resolution of projector, and M_s and N_s be the horizontal and vertical size of the visible region, the number of Fourier coefficients required is:

$$C_3 = \frac{M}{2} + \frac{N}{2} + \frac{M_s N_s}{2} + 4. \quad (\text{A-11})$$

If four-step phase-shifting sinusoidal patterns is used, the number of patterns required is: $2M+2N+2M_s N_s$.

The number of Fourier coefficients required by Naive SI can be calculated by Eq. (A-9), and the number of patterns required when four-step phase-shifting sinusoidal patterns is used is $2MN$, which is far larger than the number required by LRE method.

Acknowledgements This work was supported by National Natural Science Foundation of China (NSFC) (61875007, 61735003), National Key R&D Program of China (2020YFB2010701), and foundation (6141B061106).

Author Contributions HJ designed the system, performed the experiments, analyzed the data, and wrote the manuscript. YL designed the system, performed the experiments, analyzed the data, and wrote the manuscript. HJ and YL contributed equally to this work. HZ conceived, designed, supervised the study, and interpreted the data. XL conceived, designed, and supervised the study. YX analyzed the data, and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Compliance with Ethical Standards

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bian, L., Suo, J., Situ, G., Li, Z., F. J., Feng, C., et al. (2016). Multi-spectral imaging using a single bucket detector. *Scientific Reports*, 6, 24752.
- Chan, W. L., Charan, K., Takhar, D., Kelly, K. F., Baraniuk, R. G., & Mittleman, D. M. (2008). A single-pixel terahertz imaging system based on compressed sensing. *Applied Physics Letters*, 93, 121105.
- Chen, H., Asif, M. S., Sankaranarayanan, A. C., & Veeraraghavan, A. (2015). FPA-CS: Focal plane array-based compressive imaging in short-wave infrared. In *IEEE Conference on Computer vision and Pattern Recognition (CVPR)* 2358–2366.
- Chen, T., Lensch, H. P. A., Fuchs, C., & Seidel, H.-P. (2007). Polarization and phase-shifting for 3D scanning of translucent objects. In *IEEE Conference on Computer vision and Pattern Recognition (CVPR)* 1–8.
- Chen, T., Seidel, H.-P., & Lensch, H. P. A. (2008). Modulated phase-shifting for 3D scanning. In *IEEE Conference on Computer vision and Pattern Recognition (CVPR)* 1–8.
- Chen, H., Weng, Z., Liang, Y., Lei, C., Xing, F., Chen, M., & Xie, S. (2014). High speed single-pixel imaging via time domain compressive sampling. In *2014 Conference on Lasers and Electro-Optics (CLEO)—Laser Science to Photonic Applications*. JTh2A.132.
- Chiba, N., & Hashimoto, K. (2017). 3D measurement by estimating homogeneous light transport (HLT) matrix. *IEEE International Conference on Mechatronics and Automation (ICMA)*.
- Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., & Sagar, M. (2000). Acquiring the reflectance field of a human face. In *Proceedings of ACM SIGGRAPH 2000*. 145–156.
- Devaux, F., Moreau, P.-A., Severine, D., & Eric, L. (2016). Computational temporal ghost imaging. *Optica*, 3(7), 698–701.
- Edgar, M. P., Gibson, G. M., & Padgett, M. J. (2019). Principles and prospects for single-pixel imaging. *Nature Photonics*, 13, 13–20.
- Ferri, F., Magatti, D., Lugiato, L. A., & Gatti, A. (2010). Differential ghost imaging. *Physical Review Letters*, 104, 253603.
- Garg, G., Talvala, E.-V., Levoy, M., & Lensch, H. A. (2006). Symmetric photography: Exploiting data-sparseness in reflectance fields. In *Proceedings of the 17th Eurographics conference on Rendering Techniques*. 251–262.
- Gungor, A., Kar, O. F., & Guven, H. E. (2018). A matrix-free reconstruction method for compressive focal plane array imaging. In *25th IEEE International Conference on Image Processing (ICIP)*. 1827–1831.
- Gupta, M., Agrawal, A., Ashok, V., & Narasimhan, S. G. (2013). A Practical approach to 3D scanning in the presence of interreflections, subsurface scattering and defocus. *International Journal of Computer Vision (IJCV)*, 102, 33–35. <https://doi.org/10.1007/s11263-012-0554-3>.
- Gupta, M., & Nayar, S. K. (2012). Micro phase shifting. In *Proceedings of IEEE Conference on Computer vision and Pattern Recognition (CVPR)*. 813–820.
- Hahn, J., Debes, C., Michael, L., & Zoubir, A. M. (2014). Compressive sensing and adaptive direct sampling in hyperspectral imaging. *Digital Signal Processing*, 26, 113–126.
- Hartley, R. I., & Peter, S. (1996). Triangulation. *Computer Vision and Image Understanding*, 68(2), 146–157.
- Jiang, H., Huanjie, Z., Xu, Y., Li, X., & Zhao, H. (2019). 3D shape measurement of translucent objects based on Fourier single-pixel imaging in projector-camera system. *Optics Express*, 27(23), 33564.
- Jiang, H., Zhou, Y., & Zhao, H. (2017a). Using adaptive regional projection to measure parts with strong reflection. In *Conference on 3D measurement Technology for Intelligent Manufacturing* 104581A.
- Jiang, H., Zhu, S., Zhao, H., Xu, B., & Li, X. (2017b). Adaptive regional single-pixel imaging based on the Fourier slice theorem. *Optics Express*, 25(13), 15118–15130.

