Super-resolution computed tomography imaging spectrometry

LEI YUAN,1 QIANG SONG,2 HECONG LIU,1 KEVIN HEGGARTY,3 AND WEIWEI CAI1,*

1Key Laboratory for Power Machinery and Engineering of the Ministry of Education, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
2Lochn Optics, Shenzhen 518200, China
3Optics Department, IMT-Atlantique, Technopole Brest-Iroise, CS 83818, 29285 Brest, France
*Corresponding author: cweiwei@sjtu.edu.cn

Received 1 August 2022; revised 27 November 2022; accepted 27 November 2022; posted 1 December 2022 (Doc. ID 472072); published 23 January 2023

Computed tomography imaging spectrometry (CTIS) is a snapshot spectral imaging technique that relies on a limited number of projections of the target data cube (2D spatial and 1D spectral), which can be reconstructed via a delicate tomographic reconstruction algorithm. However, the restricted angle difference between the projections and the space division multiplexing of the projections make the reconstruction suffer from severe artifacts as well as a low spatial resolution. In this paper, we demonstrate super-resolution computed tomography imaging spectrometry (SRCTIS) by assimilating the information obtained by a conventional CTIS system and a regular RGB camera, which has a higher pixel resolution. To improve the reconstruction accuracy of CTIS, the unique information provided by the zero-order diffraction of the target scene is used as a guidance image for filtering to better preserve the edges and reduce artifacts. The recovered multispectral image is then mapped onto the RGB image according to camera calibration. Finally, based on the spectral and the spatial continuities of the target scene, the multispectral information obtained from CTIS is propagated to each pixel of the RGB image to enhance its spectral resolution, resulting in SRCTIS. Both stimulative studies and proof-of-concept experiments were then conducted, and the results quantified by key metrics, such as structural similarity index measurement and spectral angle mapping have suggested that the developed method cannot only suppress the reconstruction artifacts, but also simultaneously achieve high spatial and spectral resolutions. © 2023 Chinese Laser Press

https://doi.org/10.1364/PRJ.472072

1. INTRODUCTION

Digital cameras are ubiquitous in our daily life. For instance, they are the essential components of mobile phones and have been designed to accommodate a range of functions, such as photography, 3D time-of-flight (ToF) ranging and biometric recognition. As a sensing platform, more and more digital cameras tend to be integrated into the mobile phones to unlock more utilities. There are majorly two types of digital cameras, which are monochromatic/gray cameras and RGB cameras. In the latter, a Bayer filter array is integrated with the camera sensor chip to obtain three spectral channels. However, both types are incapable of imaging the scene with more spectral channels, i.e., multispectral imaging, which is crucial for a plethora of applications, such as combustion diagnostics [1–3], cancer detection [4], remote sensing [5,6], medical diagnostics [7,8], pollution detection [9], and agricultural applications [10,11]. Because of this, multispectral imaging has attracted enormous attention in the past decades. It essentially relates to the recovery of a data cube \((x, y, \lambda)\). The conventional multispectral camera relies on sequential sampling of the data-cube wavelength by wavelength, point by point, or line by line. There are also multispectral cameras that rely on Fourier transform infrared spectroscopy [12]. However, all these approaches require a mechanical scanning mechanism and, thus, suffer from a low temporal resolution. This disadvantage prevents the application of multispectral imaging from dynamic processes, such as turbulent combustion [13].

To overcome the aforementioned limitation, the so-called snapshot spectral imaging techniques have been developed based on which the data cube can be computationally recovered from a 2D spatially and spectrally encoded image captured within a single exposure. For example, Gehm et al. [14] proposed the so-called coded aperture snapshot spectral imaging (CASSI), which is established upon the theory of compressive sensing and encodes the spatial and spectral information with a combination of a random amplitude mask and a 1D disperser (e.g., grating/prism). However, CASSI typically requires that
the spectral datacube is sparse in both the spatial and the spectral domains. To relax the requirement for sparsity, Kittle et al. [15] developed multiexposure CASSI that can significantly improve the reconstruction accuracy at the cost of a reduced temporal resolution. However, this would comprise the applicability of CASSI to high dynamic scenarios. To improve the reconstruction fidelity of a single snapshot, Arguello et al. implemented an end-to-end model to design both a diffractive optical element and a color-coded aperture to generate shift-variant point spread functions (PSFs) so that the spatial-spectral modulation can be enhanced [16]. To increase the temporal resolution, Cao et al. [45] developed a hybrid camera system to achieve video-rate multispectral imaging. An array of dispersive spectrometers was created to uniformly and sparsely sample the multispectral pixels of the scene. Those pixels were then fused with a full RGB image of the same scene to recover the complete multispectral image. Later, similar systems with content-adaptive sampling were developed to increase system flexibility and measurement accuracy [17,18]. However, such systems rely on sparse spatial sampling and may encounter difficulties for scenes with complex spatial structures.

