Optics EXPRESS

PSLFM: a single-frame uncalibrated photometric stereoscopic light field measurement scheme based on dense convolutional neural networks

KAIYI ZHANG,[1] XING ZHAO,[1,2,3,5] YA WEN,[1] AND DA LI[1,2,3,4,*]

[1] *Institute of Modern Optics, Nankai University, Tianjin Key Laboratory of Micro-scale Optical Information Science and Technology, Tianjin 300350, China*
[2] *Nankai University Eye Institute, Nankai University Affiliated Eye Hospital, Nankai University, Tianjin 300350, China*
[3] *Prosper-Vision (Tianjin) Optoelectronics Technology Co., Tianjin 300192, China*
[4] *State Key Laboratory of Ultra-Precision Machining Technology, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Kowloon 999077, Hong Kong SAR, China*
[5] *zhaoxingtjnk@nankai.edu.cn*
[*] *da.li@nankai.edu.cn*

Abstract: In the realm of 3D measurement, photometric stereo excels in capturing high-frequency details but suffers from accumulated errors that lead to low-frequency distortions in the reconstructed surface. Conversely, light field (LF) reconstruction provides satisfactory low-frequency geometry but sacrifices spatial resolution, impacting high-frequency detail quality. To tackle these challenges, we propose a photometric stereoscopic light field measurement (PSLFM) scheme that harnesses the strengths of both methods. We have developed an integrated information acquisition system that requires only a single data acquisition and does not necessitate the light source vectors as input. This system enables uncalibrated multispectral photometric stereo reconstruction using a dense convolutional neural network (DCN). After that, the two reconstruction results are processed by frequency domain filtering, and the processed results are fused according to a certain weight, which can be adaptively determined by the algorithm according to the reconstruction error. Utilizing a light field camera as the sole acquisition device allows for natural alignment of data, mitigating registration errors. Our approach demonstrates effectiveness across both online datasets and laboratory samples, achieving an error of about 10° and lower in uncalibrated scenarios, with notable generalization. In conclusion, the proposed method facilitates single-frame measurement without calibration and exhibits strong robustness, which is expected to exert significant influence in the fields of machine vision, 3D printing and manufacturing, as well as virtual reality and augmented reality.

1. Introduction

In the late 1970s and early 1980s, Prof. Woodham pioneered the concept of photometric stereo, which revolutionized the field of surface reconstruction [1]. His groundbreaking idea harnessed the light and dark information from multiple images taken under varying lighting conditions to accurately determine the surface normal of an object, enabling its reconstruction in three dimensions. Traditionally, photometric stereo methods required controlled environments, often necessitating darkness during image capture and employing meticulous calibration techniques to regulate the light source's direction and mitigate ambient light interference. Woodham emphasized the need for at least three input images from distinct light source directions to calculate the object surface's normal direction. Once the normal vector is derived, the object's 3D shape can be computed by integrating the gradient field across its surface depth. Initially confined to laboratory settings, the technique gained momentum in the 1990s as computing power advanced, enabling researchers to explore its applicability in real-world scenarios. Researchers

#546806 https://doi.org/10.1364/OE.546806
Journal © 2025 Received 31 Oct 2024; revised 5 Dec 2024; accepted 2 Jan 2025; published 21 Jan 2025

delved into factors such as lighting conditions [2], material properties [3], and noise, seeking to refine reconstruction accuracy. The 2000s witnessed a surge in photometric stereo's accessibility and efficiency, thanks to advancements in camera technology and computing hardware. This led to the proposal of more precise lighting models [4], robust optimization algorithms [5], and enhanced normal estimation methods [6,7], driving significant strides in the technique's practical application. In recent years, researchers have endeavored to expand upon the constraints, exploring areas such as uncalibrated photometric stereo [8–10], Lambertian reconstruction in the presence of highlights and shadows [11–13], and precise reconstruction under ambient lighting conditions [14].

Traditional photometric stereo method needs to sequentially illuminate the object with different light sources, which is extremely inconvenient and cannot meet our needs in many application scenarios. To cope with this problem, researchers have proposed multispectral photometric stereo techniques. For instance, Hernandez et al. utilized three spatially separated light sources - red, green, and blue, and captured corresponding images in a single shot [7]. Similarly, Miyazaki et al. employed a 16-color multispectral light source and a multispectral camera, expanding beyond conventional three-color photometric stereo [15]. However, these methods require a predetermined association between the light source and the image, necessitating prior knowledge of the light source direction, which can be complicated. To overcome this hurdle, Hashimoto et al. proposed utilizing intrinsic images representing the object's albedo and guiding normals to approximate its shape [9]. Guo et al. introduced an equivalent directional illumination model for small surface patches with slowly varying normals, achieving results comparable to calibrated photometric stereo [16]. Despite substantial efforts by researchers to extend the calibration conditions, these methods often fail in performance for applications requiring high accuracy or in scenarios where calibration is challenging.

With the rapid development of deep learning in recent years, researchers consider applying it to photometric stereo to solve the problems that traditional methods fail to solve. Hiroaki Santo et al. pioneered this integration by using deep neural networks to map complex reflectance observations to surface normals, a milestone in the field [17]. Building on this work, Satoshi Ikehata introduced a refined photometric stereo algorithm driven by supervised convolutional neural networks, enabling the recovery of surface normal information from unstructured images and illumination data [18]. Additionally, Chen et al. developed the PS-FCN, a deep convolutional neural network that does not require predefined light directions during training and testing, enhancing flexibility and efficiency in photometric stereo [19]. However, these studies still face significant limitations in uncalibrated scenarios, resulting in relatively low final reconstruction accuracy. The average error in recovered results using FCN [19] on the publicly available Diligent dataset is approximately 16°. Therefore, we propose a novel network architecture inspired by DenseNet [20] to implement uncalibrated photometric stereo, which emphasizes feature reuse to maximize information utilization across layers, aligning with the pixel-level reconstruction focus of photometric stereo.

Simultaneously, as a pixel-by-pixel reconstruction method, photometric stereo inevitably introduces cumulative errors, which adversely affect the final measurement results, primarily manifested as distortions in low-frequency information. To address this issue, we consider light field, which offers robust low-frequency data. Light field is a crucial technique recognized for its ability to comprehensively capture scenes while offering increased degrees of freedom and flexibility [21], including the capability for post-capture adjustments of viewing angles and depth of field control.

Light field, represents a significant advancement in imaging technology by capturing detailed information about both the direction and intensity of light [22]. Unlike traditional 2D imaging methods, which primarily record the position as it passes through the lens, light field cameras provide a more comprehensive dataset by documenting the angular and intensity characteristics of light at various points. This enhanced data richness allows for more nuanced and detailed imaging [23]. In the fields of computer vision and image processing, techniques for light field camera reconstruction have become a prominent focus of research. For instance, Jeon et al introduced an algorithm adept at accurately estimating depth maps using small-lens light field cameras, leveraging a cost volume approach to achieve sub-pixel stereo correspondences [24]. Similarly, Cho et al., proposed a learning-based interpolation technique to elevate the quality of reconstruction from original images to refocused ones [25]. Furthermore, Jin et al., presented an end-to-end learning method targeting angular super-resolution of sparsely sampled light fields with significant baselines captured by light field cameras [26]. In this study, we employ a light field camera integrated with a microlens array to capture the photometric field information, enabling the entire acquisition process to be completed in a single frame.

Despite the advancements made in 3D reconstruction with light field cameras, the use of microlens arrays can lead to resolution degradation, particularly affecting high-frequency details. To address this problem, we propose a novel hybrid measurement scheme that combines photometric stereo and light field techniques. Photometric stereo, which reconstructs images on a pixel-by-pixel basis, naturally excels in capturing high-frequency details. Meanwhile, light fields, with their multi-view geometric information, offer enhanced low-frequency information. This method leverages the strengths of both approaches and introduces a novel solution to the challenges they encounter, aiming to achieve high-precision single-frame uncalibrated measurements.

2. Method

In this section, we firstly outlines the fundamental theory of photometric stereo, and then describe our proposed network architecture tailored for uncalibrated scenarios. Following by this, we present the principles underlying light field cameras. Finally, we detail our hybrid measurement scheme, which integrates the complementary strengths of both technologies.

