Spectral Clustering Super-Resolution Imaging Based on Multispectral Camera Array

Feng Huang, Yating Chen, Xuesong Wang, Shu Wang, and Xianyu Wu

***Abstract*— Although multispectral and hyperspectral imaging acquisitions are applied in numerous fields, the existing spectral imaging systems suffer from either low temporal or spatial resolution. In this study, a new multispectral imaging system—camera array based multispectral super resolution imaging system (CAMSRIS) is proposed that can simultaneously achieve multispectral imaging with high temporal and spatial resolutions. The proposed registration algorithm is used to align pairs of different peripheral and central view images. A novel, super-resolution, spectral-clustering-based image reconstruction algorithm was developed for the proposed CAMSRIS to improve the spatial resolution of the acquired images and preserve the exact spectral information without introducing false information. The reconstructed results showed that the spatial and spectral quality and operational efficiency of the proposed system are better than those of a multispectral filter array (MSFA) based on different multispectral datasets. The PSNR of the multispectral super-resolution images obtained by the proposed method were respectively higher by 2.03 and 1.93 dB than those of GAP-TV and DeSCI, and the execution time was significantly shortened by approximately 54.55 s and 9820.19 s when the CAMSI dataset was used. The feasibility of the proposed system was verified in practical applications based on different scenes captured by the self-built system.**

***Index Terms*— Snapshot multispectral camera array, computational imaging, hierarchical clustering, multiple local-geometric transformations, spectral clustering super-resolution, adaptive kernel.**

I. Introduction

CONVENTIONAL color cameras only acquire images of targets by using three broad spectral bandwidths (red, green, and blue). Although these three spectral bandwidths are sufficient to meet the needs of human vision, higher spectral resolution images are needed to detect information that cannot be directly observed by human vision. For example, multispectral imaging can acquire dozens of spectral images of targets. These are extensively used in biomedical imaging, medical diagnostics, and remote sensing [1], [2], [3], [4], [5], [6]. Multispectral or hyperspectral imaging technologies can be categorized into spatial-scanning or spectral-scanning spectrometers, snapshot multispectral imaging systems, and computational multispectral imaging systems. Spatial-scanning multispectral imaging systems collect all the spectral information from a single point or a single spatial line of the target scene [7]. Spectral-scanning multispectral imaging systems were created based on bandpass filter wheels or tunable filters, such as liquid crystal tunable filters (LCTFs) and acousto-optical tunable filters (AOTFs). Hence, the imaging quality of the spatial-scanning and spectral-scanning spectrometers is time sensitive [8]. Scanning-based multispectral imaging systems require multiple exposures to capture complete multispectral information of the target scene, thus hindering the application of multispectral imaging systems to dynamic scenes and high-speed targets [9], [10], [11]. To overcome the shortcomings of the scanning-based multispectral imaging techniques, snapshot multispectral imaging techniques and computational spectral imaging techniques are proposed to acquire the multispectral images of dynamic scenes. The snapshot multispectral imaging systems includes snapshot multispectral imaging sensors, multi-aperture multispectral imaging systems, and a coded aperture snapshot spectral imager. The coded aperture snapshot spectral imager comprises a binary coded aperture and a disperser based on a single dispersive element and employs compressive sensing methods to restore spectral images [12]. Arce's group replaced the traditional random binary coded apertures with colored coded apertures to improve the compressive capabilities of CASSI; this was achieved by optimizing the colored-coded apertures to minimize the number of projections and maximize the reconstruction quality [13], [14], [15]. The multi-aperture imaging systems can be modified for spectral imaging by using bandpass filters, band-stop filters, continuous variable filters, or the linear variable spectral filter. Recently, Hubold et al. placed a slanted linear variable spectral filter near the entrance pupil of the microlens array, multi-aperture imaging system to realize snapshot multispectral imaging [16]. The aforementioned two types of snapshot multispectral imaging consist of a dispersive element or variable filter, which requires a sophisticated optical system to focus multi-bands on a single focal plane array. Genser et al. built a multi-aperture, camera array imaging system for multispectral imaging using bandpass filters (CAMSI), wherein each aperture was equipped with a separate sensor [17]. There no non-redundant spatial motion information between images captured by different apertures in the CAMSI. Therefore, the captured multispectral images cannot be used for super-resolution image reconstruction. Limited by the transmission bandwidth of the