- Lutzke, P., Peter, K., & Notni, G. (2011). Measuring error compensation on three-dimensional scans of translucent objects. *Optical Engineering*, 50(6), 063601. <https://doi.org/10.1117/1.3582858>.
- Masselus, V., Peers, P., Dutré, P., & Willems, Y. D. (2003). Relighting with 4D incident light fields. In *Proceedings of ACM SIGGRAPH 2003*. 613–620.
- Meng, W., Shi, D., Huang, J., Yuan, K., Wang, Y., & Fan, C. (2019). Sparse Fourier single-pixel imaging. *Optical Express*, 27(22), 31490–31503.
- Nayar, S. K., Ikeuchi, K., & Kanade, T. (1991). Shape from inter-reflections. *International Journal of Computer Vision (IJCV)*, 6, 173–195.
- Nayar S. K., Krishnan, G., Grossberg, M. D., & Ramesh, R. (2006). Fast separation of direct and global components of a scene using high frequency illumination. *Proceedings of ACM SIGGRAPH, 2006*, 935–944. <https://doi.org/10.1145/1179352.1141977>.
- O'Toole, M., & Kutulakos, K. N. (2010). Optical computing for fast light transport analysis. *ACM Transactions on Graphics (TOG)*, 29(6), 1–12.
- O'Toole, M., Mather, J., & Kutulakos, K. N. (2014). 3D Shape and Indirect appearance by structured light transport. In *Proceedings of IEEE Conference on Computer vision and Pattern Recognition (CVPR)*. 3246–3253.
- O'Toole, M., Raskar, R., & Kutulakos, K. N. (2012). Primal–dual coding to probe light transport. *ACM Transactions on Graphics (TOG)*, 31(4), 1–11. <https://doi.org/10.1145/2185520.2185535>.
- Park, J., Byun, S.-C., & Byung-Uk, L. (2009). Lens distortion correction using ideal image coordinates. *IEEE Transactions on Consumer Electronics*, 55(3), 987–991.
- Peers, P., Mahajan, D. K., & Lamond, B. (2009). Compressive light transport sensing. *ACM Transactions on Graphics (TOG)*, 28(1), 1–18. <https://doi.org/10.1145/1477926.1477929>.
- Pharr, M., Jakob, W., & Humphreys, G. (2017). *Physically based rendering* (3rd ed., Vol. 7). Amsterdam: Elsevier.
- Phillips, D. B., Sun, M., JM, T., Edgar, M. P., Barnett, S. M., Gibson, G. G., et al. (2017). Adaptive foveated single-pixel imaging with dynamic supersampling. *Science Advances*, 3, 1601782.
- Radwell, N., Mitchell, K. J., Gibson, G. M., Edgar, M. P., Richard, B., & Padgett, M. J. (2014). Single-pixel infrared and visible microscope. *Optica*, 1, 285–289.
- Ren, P., Dong, Y., Lin, S., Tong, X., & Guo, B. (2015). Image based relighting using neural networks. *ACM Transactions on Graphics (TOG)*, 34(4), 1–12. <https://doi.org/10.1145/2766899>.
- Ryczkowski, P., Barbier, M., Friberg, A. T., Dudley, J. M., & Goery, G. (2016). Ghost imaging in the time domain. *Nature Photonics*, 10, 167–170.
- Schechner, Y. Y., Nayar Shree, K., & Belhumeur, P. N. (2007). Multiplexing for optimal lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)* 29, 8, 1339–1354.
- Sen, P., Chen, B., Garg, G., Marschner, S. R., Mark, H., Marc, L., et al. (2005). Dual photography. *Proceedings of ACM SIGGRAPH, 2005*, 745–755. <https://doi.org/10.1145/1186822.1073257>.
- Sen, P., & Darabi, S. (2009). Compressive dual photography. *Computer Graphics Forum*, 28(2), 609–618. <https://doi.org/10.1111/j.1467-8659.2009.01401.x>.
- Sun, M.-J., Edgar, M. P., Gibson, G. M., Baoqing, S., Neal, R., Robert, L., et al. (2016). Single-pixel three-dimensional imaging with time-based depth resolution. *Nature Communications*, 7, 12010.
- Sun, B., Edgar, M. P., William, B. R., Vittert, L. E., Welsh, S. S., Adrian, B., et al. (2013). 3D Computational Imaging with Single-Pixel Detector. *Science*, 340, 844–847.
- Sun, B., Welsh, S. S., Edgar, M. P., Shapiro, J. H., & Padgett, M. J. (2012). Normalized ghost imaging. *Optics Express*, 20(15), 16892.
- Tarjan, R. (1972). Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2), 146–160.
- Wang, Y., Suo, J., Fan, J., & Dai, Q. (2016). Hyperspectral computational ghost imaging via temporal multiplexing. *IEEE Photonics Technology Letters*, 28(3), 288–291.
- Watts, C. M., Shrekenhamer, D., Montoya, J., Lipworth, G., Hunt, J., Sleasman, T., et al. (2014). Terahertz compressive imaging with metamaterial spatial light modulators. *Nature Photonics*, 8, 605–609.
- Xu, Y., Zhao, H., Jiang, H., & Li, X. (2019). High-accuracy 3D shape measurement of translucent objects by fringe projection profilometry. *Optics Express*, 27(13), 18421.
- Zhang, Z. (1999). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334.
- Zhang, Z., Ma, X., & Zhong, J. (2015). Single-pixel imaging by means of Fourier spectrum acquisition. *Nature Communications*, 6, 6225.
- Zhang, Z., Wang, X., Zheng, G., & Zhong, J. (2017). Hadamard single-pixel imaging versus Fourier single-pixel imaging. *Optics Express*, 25(16), 19619.
- Zhao, H., Yang, X., Jiang, H., & Xudong, L. (2018). 3D shape measurement in the presence of strong interreflections by epipolar imaging and regional fringe projection. *Optics Express*, 26(6), 7117–7131.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.