2.1. Photometric stereo (PS)

Photometric stereo is a single viewpoint measurement method, which does not require the use of special conditions such as laser or structured light, but only the use of ordinary light sources for illumination. As a 3D reconstruction technique based on image luminance information, the basic principle of photometric stereo is to infer the depth information of objects in the scene by comparing the image luminance information in different viewing angles. In the classical photometric stereo assumption [1], the object under test is assumed to be a Lambertian body and the light source is assumed to be a parallel light incident. Here, we assume that the image luminance is I, the intensity of the light source is E, the incident direction of the light source is L, the reflectance of the object surface is ρ, and n represents the normal vector of the object surface. According to the Lambertian body reflection formula, we can get the following equation:

$$m = EL \cdot \rho n. \tag{1}$$

The light source direction we can obtain by calibrating, for example, the high light sphere, three non-coplanar light source angles can form a 3*3 non-singular matrix that can be inverted. As a constant ρ, we consider ρn as a whole N, $N = \rho$n, and after neglecting the unknown scaling and assuming the intensity of the light source per unit intensity, we can obtain a simplified Lambertian imaging model:

$$m = L \cdot N, \tag{2}$$

$$N = L^{-1} \cdot m, \tag{3}$$

Since n has unit length, we can estimate both the surface normal (vector direction) and the albedo (vector length) at the same time, and the paradigm of n should be 1, so we have the following equation:

$$\rho = |\rho \mathrm{n}| = \frac{N}{|N|}, \tag{4}$$

$$n = \frac{N}{\rho}. \tag{5}$$

If more than three light sources are used in the experiment, the corresponding light source direction matrix is no longer a square matrix, and the system of linear equations to be solved becomes hyper definite, and an approximate solution in the least squares sense can be obtained using the pseudo-inverse of L:

$$n = (L^T L)^{-1} L^T m. \tag{6}$$

The robustness of our results is contingent upon the diversity and abundance of light source directions at our disposal. The calculation process reveals that traditional photometric stereo methods are unable to perform accurate reconstruction in the absence of known direction vectors of light, which poses a challenge in uncalibrated scenes. To address this limitation, we consider deep learning techniques, which can effectively handle such scenarios and enhance reconstruction capabilities.

2.2. Network architecture design

Leveraging the power of deep learning, we can simulate the traditional photometric stereo by analyzing and learning from the dataset. This enables effective reconstruction in uncalibrated scenes where direct photometric stereo methods would otherwise fail. This paper introduces an end-to-end neural network architecture based on DenseNet [20] for the task of mapping surface normals of objects from multi-light source photometric stereo images. Photometric stereo, a pixel-wise reconstruction technique, demands intricate details. Traditional networks struggle in analyzing fine details accurately, and the attention mechanism is not well applied to it, which ultimately leads to the extraction of image information is not satisfactory. Motivated by DenseNet's concept, we design a network where each layer establishes direct connections with all preceding layers, facilitating efficient information flow within the network. Furthermore, each layer receives feature maps as input from all previous layers, enhancing the network's ability to acquire richer feature representations that align effectively with the principles of photometric stereo. We innovatively add the SE (squeeze excitation) attention module to the network [27], resulting in an effective improvement of the model's performance and expressive ability. The proposed network, termed DCN, offers advantages such as alleviating the gradient vanishing issue, ensuring parametric efficiency, robustness, and scalability. This perfectly solves the problems we mentioned.

In Fig. 1, we present the SE attention module integrated into our framework, which can improve the performance of convolutional neural networks by effectively enhancing the expression ability of feature channels Following a conventional procedure, we initially compress the data dimensions to a 1*1 size via global average pooling, while preserving the original channel count. Subsequently, we establish inter-channel correlations through two fully-connected layers (FC). The first layer reduces the channel count, while the second layer restores it to its original magnitude, introducing further non-linearities in the process. Ultimately, we set the activation function sigmoid that applies normalized weights to each channel's features, employing a channel-wise multiplication technique for weighting coefficients in this study.

The structure of the DCN network as depicted in Fig. 1 comprises denselayers, which are essentially "convolutional layers" housing two distinct convolutional operations. The first involves a 1*1 convolutional kernel to preprocess the image and augment the channel count, while the second utilizes a 3*3 convolutional kernel for the core convolutional processes. The two activation functions use a variant of rectified linear unit (ReLU), LeakyReLU, as well as

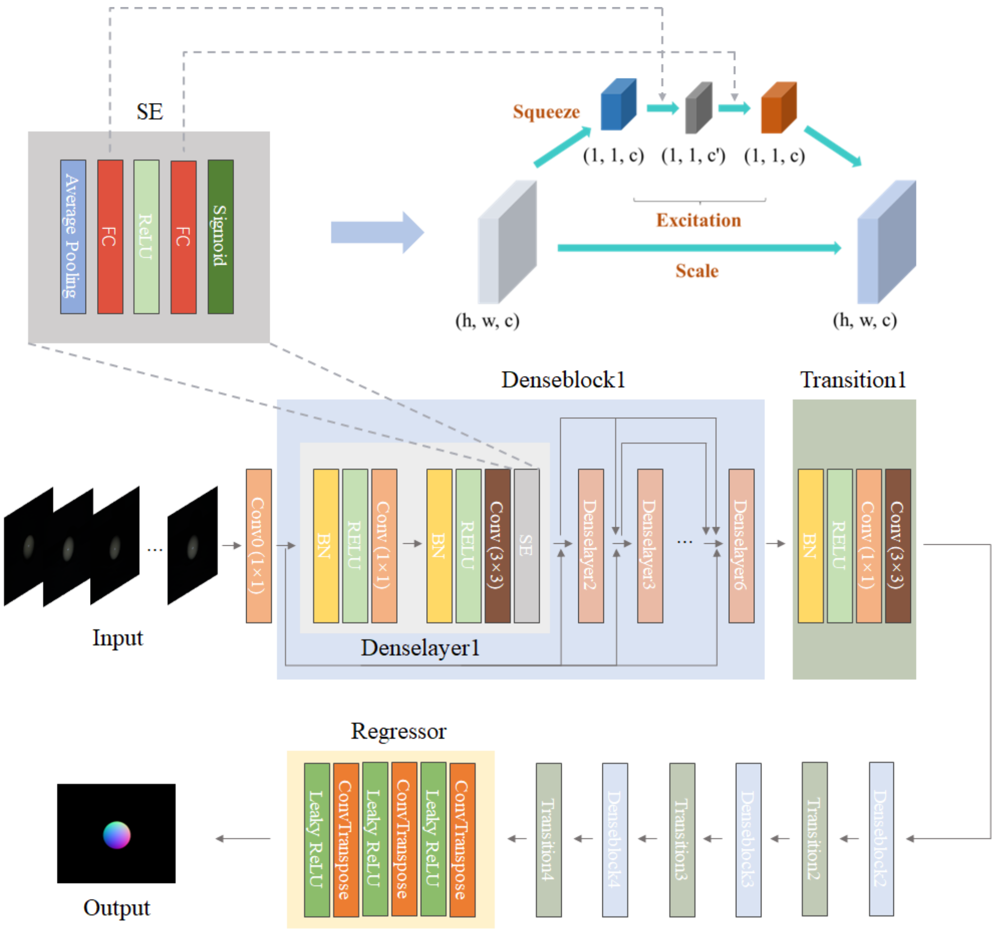

Fig. 1. Main modules required in the DCN and the flowchart of the network. The output of each denselayer in the denseblock is fed into each layer behind it, the bottom right corner is schematic diagram of the SE Attention Module (The *h* and *w* represent the size of the input, *c* denotes the number of input and output channels, *c'* signifies the number of channels after convolution).

two batch normalization (BN) layers to enforce regularization and mitigate overfitting, and finally, insert our SE module. This module employs two fully connected layers for non-linear learning to capture the importance of each channel. Specifically, it trains a set of weights that are applied to the feature maps of each channel, enabling multiplicative feature fusion. This process enhances the reconstruction quality with minimal additional computational cost. The default value for padding is 1. Each denseblock encompasses multiple denselayers, and transitions serve as intermediate layers between consecutive denseblocks. Each transition unit incorporates a BN layer, an activation function, a convolutional layer (1*1 kernel, stride 1), primarily facilitating channel adjustments and downsampling via another convolutional layer (3*3 kernel, stride 2) to enhance training efficiency. The final component, the Regressor, is a regression layer comprising three inverse convolution layers (4*4 kernel, stride 2) and one convolution layer (3*3 kernel). This segment primarily focuses on channel reduction and image size up-sampling, corresponding to the threefold downsampling in the preceding transition layers, thereby restoring the original image dimensions.