The so-called computed tomography imaging spectrometry (CTIS) is another type of snapshot spectral imaging technique that shares the same principle as X-ray CT, which is a commercially mature technique and enjoys enormous amount of reconstruction algorithms that have been developed during the past half century [19–21]. However, CTIS suffers from a low spatial resolution as the imaging sensor is divided into an array of regions, each of which contains the projection of the target data cube along a distinct angle. Also, there are severe artifacts in the reconstruction due to the limited number of projections and the minor angular difference between them. This is also referred to as the missing-cone problem in the literature [22].

In order to improve its performance, extensive efforts have been devoted to either optimizing the optical components or the reconstruction algorithms. In terms of optical components, most researchers focused on the optimal design of the disperser as it determines the number of projections, the angular difference, and the energy distribution among them. Descour [23] distributed effectively the incident spectrum over the detector array by designing the original disperser composed of three cross cosine gratings, which had the problem of misalignment between the gratings and low diffraction efficiency of the higher-order projections. Volin et al. [24] fabricated a single disperser, i.e., a computer generated hologram (CGH) for a midwave infrared computed-tomography spectrometer to improve the signal intensity of the higher orders and maintain uniform irradiance. Later, Scholl et al. [25] used singular value decomposition to optimize the diffraction efficiency of the disperser and designed two CGHs, each of which works at both the midwave and long-wave infrared ranges, resulting in the so-called dual-band CTIS system. Kudenov et al. [26] improved the spectral resolution and increased the light throughput into each projection by applying the division of aperture method to design a faceted radial disperser.

With respect to reconstruction algorithms, improving reconstruction accuracy and convergence speed is the major research interests. The most widely adopted reconstruction algorithms so far are the multiplicative algebraic reconstruction technique [27] and the expectation maximization (EM) [28], which have advantages including easy implementation and fast convergence [29]. Based on the noise source of the actual imaging system, Garcia and Dereniak [30] combined Poisson-distributed photon noise in the image and signal-independent system noise by using the standard maximum likelihood to design a mixed-expectation reconstruction technique that can effectively mitigate the influence of noise. An et al. [31] developed the subspace constraint algorithm to improve the spectral resolution of the reconstruction by reducing the nonuniqueness of the pseudoinverse solution. Later, Vose and Horton [32] extended the research of Garcia and Dereniak [30] by assuming that the system matrix has a shift-invariant structure to simplify the calibration process of PSFs and converting the result of the maximum likelihood estimator into an explicit closed form, which significantly improves the reconstruction accuracy and efficiency. Hagen et al. [33] proposed an optimized reconstruction algorithm based on the spatial shift invariance of the CTIS system, which greatly reduces the dimension of the system matrix and realizes image reconstruction with a near real-time response. Li et al. [29] developed a low-rank estimation [34] method for CTIS reconstruction by dividing the 3D data cube into numerous overlapping cubic patches, which can improve effectively the spatial and the spectral qualities of the reconstruction at the cost of increased computational complexity. By applying superiorization and guided image filtering (GIF) in the iterative process, Han et al. [35] effectively suppressed the artifacts and improved the reconstruction quality but failed to improve its spatial resolution. Wei et al. [36] proposed a method based on the 3D low-rank constraint and the tensor analysis to reduce the computational time and facilitate real-time applications of multispectral imaging. Despite the aforementioned progress, the severe artifacts caused by the limited number of projections and the inefficient utilization of the sensor array due to space division multiplexing of the projections have not been effectively solved. Therefore, the limitation in the spatial resolution and the severe reconstruction artifacts remain key issues limiting the application of CTIS.

In this paper, we develop a super-resolution CTIS (SRCTIS) by combining a conventional CTIS system and an RGB camera. The former can reconstruct a multispectral data cube with a low spatial resolution, and the latter can capture an RGB data cube with a high spatial resolution. In order to effectively assimilate the information of both data cubes, we first introduce GIF into each iteration of the CTIS reconstruction process to reduce severe artifacts due to the limited number of projections and angle span. The multispectral data cube is then mapped onto the RGB image through camera calibration. Finally, according to the spectral and the spatial continuity of the sought target, the multispectral information is propagated to each RGB pixel by applying the spectral propagation algorithm to obtain an image with a high resolution in both the spectral and the spatial domains. Different from aforementioned hybrid systems, which sample multispectral pixels at sparsely distributed locations, our hybrid system relies on CTIS which can recover the multispectral information of all pixels within a
continuous region. Thus, it can be applied to scenes with more complex spatial structures. In addition, it can effectively avoid the problem of metamerism. The details of the proposed technique along with the simulative studies and proof-of-concept experiments are discussed in the following sections.

2. MATHEMATICAL FORMULATION AND RECONSTRUCTION

A. Modeling of the Forward Process

Figure 1 illustrates the layout of the proposed SRCTIS system, which includes two arms, i.e., a conventional CTIS system and an RGB camera. The incoming spectral image is first filtered with a filter to truncate the spectral information outside of the target range. The image is then divided into two identical parts, each delivered along one arm. The CTIS arm consists of an objective lens, a collimation lens, a diffractive optical element (DOE), and a gray camera. The objective lens collects the filtered spectral image, and the collimation lens then converts the image into planar waves. The DOE is used to diffract each incoming monochromatic planar wave toward a $3 \times 3$ point array. The intensity distribution is uniform at the design wavelength.