The flowchart of the network is also shown in the Fig. 1, first through a convolutional kernel of 3*3 convolutional layer for preprocessing, and then successively through the four denseblock and three transition, and finally through the Regressor recovery to become the final result, that is, for the surface of the object's normal vector picture. The choice of kernel sizes is based on established practices in similar tasks such as photometric stereo and image processing. The 1*1 kernel is commonly used in preprocessing layers as it introduces nonlinearity without expanding the receptive field, which helps retain computational efficiency. The 3*3 kernel, in contrast to larger kernels (e.g., 5*5 or 7*7), contains fewer parameters, facilitating faster convergence during training and reducing the risk of overfitting. This configuration enhances both the training efficiency and the generalization capability of the model. In the DenseNet architecture described in the Fig. 1, the denseblock configurations are 8, 16, 16, and 8 denselayers, respectively. These values can be adjusted based on the input image size, providing flexibility in the design. After the final denseblock, the number of channels totals 676. The subsequent regression part consists of layers with channel counts of 256, 128, 64, and 3. The first three layers perform up-sampling while gradually reducing the number of channels. The final convolutional layer then reduces the number of channels to three, corresponding to the RGB output. The model is optimized using the Adam optimizer with a learning rate set to 0.001, and the cosine similarity loss function is employed during training. The dataset is split into a training set and a test set at a 4:1 ratio. Training is conducted on two 16GB GPUs, with a total training time of 4 hours and 49 minutes. Our training was conducted over 15 epochs, with the loss curve showing a steady decrease during the first five epochs, after which it begins to stabilize and converge.

While DCN effectively addresses photometric stereo measurements in uncalibrated scenes, it continues to face challenges with low-frequency distortions due to inherent physical limitations. To address this issue, we incorporate light field, which offer a more robust solution for mitigating such distortions.

2.3. Light field camera

A light field camera is an advanced imaging device that captures information about the direction and intensity of light at each pixel in a scene, enabling applications such as depth estimation, depth of field adjustment, and 3D reconstruction. The core principle is to utilize a microlens array to capture the incident direction and intensity information of light. In this paper we use Lytro's light field camera for our experiments.

The Lytro Light Field Camera utilizes a microlens array arranged in a hexagonal shape, as shown in the Fig. 2. This microlens array is placed on the imaging plane of the main lens of a conventional camera. At the same time, the photosensitive devices are aligned parallel to the microlens array, and the distance between them is exactly equal to the focal length of the microlens array. Finally, through the mapping effect of the microlens array, light is projected onto the corresponding area to realize image capture.

We describe the working principle of the light field camera in terms of a two-dimensional coordinate system. Assume that there is a pixel point (x, y) in the image plane, corresponding to the coordinates (u, v) in the lens plane. At the same time, we introduce a parameter z to represent the depth of the object in the camera coordinate system. Then, the coordinates on the lens plane can be calculated by the following equation:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{pmatrix} f_x & 0 \\ 0 & f_y \end{pmatrix} \begin{bmatrix} \frac{x}{z} \\ \frac{y}{z} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \tag{7}$$

where f_x and f_y are the focal lengths of the lens, c_x and c_y are the coordinates of the lens center. This equation describes the relationship between pixel coordinates and object depth in conventional cameras.

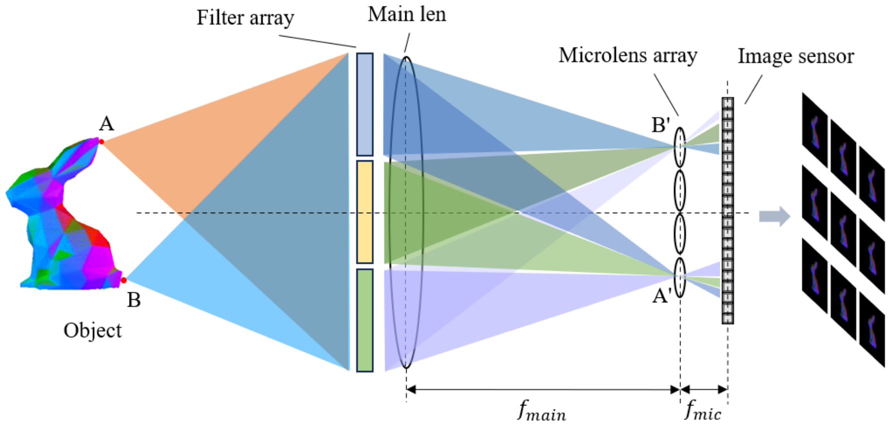

Fig. 2. The image formation for Lytro light field.

In order to obtain information about the direction of the light, the light field camera introduces an array of microlenses. Suppose we have an array of microlenses with sampling rate N in the lens plane. For each pixel point (u, v), we can calculate the angle of incidence of the light (θ, ϕ) as follows:

$$\begin{bmatrix} \theta \\ \phi \end{bmatrix} = \begin{bmatrix} \frac{u-c_x}{f_x N} \\ \frac{v-c_y}{f_y N} \end{bmatrix}. \tag{8}$$

This formula expresses information about the direction of light at each pixel point and is known as the light vector in light field cameras. By analyzing the angular difference between different pixel points, we can infer the depth information of objects in the scene. Since the light field camera is able to capture the direction of the light, it is able to better distinguish the overall shape of the object's surface and the slow part of the light change, which is valuable for our reconstruction work.

As illustrated in Fig. 3, the light field camera is employed to reconstruct the imaging process [24]. Initially, image decoding is conducted using Lytro Desktop, producing a refocused image on the computer and the quality of the demodulated image is directly related to the accuracy of the reconstructed depth map. The light field camera employed in this study features a 15 × 15 microlens array, enabling the demodulation of 225 viewpoints, with the central viewpoint image reconstructed for input into photometric stereo analysis. In general, more views lead to better quality. To evaluate this, we also tested 12*12 and 9*9 views from the 15*15 configuration. The results show that more views improve the depth estimation as well as the reconstruction quality of the final result. Next, camera calibration is performed using the camera's internal files, white images (usually taken under uniform illumination and intended to serve as a baseline for subsequent image processing). Simultaneously, subaperture images are analyzed for aberration correction. Additionally, the correlation function is employed to demodulate the original data, while frequency-domain filtering and color correction algorithms are applied for optimization, resulting in the acquisition of multi-view images. After these preprocessing steps, we apply a phase-shift-based sub-pixel displacement method for depth map estimation, optimizing the cost function alongside subsequent inspection and enhancement to derive the corresponding depth map. Finally, the depth maps are processed to calculate the gradients in both horizontal and vertical directions, from which the normal vectors are derived based on the definition of normal vector direction, ensuring the normal vectors is perpendicular to the gradient of the depth image.

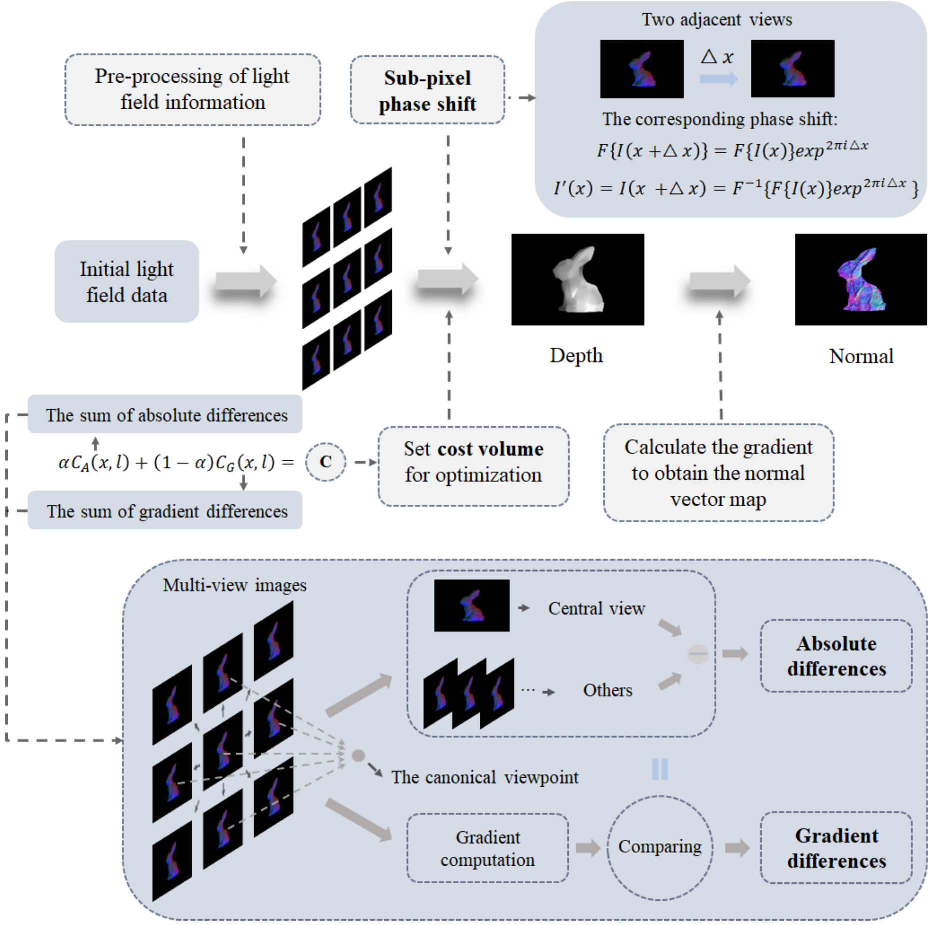
Two adjacent views
$\triangle x$
The corresponding phase shift:
$F\{I(x + \triangle x)\} = F\{I(x)\}exp^{2\pi i \triangle x}$
$I'(x) = I(x + \triangle x) = F^{-1}\{F\{I(x)\}exp^{2\pi i \triangle x}\}$
Pre-processing of light field information
Sub-pixel phase shift
Initial light field data
Depth
Normal
The sum of absolute differences
$\alpha C_A(x,l) + (1-\alpha)C_G(x,l) =$ C
The sum of gradient differences
Set cost volume for optimization
Calculate the gradient to obtain the normal vector map
Multi-view images
Central view
Others
Absolute differences
The canonical viewpoint
Gradient computation
Comparing
Gradient differences

Fig. 3. The flowchart of light field camera reconstruction (F denotes the discrete 2D Fourier transform, $I'(x)$ is the sub-pixel shifted image, $\triangle x$ is the shift of image I, the cost volume C is defined as a function of x and cost label l, α represents the weight).

According to the principles governing light field cameras, the microlens array enables the acquisition of multi-perspective images, significantly enhancing the reconstruction of the overall shape and outline of the object, commonly referred to as low-frequency information. However, this advantage is accompanied by a trade-off in resolution, resulting in suboptimal high-frequency details in the reconstruction outcomes. In this regard, we combine it with photometric stereo method using DCN, which excels in capturing high-frequency details, to create a hybrid measurement approach.

2.4. Theory of PSLFM

After obtaining the normal vector maps recovered from both photometric stereo and light field, our next step involves their integration. It is imperative to highlight that a mere summation and averaging of pixel values is insufficient for this task. In the reconstruction result, the high frequency information represents the edges and texture details of the object, and the corresponding low frequency part represents the overall shape and main characteristics of the object.

As shown in the Fig. 4, building upon established principles of our innovative fusion theory, we commence our approach by transforming the two mappings into the frequency domain using Fourier transform. In our experiments, the frequency domain threshold was set to 40, 60, or 80, depending on the object. A parameter scanning is conducted at 5-unit intervals to identify the trend of change, which is then used to select the optimal threshold corresponding to the best result. This step is crucial, as it allows us to analyze the frequency components of the data more effectively. Following this transformation, we apply filtering techniques to the two reconstruction results. Specifically, we utilize a high-pass filter to selectively retain the high-frequency details present in the photometric stereo outputs, ensuring that essential fine details are not lost. Simultaneously, we implement a low-pass filter to preserve the low-frequency components from the light field recovery results, capturing broader information.

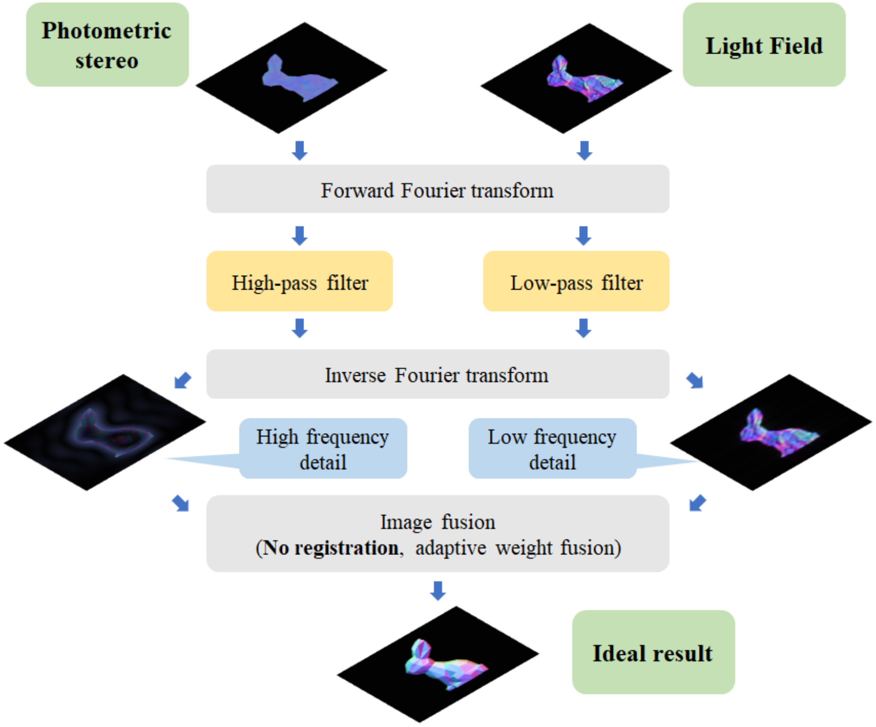

Fig. 4. Overall flowchart of image fusion.

To optimize the image fusion process, we assign different weights to the fusion operation based on the characteristics of the objects being tested. This weighting process is informed by our empirical findings, leading to the identification of the most effective parameters for fusion, which is the key step of adaptive weighting, namely setting initial values and iterative optimization. After conducting a series of rigorous fusion tests, we determine the optimal weights that yield the most accurate and cohesive fusion results. Importantly, we set different initial thresholds for high-pass filtering and low-pass filtering when filtering in the frequency domain. At the algorithmic level, the final outputs corresponding to each threshold can be compared with error analyses, facilitating the automated identification of the optimal solution and streamlining the data fusion process.

Just like in Fig. 5, the overall process of PSLFM is as follows: under RGB three light sources, the data is collected by light field camera, and then the corresponding picture of the central view is preprocessed by channel separation and used as the input of the network for uncalibrated photometric stereo reconstruction. Then, the light field recovery algorithm is used to reconstruct the image. Finally, the results of the two images are fused by frequency domain conversion.