Figure 2 illustrates the principle of SRCTIS for the generation of a data cube (i.e., a discretized version of the target spectral image) with both high spatial and spectral resolutions. The target data cube is diffracted toward the detector array along nine viewing angles, forming the same number of 2D projections. This forward imaging process can be mathematically described in a matrix form as

$$\vec{g} = H \cdot \vec{f}, \quad (1)$$

where $\vec{g}$ is a vector with a length of $M$ that contains all the measured pixel values of the CTIS camera, $\vec{f}$ is the vector with a length of $N$ that contains all the voxel values of the sought data cube, and $H$ is an $M \times N$ weight matrix that linearly approximates the forward imaging process.

B. Reconstruction Algorithm

Typically Eq. (1) is an ill-posed linear equation system, which can be solved with numerous mathematical techniques, such as optimization, iterative reconstruction, and machine learning algorithms. The gradient-based optimization methods are efficient but can be easily trapped in local minima; and the global optimizers, such as simulated annealing [37,38] and genetic algorithms [39], suffer from formidable computational costs. In addition, the machine learning algorithms usually require a large set of high-quality training data samples, which are not readily available [40–42]. Due to the ease of implementation and good performance for limited-projection tomography, a well-established iterative algorithm, i.e., the maximum likelihood expectation maximization (MLEM) [43], is adopted, and its iteration process can be described as

$$f_j^{(k+1)} = \frac{\sum_{i=1}^{M} (H_{ij} \sum_{l=1}^{N} (H_{lj} f_l^{(k)}))}{\sum_{i=1}^{M} H_{ij} f_j^{(k)}}, \quad (2)$$

where the superscript $k$ indicates the number of iterations, $j$ indicates the index of the voxel in vector $\vec{f}^k$, $H_{ij}$ indicates an element of the weight matrix $H$ at the $i$th row and $j$th column, and $g_i$ represents the $i$th element of vector $\vec{g}$.

As can be seen from Fig. 2, the CTIS branch can recover more spectral details but has a lower spatial resolution, whereas, the RGB branch can capture more spatial information but with only three broadband spectral channels. The assimilation of the two data cubes can lead to the enhancement of the resolution in both the spectral and the spatial dimensions. For this purpose, we propose a hybrid algorithm that combines tomographic reconstruction aided with GIF [44] and a spectral propagation algorithm [45]. The first step of the method is a regular CTIS reconstruction but with an additional GIF step added within each MLEM iteration, and the corresponding flow chart is shown in Fig. 3. The reconstructed multispectral data cube is then mapped to the RGB data cube according to the geometrical relationship established based on camera calibration.

**Fig. 1.** Schematic of the SRCTIS system. The input image is split into two by a beam splitter and collected by an RGB camera and a conventional CTIS system, respectively. In the CTIS branch, after collimation, the input image is diffracted by the DOE and received by a gray camera.
Finally, the spectral resolution of the RGB image is enhanced by propagating the multispectral details mapped to the specific RGB pixels to their contiguous RGB pixels through the spectral propagation algorithm, which is guided by the continuity in both the spatial and the spectral domains. The continuity condition plays a critical role for the successful implementation of the method. However, the conventional CTIS reconstruction usually suffers from severe artifacts mainly due to the insufficient number of projections and their minor angular difference, which would jeopardize the continuity condition and undermine the subsequent propagation process. To mitigate the limitations and take the full advantage of the information provided by the zero-order diffraction of the CTIS system, a smoothing operation, i.e., GIF, is introduced into each iteration step to better preserve the edge information in the spatial domain. The filtered image $q$ is regarded as the linear transformation of the guidance image $I$ in a local square window $\omega_k$ (represented as a set of pixel indices) of size $2r + 1$ centered at pixel $k$ [44] and can be expressed as

$$q_i = a_k I_i + b_k, \quad \forall i \in \omega_k,$$

(3)

where the size of filtered image $q$ and the guidance image $I$ is $m \times n$, $i$ is the pixel index, $a_k$ and $b_k$ are the constant coefficients of the local square window and are determined by minimizing the difference between the input image $p$ with a dimension of $m \times n$ and the filtered output image $q$, which can be calculated as

$$E(a_k, b_k) = \sum_{i \in \omega_k} [(q_i - p_i)^2 + \epsilon a_k^2],$$

(4)

$$a_k = \frac{1}{|\omega_k|} \sum_{i \in \omega_k} (I_i \hat{p}) - \mu \hat{p}/(\sigma^2 + \epsilon),$$

(5)

**Fig. 2.** Principle of SRCTIS for reconstruction of a data cube with both high spatial and spectral resolutions. First, the CTIS reconstruction and an RGB image are obtained separately; then the CTIS multispectral pixels and the RGB image pixels are aligned by position calibration; and finally, a spectral propagation algorithm is used to fuse the two images.