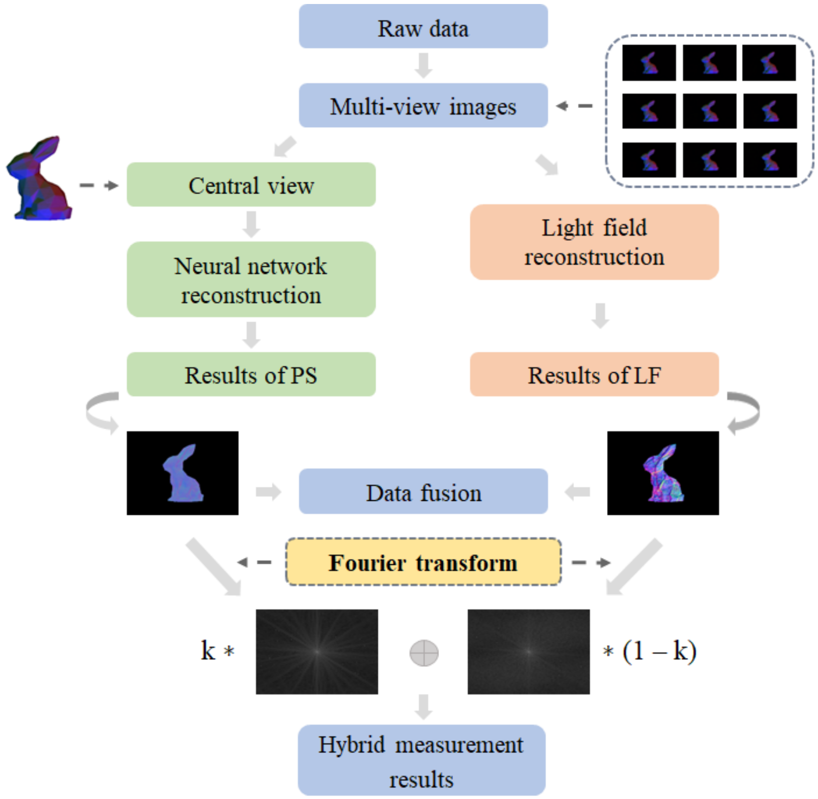

Fig. 5. The schematic diagram of PSLFM (*k* represents the weight coefficient during fusion, which can be set adaptively according to the algorithm).

Following these steps, we can obtain the final desired fusion results of PSLFM, which are subsequently analyzed and evaluated for errors. It is important to note that the images processed using both techniques are sourced from the light field camera throughout the entire process. The input for photometric stereo consists of a central view image, and the light field is also reconstructed from this central view. so the reconstruction results are inherently aligned, eliminating any issues related to data matching between different sensors or variations in acquisition angles. Consequently, image alignment is unnecessary, effectively reducing potential errors associated with this step. In addition, we use DCN for photometric stereo reconstruction without pre-calibration of light source directions, which also broadens the limitation of measurement scene to some extent.

3. Experiments

3.1. Experimental system setup

Based on the theory and methodology proposed in this paper, we designed an experimental system, as illustrated in Fig. 6. This system is a photometric field information acquisition system. To enhance application versatility and mitigate stray light interference, the entire setup is housed within a semi-circular, rotationally symmetric light absorption mask. Multi-spectral technology utilizing red, blue, and green LEDs is employed to facilitate rapid single-frame measurement. The mask is equipped with a number of sliders to adjust the height and inclination angle of the light source. According to the experiments of a large number of researchers, the incident inclination Angle of the light source is optimal between 40° and 60° [28], while considering the specific object tested in our experiment, after multiple tests, we finally choose 45° as the inclination angle for testing according to the error corresponding to each angle. The light field camera is placed directly on top of the device. Since it can simultaneously acquire 2D and 3D information, it is used as the data acquisition device in the whole system.

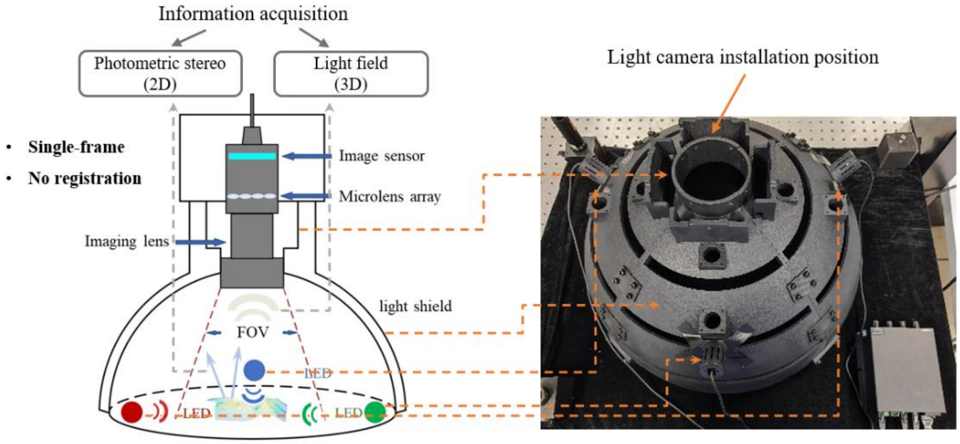

Fig. 6. Theoretical diagram of the experimental setup designed according to our method.

The three lights are simultaneously illuminated by a light controller to maintain the same intensity and to avoid the introduction of unwanted errors, and are captured by the top camera. The object should be positioned at the center of the shield, which corresponds to the central point of the light field camera's field of view (FOV), the shield and the experimental platform placed between a background light-absorbing cloth, and also to a certain extent to avoid the interference of stray light.

Figure 7 illustrates the typical setup and operating conditions of our experimental system. Notably, in environments outside the laboratory or in situations where creating a controlled dark room is challenging, we can mitigate ambient light interference by sealing excess apertures and openings.

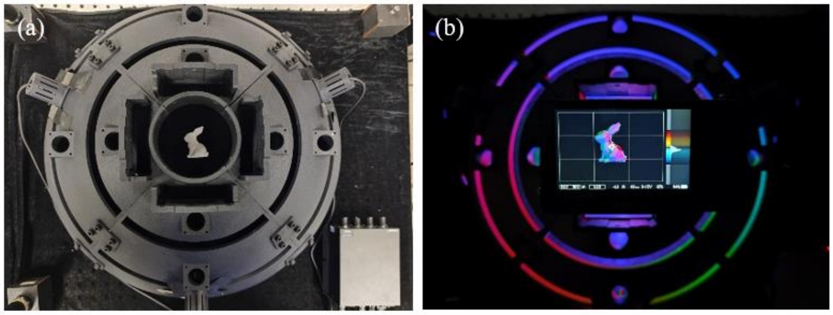

Fig. 7. Physical drawings of the experimental setup, (a) showcases a top-down view of the device, (b) illustrations the acquisition process of the bunny model by the light field camera, and the screen is the acquisition interface.

3.2. Neural network reconstruction

Due to the large dataset required for training the neural network, we use an existing dataset for training (our input for training is only the images, and we follow the assumption of traditional photometric stereo and ignore the unknown scaling relationship), the dataset is called CyclesPS [18], and in this work we use the CyclesPS_Dichromatic dataset from it for training, which contains 25 kinds of objects, each object contains 740 direct images and 740 indirect images, both of which are 256*256 in size, and also contains the surface normal vector truth map corresponding to each object. This dataset covers a wide range of objects, a large number of light sources, and a high reconstruction quality, which is crucial to improve the generalization of our method. In the training process, we divide the data set into training set and validation set according to the ratio of 4:1, and the experimental hardware conditions are two GPU T4s with 16 G memory. Preprocessing module we have set up in the network, mainly through the convolution layer with the size of 1*1 convolution kernel.

For the effectiveness of the DCN we built, DiLiGenT [29], as a recognized test and evaluation dataset in the field, we first carry out the relevant tests on this dataset, and the evaluation criterion is the size of the angle between the recovered normal vector and the real normal vector (the lower the better). The dataset contains ten real objects with normal reflectance, the image size is 612*512, and it contains 96 viewpoint maps with different orientations. Since we work with uncalibrated photometric stereo, we do not take the light source orientation in the dataset as an input, which means that we only have a few images as input. It should be noted that many network models involve complex and tedious processes, where light source intensities and orientations are estimated by means of complex mathematical models. These estimates are then used for photometric stereo reconstruction. The accuracy of these prior estimates can significantly affect the quality of the final results. Moreover, some uncalibrated network models simply prune the light source direction from the input, but there are still inputs other than the image present, which means that their method does not represent a completely uncalibrated reconstruction. Recently [30–32], while some network models have achieved errors below 10° on the DiLiGenT dataset, their test times exceed 14 minutes per object, a stark contrast to our test time of only a few seconds. This discrepancy arises from their focus on reducing error at the cost of increasing model complexity and introducing additional constraints. However, DCN significantly reduces the time cost, making it more suitable for practical industrial applications. After consulting most of the current mainstream models for similar scenarios and inputs, we investigate some uncalibrated approaches [19,33–37], compared with them and summary them. DCN(W/O) refers to the DCN without the attention mechanism. The data indicate that incorporating the attention mechanism leads to a slight reduction in the error for most objects, as well as a modest improvement in the overall average error. The results in Table 1 show that WT13 [35] of them performs better for the first few simpler objects reconstructed, but our method shows superior performance for the later complex objects, and the results obtained by our approach have a better average error on the overall dataset than most of the previous approaches, while showing better robustness. We show some samples with low recovery errors in Fig. 8.

Based on our foundational framework, we conduct laboratory tests using samples primarily fabricated through 3D printing using PLA (poly lactice acid), a material renowned for its favorable Lambertian properties. PLA closely aligns with our initial assumptions and possesses a white coloration, ideal for multispectral photometric stereo analyses, as it minimizes the absorption of multicolored light associated with other materials. To assess our findings, we engage the services of Nothton Metrology Technology (Beijing) Co., Ltd. They supply us with a truth map of surface normal vectors for the samples. Here's a simplified overview of the scanning process for determining normal vectors in a 3D model: The vertex normal vectors, which indicate surface orientation, are calculated using the triangulated surface information of the model. The Zeiss ATOS Q measuring system, with a maximum accuracy of 0.007 mm, is employed for scanning.

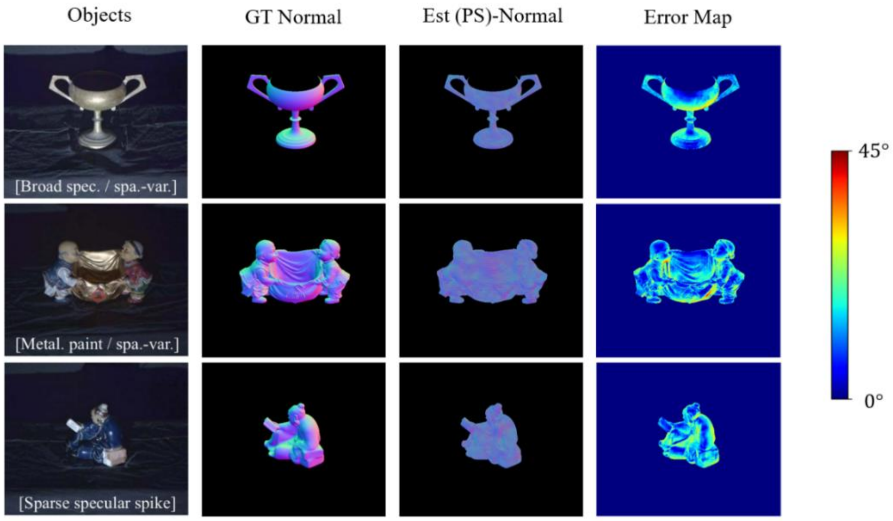

Fig. 8. The effect of some samples recovered by DCN in DiLiGenT (From top to bottom, the objects are goblet, harvest, reading). GT Normal represents the true normal vector of the object surface, Est (PS)-Normal is defined as the photometric stereoscopic counterpart recovered from the DCN, and Error Map shows the angle between the predicted normal and the true normal of the object surface based on the prediction.

Table 1. Performance of DCN versus some uncalibrated photometric stereo methods against the DiLiGenT dataset (in degrees)

Method	ball	cat	pot1	bear	pot2	buddha	goblet	reading	cow	harvest	Avg
AM07	7.27	31.45	18.37	16.81	49.16	32.81	46.54	53.65	54.72	61.7	37.25
SM10	8.90	19.84	16.68	11.98	50.68	15.54	48.79	26.93	22.73	73.86	29.59
WT13	**4.39**	36.55	**9.39**	**6.45**	14.52	**13.19**	20.57	58.96	19.75	55.51	23.93
PF14	4.77	**9.54**	9.51	9.07	15.9	14.92	29.93	24.18	19.53	29.21	46.66
LC18	9.30	12.60	12.40	10.90	15.70	19.00	18.30	22.30	15.00	28.00	16.30
FCN	6.62	14.68	13.98	11.23	**14.19**	15.87	20.72	23.26	**11.91**	27.79	16.02
DCN(W/O)	15.73	14.31	13.53	13.89	16.21	14.95	**13.78**	15.58	12.25	13.68	14.40
DCN	14.51	13.37	14.69	12.96	15.35	14.36	14.99	**14.84**	12.89	**13.61**	**14.16**

The process involves iterating through all triangular facets, computing each facet's normal vector (often using methods like cross multiplication), and adding these vectors to the corresponding vertex normals. After traversing all facets, the vertex normals are normalized. This approach allows for accurate light calculations, shadow generation, and other rendering operations, though factors like triangular face topology and normal vector smoothing may influence the results.

In this operational segment, we utilize the image corresponding to the center lens of the light field camera as the input for multispectral photometric stereo. This involves dividing the color-corrected image obtained after various processing steps into three images based on their respective color channels. Hence, the multispectral photometric stereo reconstruction of these samples comprises only three images, each with a size of 625*433 pixels. The effect is shown in Fig. 9. In our experimentation, we evaluate three 3D printed objects, yielding results indicating an average reconstruction error of approximately 10° with a limited number of image inputs. Specifically, the bunny model exhibits an error of about 11.67°, while the butterfly model displays superior performance with an error of approximately 6.05°, and the block model yields an error

Optics EXPRESS

of around 8.44°. Notably, these results surpass the performance of most uncalibrated methods, approaching the accuracy of some calibrated methods. This outcome underscores the robust generalizability of the network architecture.

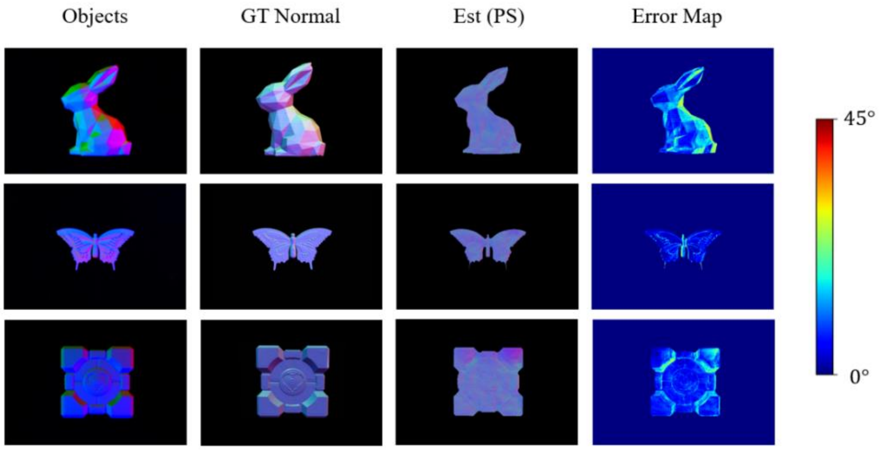

Fig. 9. Effectiveness of lab print samples recovered by DCN (From top to bottom, the objects are bunny, butterfly, block).

3.3. Data fusion

After obtaining the normal vector maps recovered and reconstructed in both ways, as shown in Fig. 10, we use the means of frequency domain filtering for image fusion and evaluate the obtained fused images by comparing them with the previously scanned true normal vector maps.

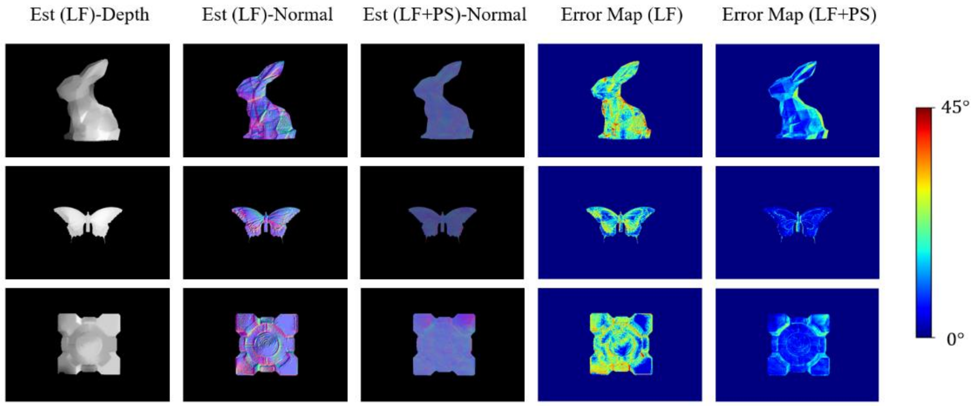

Fig. 10. Comparison of the results from the light field reconstruction alone and the fusion reconstruction. Est (LF)-Depth is the depth map reconstructed from the light field, from which we can process the corresponding surface normal vector map Est (LF)-Normal, next after data fusion we get the fusion reconstruction result Est (LF + PS)-Normal, and the reconstruction results of the light field and the fusion reconstruction results are evaluated.