**Fig. 3.** Flowchart of the MLEM algorithm with GIF. After each MLEM iteration, GIF is applied to suppress the reconstruction artifacts.
where $|\omega_k|$ is the total number of pixels in the local window $\omega_k$, $\sigma^2$ and $\mu_k$ indicate the variance and the mean of the guidance image $I$ in the window $\omega_k$, $\hat{p}_k$ is the mean of the filtered image $p$ in the local window $\omega_k$, and $\epsilon$ is a regularization parameter, which balances the posterior, i.e., the measurements and a priori information, i.e., smoothness condition. Generally, parameter $\epsilon$ regulates the pixel intensity variation at different areas and affects the filtering effect of the image. Its value should be larger than 0 and smaller than 1. As can be seen from Eq. (4) when $\epsilon = 0$, $a_k = 1$, $b_k = 0$, and $E(a_k, b_k) = 0$. In this case, GIF does not work, and the output image is the same as the original input. If $\epsilon > 0$, $a_k$ is close to 0 and $b_k$ is close to $\hat{p}_k$ in an area with a small pixel intensity variation. In this case, the function of GIF is equivalent to weighted mean filtering. In an area where the pixel intensity varies significantly, $a_k$ is close to 1, $b_k$ is close to 0, and the filtering effect is weak at this time, which helps to maintain the edge. Under the condition of a fixed window size as $\epsilon$ increases the filtering effect becomes stronger. Thus, the value of $\epsilon$ should be determined according to different application scenarios. In our case, we find the reconstructions are good and insensitive to $\epsilon$ when it is selected between 0.01 and 0.1. It can be seen from Eq. (3) that the selection of the guidance image directly affects the filtering effect since the filtered image $q$ is its linear transformation. The edge information of $q$ can be preserved as long as $I$ contains accurate edge information. The zero-order diffraction at the center among the nine projections of CTIS is not smeared in either the spectral or in the spatial domains. It is essentially an image taken by a gray camera, which perfectly reserves detailed edge information of the target scene, making it an ideal guidance image. Thus, the reconstruction artifacts of CTIS that cause edge smearing due to the limited projection angle span can be suppressed to the maximum extent, significantly improving the reconstruction quality.

The second step of the hybrid algorithm is to align the multispectral image recovered by the CTIS branch to the RGB image. Considering the different magnifications of the two branches, a pixel in the multispectral image essentially corresponds to a certain area in the RGB image. To make an accurate matching, we first approximately determine the center positions of the multispectral pixels on the RGB image according to the camera calibration. Then, the squares, each centered at those positions with a side length of magnification, are considered to be the corresponding area for each multispectral pixel in the RGB image as indicated by a green square in Fig. 4(b). Furthermore, we take the spectral angle mapping (SAM) [46] as the quantitative indicator for positioning of the multispectral pixels, which is commonly applied to measure the similarity between two spectra. SAM treats the spectra as vectors whose dimensions equal the number of spectral bands and can be mathematically described as

$$D_{\text{SAM}} = \arccos \left( \frac{\hat{S}_1 \cdot \hat{S}_2}{(\hat{S}_1 \cdot \hat{S}_1)^{1/2}(\hat{S}_2 \cdot \hat{S}_2)^{1/2}} \right),$$

where $\hat{S}_1$ and $\hat{S}_2$ are two spectra represented as vectors.

In order to determine the exact position of the multispectral pixel on the RGB image, we first convert the multispectral spectrum into an RGB vector by integrating the spectrum over the RGB camera response curves [see Fig. 4(a)]. Then, the resulting RGB vector is used to calculate the SAM values for all the RGB pixels within the green square. The position where the SAM value is minimum is taken as the position of the multispectral pixel in the RGB image. Figure 4(b) illustrates the distribution of the multispectral pixels on the RGB image. The distribution of blue pixels represents the actual positions of the CTIS multispectral pixels on the RGB image after matching the pixel points. The yellow pixels represent the remaining RGB pixels for which the multispectral information can be inferred from the neighboring blue pixels based on the spectral and spatial continuity conditions using a spectral propagation algorithm [45].

The spectral propagation algorithm takes into account not only the distance between the pixels in the spatial domain, but also the distance in the RGB domain. Since each RGB pixel has three values, the spectral propagation is conducted for each of the RGB channels. As illustrated in Fig. 4(b), assuming the spatial dimension of the RGB image is $p \times q$ and that of the corresponding multispectral image is $m \times n \times L(m < p, n < q, L \gg 3)$, then, the multispectral information of a specific pixel (i.e., a yellow one) in the RGB image can be calculated from the neighboring multispectral pixels (i.e., blue pixels) as

$$\tilde{S}_{ij} = \sum_{c \in R, G, B} \frac{\sum_{k \in \Omega} (d_{c_{RGB}}^k)^2 (d_{c_{RGB}}^k)^2}{\sum_{k \in \Omega} (d_{c_{RGB}}^k)^2} \cdot \tilde{S}_k,$$

where $\tilde{S}_{ij}$ represents the vector of the spectrum corresponding to pixel (i, j) in the RGB image, $k$ indicates the multispectral pixels within the neighborhood $\Omega$ centered on pixel (i, j), and $g_{\sigma}$ denotes the Gaussian operator with a standard deviation $\sigma$ and zero mean where the subscripts $r$ and $s$ denote the RGB and geometric space, respectively. $d_{c_{RGB}}^k$ and $d_{c_{RGB}}^s$, respectively, represent the Euclidean distance between two pixels (i, j) and k in the RGB vector space and the geometric space. $\rho_k$ denotes the weighting factor which represents the ratio of the corresponding value in a given color channel (indicated by the subscript $c$) between pixels (i, j) and k (e.g., $\rho_k = G_k/G_i$ for the green channel) and is applied to match the intensity of the two pixels at different channels. For each channel, $\tilde{S}_k$ is calculated as

$$\tilde{S}_k = \tilde{S}_k \otimes \tilde{w}',$$

where

$$\tilde{w}' = \frac{q_i}{\sum_{c} q_c},$$

and $\tilde{S}_k$ is the spectrum of pixel $k$ in $\Omega$, $q'_c$ represents the corresponding transmission of the filter $c$ at wavelength $\lambda$, $\tilde{w}'$ represents the relative response efficiency of filter $c$, and $\otimes$ represents the element-wise multiplication. The complete spectrum vector, i.e., $\tilde{S}_{ij}$ can be obtained by summing the three channels after obtaining multispectral data for each channel.

Based on the above spectral propagation principle, it is obvious that the spectral information transferred from each multispectral pixel to the target RGB pixel is determined by the
spatial distance between them, the similarity of RGB vectors, and the intensity difference between the two pixels in the corresponding channel. The spatial distance and the RGB vector difference determine the weight factors $g_{\sigma_1}(d_{xy})$ and $g_{\sigma_2}(d_{\text{RGB}})$ in Eq. (8). The larger the spatial distance is, or the greater the RGB vector difference is, the smaller the corresponding weight factor would be, indicating that the spectral difference between the target RGB pixel and the multispectral pixel is greater, and the spectral component of the corresponding multispectral pixel propagating to the RGB pixel is smaller. When the multispectral pixel is very close to the target RGB pixels, their spectral information is very similar due to the continuity of the spectral domain, and it is unlikely that there is the metamerism problem. When the multispectral pixel is far away from the target RGB pixel, although there may be the metamerism problem, the corresponding weight factor $g_{\sigma_1}(d_{xy})$ is very small due to the long spatial distance, which means that the spectral information of the multispectral pixel contributes little to the spectral information of the target RGB pixel. Thus, our method can effectively avoid the metamerism problem and the corresponding registration artifacts.

3. RESULTS AND DISCUSSION

To demonstrate the reliability and superiority of our proposed SRCTIS system, we compare it to the conventional CTIS system with both numerical simulations and proof-of-concept experiments. In this paper, MLEM is adopted to invert the 2D CTIS diffraction patterns to recover a multispectral data cube, which has a low spatial resolution. To illustrate the effectiveness of GIF in suppressing the reconstruction artifacts of CTIS as well as the capability of SRCTIS in obtaining multispectral images with a high spatial resolution, we took the reconstructions from MLEM, MLEM with GIF (denoted as MLEM_GIF), and SRCTIS for comparison. In addition to qualitative analysis, we performed a detailed comparison between the algorithms in terms of quantitative image quality metrics, such as SAM, average normalized root-mean-square error (RMSE) [47], average peak signal-to-noise ratio (PSNR), and average structural similarity index measurement (SSIM) [47]. SAM quantifies the similarity between two spectra, and a smaller SAM value indicates a better spectral fidelity. RMSE is adopted to quantify the difference between the ground truth and the reconstructed data cube. A smaller RMSE suggests a closer reconstruction to the ground truth. On the other hand, PSNR and SSIM are mainly used to quantify the consistency between the spatial details of the reconstruction and that of the ground truth. PSNR is a quantitative indicator commonly used to evaluate quality of image reconstruction, and SSIM reflects the similarity between the ground truth and the reconstruction by fully considering the image composition, such as brightness, contrast, and structure.