In our experiment, we categorized the three selected objects into distinct types. The bunny, characterized by minimal surface texture and gradual surface changes, represents objects with less high-frequency detail. The block exhibits both gradual texture changes and notable step discontinuities, introducing more high-frequency information. The butterfly, with its extensive high-frequency details, falls into the category of objects with the most intricate surface features. As detailed in Table 2, we analyze and evaluate the measurement across three laboratory samples.

Table 2. Comparison of mean and median error of three objects in single light field measurement and hybrid measurement

Objects	Error Index	LF	PS	LF + PS
Bunny	Mean	21.42°	11.67°	11.67°
	Median	21.11°	11.01°	11.00°
Butterfly	Mean	17.65°	6.05°	5.87°
	Median	16.75°	4.38°	4.05°
Block	Mean	19.01°	8.44°	8.41°
	Median	19.33°	7.49°	7.44°

The data clearly demonstrate that, following the synthesis of comparison cases for the two types of errors, our proposed method significantly enhances the performance of the light field technique, resulting in substantial improvements in measurement accuracy within uncalibrated scenes. Furthermore, our hybrid measurements exhibit certain enhancements over photometric stereo for butterflies and blocks. However, this improvement is less pronounced in the case of rabbits.

3.4. Analysis of results

After obtaining the fusion results, we conducted a further analysis, specifically focusing on the evaluation of low-frequency information within the results. We quantified the number of pixels associated with each error angle, determining that an increase in the pixel count corresponding to low-error areas after fusion serves as evidence of the effectiveness of our method. To achieve this, we employed a Gaussian filter as a low-pass filter to process the three reconstruction results. This approach effectively eliminates high-frequency details while preserving low-frequency information, thereby enhancing the accuracy of our error evaluation.

In Fig. 11–13, the error curve distributions for the three objects exhibit a notable change. Specifically, the relative increase in the number of pixels concentrated in the low-error region, along with the leftward deviation of the error curve, indicates an improvement in the low-frequency

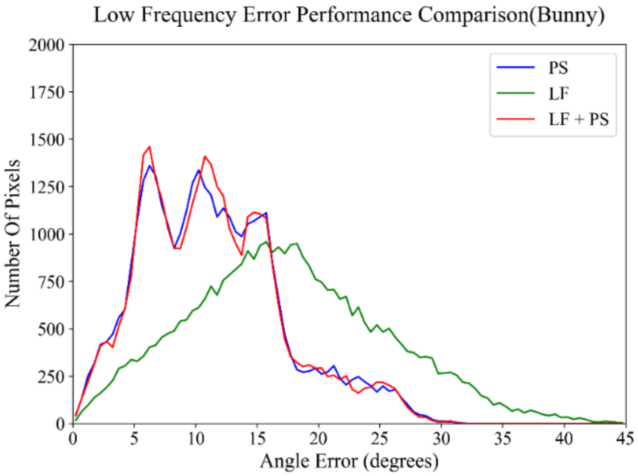

Fig. 11. Low frequency error of Bunny.

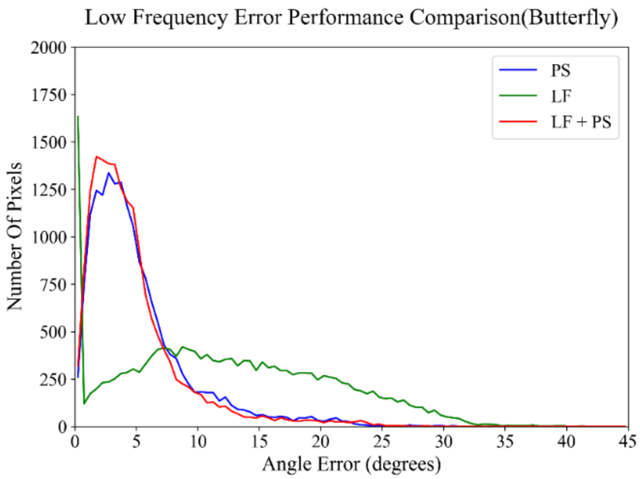

Fig. 12. Low frequency error of Butterfly.

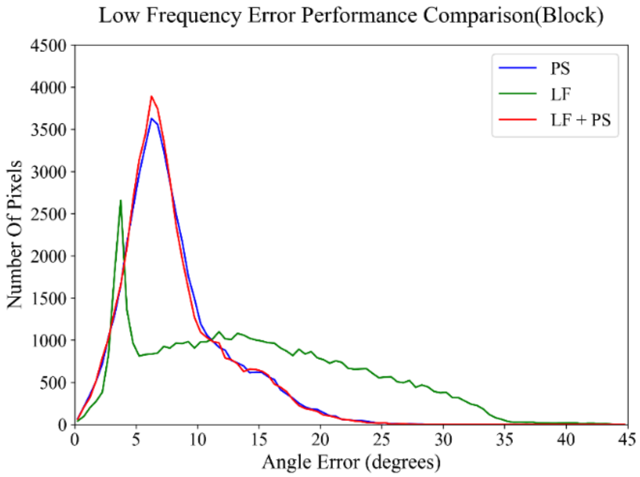

Fig. 13. Low frequency error of Block.

errors of the three objects. This finding supports the effectiveness of our fusion theory. However, it is worth noting that the average angle error for the bunny, as reported in Table 2, does not surpass the performance achieved by the DCN alone. We speculate that this outcome may be attributed to cumulative errors, and the limited high-frequency details inherent to the rabbit lead to relatively minor low-frequency distortions, suggesting that light field is not fully compensated. Additionally, this may introduce more high-frequency errors. The proportions of high-frequency details in the other two objects differ from those in the bunny, which is reflected in the overall improvement observed in the final fusion results. To sum up, our method is not applicable to all types of objects. Due to the characteristics of the light field, while bringing some excellent low-frequency information, it may also damage the high-frequency part. Therefore, in practical applications, it is crucial to determine the optimal balance for different objects—specifically, the trade-off between the benefits of preserved low-frequency information and the loss of high-frequency details. In addition, our method uses information from a single viewpoint, avoiding errors from changes in viewing angle or images alignment, which simplifies the process.

The test results demonstrate substantial enhancements in hybrid measurements across all three objects, particularly benefitting from the high-frequency details enabled by photometric stereo. The proposed system and methodology significantly extend the flexibility of the environment and offer considerable research value by employing multispectral techniques to achieve high-quality single-frame reconstruction from a single data acquisition, all without the need for calibration.

4. Conclusion

In this paper, we propose PSLFM, a novel hybrid measurement scheme that integrates the high-frequency details of photometric stereo with the comprehensive low-frequency information of light field imaging, leveraging their complementary strengths to yield robust results. The hybrid measurement achieves an angle error of approximately 10° or less, significantly outperforming light field measurements, while also providing compensation for low-frequency information in photometric stereo. Our method employs multispectral techniques to facilitate single-frame measurements and incorporates a deep learning-based DCN for uncalibrated photometric stereo reconstruction. Testing on the publicly available Diligent dataset demonstrates that our approach surpasses most existing mainstream methods. Additionally, we design and develop an experimental system based on hybrid measurement theory to achieve single measurements. Throughout the experiment, we utilize a light field camera as the same sensor for data acquisition, ensuring consistency in image registration and alignment and minimizing errors related to device variations. This method is straightforward and efficient, requiring no repeated image acquisition steps and utilizing fewer input images. It is easy to scale and can accommodate various numbers of input images for different task scenes. Additionally, it does not require light source vector input, making it suitable for measurements in scenes that are challenging to calibrate. However, our current work is still based on the measurement of Lambertian hypothesis, and the complex surface reflectance of objects has not been taken into account. It is also possible to further optimize the process, such as using a neural network for the entire reconstruction task. In the future, we will try to broaden the range of objects under test and further validate the effectiveness of our method. At the same time, we will try to learn from the Transformer network architecture to optimize our process.