For numerical validation, we randomly selected 30 representative scenes with distinct spatial and spectral features from three data sets, i.e., cave dataset [48], hyperspectral images for local illumination in natural scenes 2015 dataset [49], and ICVL hyperspectral dataset [50]. The scenes in these data sets all have spectral resolutions of 10 nm but have different spectral ranges of 400–700 nm, 400–720 nm, and 400–1000 nm, respectively. The spectral range of the selected scenes is tailored to be 420–700 nm, and a wavelength step of 5 nm is achieved by linear interpolation in the spectral dimension. To mimic the practical situations, we artificially added Gaussian noise to the simulated CTIS projections to obtain an SNR of 25–50 dB. The dimension of the reconstructed data cube is assumed to be 250 × 250 × 57. All the simulations were implemented with Intel visual FORTRAN and ran on a workstation equipped with Intel Xeon CPU E5-2630 v4. Figure 5 compares the reconstructions of MLEM and MLEM_GIF from both qualitative and quantitative perspectives. Figure 5(a) qualitatively presents the spatial details of both the reconstructions and the corresponding ground truths of different scenes for which only the image at 460 nm is shown to better illustrate the anti-artifact function of GIF at a specific spectral band. Visually, the multispectral images reconstructed by
MLEM_GIF are closer to the ground truths in spatial details and do not have obvious artifacts and noise in contrast to that reconstructed from MLEM. From a quantitative point of view, MLEM_GIF has larger PSNR and SSIM values and a smaller SAM value compared with MLEM, indicating that the application of GIF can effectively improve reconstruction fidelity.

In order to assess the performance of GIF in suppressing artifacts under different noise levels, we further analyzed the variations of the quantitative image quality metrics of MLEM and MLEM_GIF when the noise level changes from 25 to 50 dB as shown in Fig. 5(b). Overall speaking, the reconstruction quality of MLEM_GIF is consistently better than that of MLEM under different noise levels, specifically, the PSNR and SSIM values of MLEM_GIF are larger, and the RMSE and SAM values are smaller compared with those of MLEM. As the noise level increases, the reconstruction quality of MLEM_GIF maintains at a high level, whereas, that of MLEM decreases significantly. Thus, GIF can effectively suppress artifacts and is robust against the noise.

For the need of SRCTIS simulation, we further constructed 2D diffraction patterns and the high-spatial-resolution RGB image for the CTIS and the RGB branch, respectively. On one hand, we down-sampled the multispectral data cubes of size 1000 × 1000 × 57, obtained from the database by moving average method [51] to a size of 200 × 200 × 57. Then, the 2D diffraction pattern is simulated according to the diffraction efficiency and diffraction angle of the DOE, and 25-dB Gaussian noise is then added to the pattern to better simulate the practical scenarios. On the other hand, the high-spatial-resolution (1000 × 1000 pixels) RGB image is simulated by integrating the multispectral image over the spectral response curve of each RGB channel. The size of the multispectral images reconstructed by MLEM and MLEM_GIF is 200 × 200 × 57, whereas the size of SRCTIS is 1000 × 1000 × 57. Thus, SRCTIS can enhance the spatial dimension of the reconstructed data of MLEM and MLEM_GIF by 25 times.

Figure 6 shows the spatial details of both the reconstructions from different algorithms and the corresponding ground truths. In order to better compare the spatial details among MLEM, MLEM_GIF, and SRCTIS, all reconstructions are integrated over each of the RGB channels and then presented in the RGB form, and the local details are enlarged and placed in the upper left corner of each image. In the CTIS branch, as the input image is diffracted toward nine different angles after passing the DOE; the spatial resolution is essentially sacrificed for an improved spectral resolution. Therefore, it can be seen from Fig. 6 that MLEM reconstruction obviously loses considerable spatial details compared with the ground truth, and the relatively lower PSNR and SSIM values further suggest its reduced reconstruction quality in the spatial dimension from a quantitative point of view. For MLEM_GIF, although it can effectively suppress the artifacts in the CTIS reconstructed image and improve the reconstruction quality, it basically has the same discretization with MLEM and cannot increase the spatial resolution of the reconstructed image. On the contrary, as can be clearly seen from the local magnification of images recovered by SRCTIS, they have almost identical spatial details to the ground truth, which is far superior to MLEM and MLEM_GIF. It is worth noting that when calculating the quantitative image quality metrics, SRCTIS takes the original image with a spatial size of 1000 × 1000 as the reference, whereas, MLEM and MLEM_GIF take the down-sampled image with a spatial size of 200 × 200 as the reference. Although the PSNR and SSIM values of the SRCTIS reconstructed images are slightly smaller than those of MLEM_GIF, they are still larger than those of MLEM as can be seen from

Fig. 5. (a) Ground truth and the corresponding reconstructions of different scenes (denoted in sequence from left to right as "flowers," "oil painting," "cloth," and "toys") from MLEM and MLEM_GIF with the quality metrics PSNR/SSIM/SAM labeled at the bottom of each reconstructed image; (b) the relationships between the noise level and the different reconstruction quality metrics.
Fig. 6, suggesting that SRCTIS effectively preserves the spatial details of the RGB images.