Funding. National Natural Science Foundation of China (52375549); Nankai University Eye Institute (NKYKD202204, NKYKK202209); China Postdoctoral Science Foundation (Grant No.2022M721695); Enterprise R&D Special Project of Tiankai Higher Education Science and Technology Innovation Park (No.23YFZXYC00041); Fundamental Research Funds for the Central Universities, Nankai University (63241331).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. R. Woodham, "Photometric Method For Determining Surface Orientation From Multiple Images," Opt. Eng **19**(1), 139–144 (1980).
2. H. D. Tagare, *A theory of photometric stereo for a general class of reflectance maps* (1990).
3. A. Matsumoto, H. Saito, S. Ozawa, *et al.*, "3D reconstruction of skin surface from photometric stereo images with specular reflection and interreflection," Elect. Eng. Jpn. **129**(3), 51–58 (1999).
4. R. Basri, D. Jacobs, I. Kemelmacher, *et al.*, "Photometric stereo with general, unknown lighting," Int J Comput Vision **72**(3), 239–257 (2007).
5. T. Kuparinen and V. Kyrki, "Optimal Reconstruction of Approximate Planar Surfaces Using Photometric Stereo," IEEE Trans. Pattern Anal. Mach. Intell. **31**(12), 2282–2289 (2009).
6. J. Sun, M. Smith, L. Smith, *et al.*, "Examining the uncertainty of the recovered surface normal in three light photometric stereo," Image and Vision Computing **25**(7), 1073–1079 (2007).
7. C. Hernández, G Vogiatzis, R Cipolla, *et al.*, "Overcoming Shadows in 3-Source Photometric Stereo," IEEE Trans. Pattern Anal. Mach. Intell. **33**(2), 419–426 (2011).

8. D. Miyazaki and S. Hashimoto, "Uncalibrated photometric stereo refined by polarization angle," Opt. Rev. **28**(1), 119–133 (2021).
9. S. Hashimoto, D. Miyazaki, S. Hiura, *et al.*, "Uncalibrated photometric stereo constrained by intrinsic reflectance image and shape from silhoutte," in *16th IAPR International Conference on Machine Vision Applications (MVA)*, (2019).
10. Y. J. Sun, J. Dong, M. Jian, *et al.*, "Fast 3D face reconstruction based on uncalibrated photometric stereo," Multimedia Tools and Applications **74**(11), 3635–3650 (2015).
11. Y. M. Wang, Q. Zhang, W. Feng, *et al.*, "Shadow-aware Uncalibrated Photometric Stereo Network," in *14th International Conference on Computer and Automation Engineering (ICCAE), International Conference on Computer and Automation Engineering* (2022), 80–85.
12. M. Khanian, A. S. Boroujerdi, M. Breuß, *et al.*, "Photometric stereo for strong specular highlights," arXiv (2017).
13. A. R. Khadka, P. Remagnino, V. Argyriou, *et al.*, "Object 3D Reconstruction based on Photometric Stereo and Inverted Rendering," in *14th International Conference on Signal Image Technology & Internet Based Systems (SITIS)*, (2018), 208–215.
14. Y. Braun and H. Guterman, "Light invariant photometric stereo," Opt. Express **31**(4), 5200–5214 (2023).
15. D. Miyazaki and K. Uegomori, "Example-Based Multispectral Photometric Stereo for Multi-Colored Surfaces," J. Imaging **8**(4), 107 (2022).
16. H. Guo, Z. Mo, B. Shi, *et al.*, "Patch-Based Uncalibrated Photometric Stereo Under Natural Illumination," IEEE Trans. Pattern Anal. Mach. Intell. **44**(11), 7809–7823 (2022).
17. H. Santo, M. Samejima, Y. Sugano, *et al.*, "Deep Photometric Stereo Networks for Determining Surface Normal and Reflectances," IEEE Trans. Pattern Anal. Mach. Intell. **44**(1), 114–128 (2022).
18. S. Ikehata, "CNN-PS: CNN-Based Photometric Stereo for General Non-convex Surfaces," in *15th European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science* (2018), 3–19.
19. G. Y. Chen, K. Han, K. Y. K. Wong, *et al.*, "PS-FCN: A Flexible Learning Framework for Photometric Stereo," in *15th European Conference on Computer Vision (ECCV)*, Lecture Notes in Computer Science (2018), 3–19.
20. G. Huang, Z. Liu, L. V. Der Maaten, *et al.*, "Densely Connected Convolutional Networks," in *30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Computer Vision and Pattern Recognition* (2017), 2261–2269.
21. W. X. Fu, X. Tong, C. Shan, *et al.*, "Implementing Light Field Image Refocusing Algorithm," in *2nd International Conference on Opto-Electronics and Applied Optics (IEM OPTRONIX 2015)*, (2015).
22. X. M. Hu, Z. Li, L. Miao, *et al.*, "Measurement Technologies of Light Field Camera: An Overview," Sensors **23**(15), 6812 (2023).
23. H. X. Duan, J. Wang, L. Song, *et al.*, "Imaging Model and Calibration of Lytro Light Field Camera," in *2nd CCF Chinese Conference on Computer Vision (CCCV), Communications in Computer and Information Science* (2017), 389–400.
24. H. G. Jeon, J. Park, G. Choe, *et al.*, "Accurate Depth Map Estimation from a Lenslet Light Field Camera," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Computer Vision and Pattern Recognition* (2015), 1547–1555.
25. D. Cho, M. Lee, S. Kim, *et al.*, "Modeling the calibration pipeline of the Lytro camera for high quality light-field image reconstruction," in *IEEE International Conference on Computer Vision (ICCV), IEEE International Conference on Computer Vision* (2013), 3280–3287.
26. J. Jin, J. Hou, H. Yuan, *et al.*, "Learning Light Field Angular Super-Resolution via a Geometry-Aware Network," in *34th AAAI Conference on Artificial Intelligence / 32nd Innovative Applications of Artificial Intelligence Conference / 10th AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI Conference on Artificial Intelligence* (2020), 11141–11148.
27. J. Hu, L. Shen, G. Sun, *et al.*, "Squeeze-and-Excitation Networks," in *31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Computer Vision and Pattern Recognition* (2018), 7132–7141.
28. H. Fan, L. Qi, N. Wang, *et al.*, "Deviation correction method for close-range photometric stereo with nonuniform illumination," Opt. Eng **56**(10), 1 (2017).
29. B. X. Shi, Z. Mo, D. Duan, *et al.*, "A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Computer Vision and Pattern Recognition* (2016), 3707–3716.
30. Z. R. Li, Q. Qian, B. Pan, *et al.*, "DANI-Net: Uncalibrated Photometric Stereo by Differentiable Shadow Handling, Anisotropic Reflectance Modeling, and Neural Inverse Rendering," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Computer Vision and Pattern Recognition* (2023), 8381–8391.
31. G. Chen, M. Waechter, B. Shi, *et al.*, "What Is Learned in Deep Uncalibrated Photometric Stereo?" in *Computer Vision – ECCV 2020* Springer International Publishing, (2020), 745–762.
32. J. X. Li and H. D. Li, "Self-calibrating Photometric Stereo by Neural Inverse Rendering," in *17th European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science* (2022), 166–183.
33. N. G. Alldrin, S. P. Mallick, D. J. Kriegman, *et al.*, "Resolving the generalized bas-relief ambiguity by entropy minimization," in *IEEE Conference on Computer Vision and Pattern Recognition, IEEE Conference on Computer Vision and Pattern Recognition* (2007), 1822–+.

34. B. X. Shi, Y. Matsushita, Y. Wei, *et al.*, "Self-calibrating Photometric Stereo," in *23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Computer Vision and Pattern Recognition* (2010), 1118–1125.
35. Z. Wu and P. Tan, "Calibrating Photometric Stereo by Holistic Reflectance Symmetry Analysis," in *26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Computer Vision and Pattern Recognition* (2013), 1498–1505.
36. T. Papadhimitri and P. Favaro, "A Closed-Form, Consistent and Robust Solution to Uncalibrated Photometric Stereo Via Local Diffuse Reflectance Maxima," Int J Comput Vis **107**(2), 139–154 (2014).
37. F. Lu, X. Chen, I. Sato, *et al.*, "SymPS: BRDF Symmetry Guided Photometric Stereo for Shape and Light Source Estimation," IEEE Trans. Pattern Anal. Mach. Intell. **40**(1), 221–234 (2018).