In addition to the spatial details, we also randomly selected three points (marked A, B, and C) from different scenes as shown in Fig. 6 to assess the spectral reconstruction quality, and the relevant results are presented in Fig. 7(a). As can be seen, the spectra reconstructed by MLEM deviate significantly from the ground truth (indicated by a large SAM value) and are accompanied by considerable noise. However, the reconstructed spectra of MLEM_GIF closely approximate the ground truth, and the SAM value is also far less than that of MLEM, further suggesting that GIF cannot only suppress the artifacts, but also improves the spectral quality. For SRCTIS, although the SAM value is slightly reduced compared with MLEM_GIF after propagation, the recovered spectrum of SRCTIS is basically consistent with that of MLEM_GIF and in good agreement with the ground truth. The quality of the spectrum is much higher than that of MLEM. Figure 7(b) presents the quantitative image quality metrics and the computational time corresponding to the reconstruction results of 10 different scenes with a noise level of 25 dB. It can be seen that in all cases, the quantitative metrics of the multispectral images obtained by SRCTIS are similar to those of MLEM_GIF and are better than that of MLEM, fully demonstrating the superiority of SRCTIS. In terms of computational time, it can be seen that SRCTIS only needs slightly longer computational time than MLEM_GIF.

In order to further validate the effectiveness of SRCTIS, we built a hybrid system according to the layout shown in Fig. 1 for a proof-of-concept experimental demonstration. A long-wave pass filter with a cutoff wavelength of 425 nm and a short-wave pass filter with a cutoff wavelength of 650 nm are combined to tailor the spectral information of the target scenes. The DOE is a binary phase element, etched into a layer (thickness of $\sim$410 nm) of spin-coated S1813 photoresist (MicroChem) on 1.5-mm thick borosilicate glass substrate using a custom-built massively parallel-write photoplotter [52,53]. The light efficiency of the CTIS branch and the RGB branch is 30.23% and 48.38%, respectively. The scenes were displayed on a monitor to make it convenient to switch
between different images and measure the corresponding ground truths with a spectrometer (Ocean Optics HR4000). The spectrometer used to collect the ground truth [denoted by the dark solid line in Fig. 9(c)] is the Ocean Optics HR4000. In this proof-of-concept experiment, the wavelength step of the images captured by the CTIS branch is 5 nm, and the spatial size is $174 \times 174$ pixels. The images produced by SRCTIS have a spatial dimension of $947 \times 947$ pixels, and the spectral resolution and range are the same as the CTIS reconstruction results.

We quantitatively analyzed the spatial and spectral resolutions of SRCTIS, and the corresponding results are shown in Fig. 8. In order to assess the spatial resolution of these methods, a USAF 1951 resolution test chart was adopted in our experiment. The spatial resolution of CTIS, SRCTIS, and the simple up-scale of CTIS using the nearest-neighbor interpolation (NNI) [54], and the RGB camera were compared against each other using the modulation transfer function (MTF) as shown in Fig. 8(a). As can be seen from the figure, the MTF of NNI largely overlaps with that of CTIS, suggesting that the direct application of a simple up-scale method [54] cannot improve the spatial resolution. Also, the MTF of SRCTIS is very close to that of the RGB camera, indicating that SRCTIS can effectively inherit the spatial resolution of the latter. These results imply that SRCTIS has a significant advantage in spatial resolution over conventional CTIS. According to the Rayleigh criterion and taking 20% contrast as the threshold, the spatial resolutions of CTIS, SRCTIS, NNI, and the RGB camera are determined to be 1.63 lp/mm, 6.46 lp/mm, 1.63 lp/mm, and 6.77 lp/mm, respectively. In addition, to quantify the spectral resolution, these methods were tested by reconstructing a uniform square region whose spectrum contains two adjacent peaks that are separated with different spacing. Figures 8(b) and 8(c) present the reconstructed spectra with minimum spac-
and a slightly lower spectral resolution than CTIS. Thus, SRCTIS can achieve a much better spatial resolution than CTIS are similar, i.e., 12 nm and 10 nm, respectively. The results show that the spectral resolutions of SRCTI and CTIS are 12 nm and 10 nm, respectively. The peaks are assumed to be resolved if the valley between the peaks is smaller than 50% of the peak values. Panel (a) presents the reconstructed spectrum of CTIS where the reference spectrum contains two peaks at 495 nm and 505 nm, respectively; panel (c) presents the reconstructed spectrum of SRCTIS where the reference spectrum contains two peaks at 493 nm and 505 nm, respectively.

Quantitative comparison of different algorithms in terms of spatial and spectral resolutions. Panel (a) presents the modulation transfer function corresponding to CTIS, SRCTIS, CTIS with the nearest-neighbor interpolation, and the RGB camera; panel (b) presents the reconstructed spectrum of CTIS where the reference spectrum contains two peaks at 495 nm and 505 nm, respectively; panel (c) presents the reconstructed spectrum of SRCTIS where the reference spectrum contains two peaks at 493 nm and 505 nm, respectively.

The results from both the simulative and the proof-of-concept experiments suggested that SRCTIS can inherit the advantages of both the CTIS and the RGB branches. For the CTIS system, the complementary information was collected from both the CTIS and the RGB branches. For the CTIS reconstruction, we took full advantage of the zero-order pattern as the guidance image for filtering within each iterative step to effectively suppress the reconstruction artifacts and improve the reconstruction quality. After mapping the multispectral information reconstructed to the RGB image, which had a finer spatial resolution, we further propagated the multispectral information to each pixel in the RGB image based on the continuity condition in both the spectral and the spatial dimensions. The results from both the simulative and the proof-of-concept experiments suggested that SRCTIS can inherit the advantages of both the CTIS and the RGB branches and fulfill the high-quality reconstruction performance. Compared with the conventional CTIS system, the SRCTIS system could not only effectively suppress artifacts, but also greatly improve the spatial resolution, whereas, maintaining a high spectral fidelity.

In this paper, a super-resolution CTIS, which was capable of capturing images with both high spatial and spectral resolutions, has been developed and verified by both numerical simulations and proof-of-concept experiments. In the SRCTIS system, the complementary information was collected from both the CTIS and the RGB branches. For the CTIS reconstruction, we took full advantage of the zero-order pattern as the guidance image for filtering within each iterative step to effectively suppress the reconstruction artifacts and improve the reconstruction quality. After mapping the multispectral information reconstructed to the RGB image, which had a finer spatial resolution, we further propagated the multispectral information to each pixel in the RGB image based on the continuity condition in both the spectral and the spatial dimensions. The results from both the simulative and the proof-of-concept experiments suggested that SRCTIS can inherit the advantages of both the CTIS and the RGB branches and fulfill the high-quality reconstruction performance. Compared with the conventional CTIS system, the SRCTIS system could not only effectively suppress artifacts, but also greatly improve the spatial resolution, whereas, maintaining a high spectral fidelity.

Finally, we would like to point out some future research directions to further improve our method. The optimization of DOE effectively improved the intensity uniformity between different diffraction orders at the entire working wavelength range and improved the transmission efficiency of DOE, so as to obtain a more stable and well-behaved weight matrix and reduce the impact of noise to make reconstruction more accurate. In addition, the application of super-resolution and advantage of GIF. It is worth noting that there are also the mosaiclike artifacts in the MLEM reconstruction, which are caused by dividing the target scene into 5 × 5 regions and using different PSFs for each region in order to improve the quality of the spectral reconstruction. Nevertheless, it can be seen from the results of MLEM_GIF and SRCTIS that this problem can be also effectively suppressed by GIF. For SRCTIS, the rich spatial details of the high spatial resolution RGB image are fully inherited. As can be seen from Fig. 9(a) that the spatial details of the reconstruction from SRCTIS are far richer than those from MLEM and MLEM_GIF. Moreover, a few reconstructed slices from SRCTIS at different wavelength bands of the hills scene are presented in Fig. 9(b), and the images are exhibited in colors corresponding to respective wavelengths.

In addition, the reconstructions from SRCTIS also have a higher spectral fidelity compared with those from MLEM. Four random points (marked A, B, C, and D) in different scenes are shown in Fig. 9(a) are selected and their corresponding spectra are shown in Fig. 9(c). It can be seen that, although the SRCTIS results are derived based on the MLEM_GIF reconstructions, the spectral information is not severely impaired by the propagation process and closely approximates the ground truth. Compared with MLEM, its spectra can better capture the variations and peak positions of the real spectrum. These results firmly suggest that SRCTIS can effectively inherit the spectral details of CTIS reconstructions optimized by GIF. The results from both the simulative and the proof-of-concept experiments suggested that SRCTIS can inherit the advantages of both the CTIS and the RGB branches and fulfill the high-quality reconstruction performance. Compared with the conventional CTIS system, the SRCTIS system could not only effectively suppress artifacts, but also greatly improve the spatial resolution, whereas, maintaining a high spectral fidelity.

4. CONCLUSIONS AND OUTLOOK

In this paper, a super-resolution CTIS, which was capable of capturing images with both high spatial and spectral resolutions, has been developed and verified by both numerical simulations and proof-of-concept experiments. In the SRCTIS system, the complementary information was collected from both the CTIS and the RGB branches. For the CTIS reconstruction, we took full advantage of the zero-order pattern as the guidance image for filtering within each iterative step to effectively suppress the reconstruction artifacts and improve the reconstruction quality. After mapping the multispectral information reconstructed to the RGB image, which had a finer spatial resolution, we further propagated the multispectral information to each pixel in the RGB image based on the continuity condition in both the spectral and the spatial dimensions. The results from both the simulative and the proof-of-concept experiments suggested that SRCTIS can inherit the advantages of both the CTIS and the RGB branches and fulfill the high-quality reconstruction performance. Compared with the conventional CTIS system, the SRCTIS system could not only effectively suppress artifacts, but also greatly improve the spatial resolution, whereas, maintaining a high spectral fidelity.

Finally, we would like to point out some future research directions to further improve our method. The optimization of DOE effectively improved the intensity uniformity between different diffraction orders at the entire working wavelength range and improved the transmission efficiency of DOE, so as to obtain a more stable and well-behaved weight matrix and reduce the impact of noise to make reconstruction more accurate. In addition, the application of super-resolution and
denoising neural networks in CTIS reconstruction effectively improved the efficiency as well as simplified the structure of the entire optical system.

**Funding.** National Natural Science Foundation of China (51976122, 52061135108).

**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**