Manuscript received 2 August 2022; revised 1 December 2022; accepted 27 January 2023. Date of publication 9 February 2023; date of current version 15 February 2023. This work was supported in part by Fuzhou University under Grant 2019T009 and Grant GXRC-18066; and in part by the Department of Education, Fujian Province, under Grant JAT190005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Daniel L Lau. *(Corresponding author: Xianyu Wu.)*

The authors are with the School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China (e-mail: xwu@fzu.edu.cn).

Digital Object Identifier 10.1109/TIP.2023.3242589

system, CAMSI is not conducive to multi-spectral imaging with high spatial resolution. To align multispectral images captured by different apertures, Genser et al. [17] proposed a new registration and reconstruction algorithm based on the spatial symmetry of the aperture layout and correlation of different spectra. However, the registration algorithm requires a system with superior assembly accuracy. The snapshot multispectral sensors directly capture multispectral images using a division-of-focal-plane micro bandpass filter array. Geelen et al. [18] and Kanaev et al. [19] demonstrated a 16-band, visible spectrum, snapshot multispectral sensor, and 9-band near-infrared snapshot multispectral sensor, respectively. As the spectral data captured by a snapshot multispectral sensor are multiplexed onto a single imaging sensor, the spatial resolution of each spectral image was low. To improve the spatial resolution of the spectral image captured by snapshot multispectral imaging sensors, dozens of demosaicking algorithms have been proposed and demonstrated, including the edge-sensing interpolation algorithm [20], adaptive residual interpolation algorithm [21], [22], [23], [24], and compressed sensing algorithms [25], [26]. However, when multispectral imaging sensors are designed using a large number of spectral bands, the reconstruction effectiveness and imaging quality of most demosaicking algorithms become limited. This is owing to the image mechanism of the snapshot multispectral imaging sensors that result in low-spatial resolution, single-band spectral images.

In this study, camera array based multispectral super resolution imaging system (CAMSRIS) is proposed; this system adopts bandpass filters to capture spectral information and exploits nonredundant spatial motion information among apertures to reconstruct super-resolution images. CAMSRIS can simultaneously achieve multispectral imaging with high temporal and spatial resolutions. The flexibility of the system in bandpass filter and focal plane array selection is suitable for the realization of low-cost CAMSRIS. Given that the geometrical rectification between pairs of apertures is expressed by multiple local-geometric transformations, and given that different band images have different texture structures in real scenes, the coarse-to-fine registration algorithm based on the hierarchical clustering of feature points (Crab-HFP) is proposed to achieve fast global subpixel registration between pairs of different band images. Although many demosaicking algorithms have been proposed, they are not suitable for CAMSRIS, because most of these methods are specially designed for multispectral filter array patterns. Because the spectral information captured by the apertures is different from each other, the conventional multi-image super-resolution reconstruction algorithms [27], [28], [29], [30] are not applicable for the proposed CAMSRIS. To the authors' best knowledge, image super-resolution algorithms for camera array based multi-spectral imaging systems have not been published. In this study, a novel super-resolution image reconstruction algorithm for the proposed CAMSRIS is developed in which the spectral clustering based super-resolution reconstruction algorithm (SCSR) can improve the spatial resolution of the acquired images and preserve the exact spectral information without introducing false information. Simulations and field tests are conducted to generate multispectral super-resolution imaging for verification.

In Section II, the principles and observation model of the proposed CAMSRIS are introduced. Section III presents the prototype CAMSRIS built for demonstration. The specific implementation process of Crab-HFP is elaborated in Section IV. In Section V, the proposed SCSR for the CAMSRIS prototype is introduced. In Section VI, the proposed Crab-HFP and SCSR algorithms are verified based on simulations, indoor experimentation, and field tests; experimental results are then compared with the multispectral images reconstructed by the state-of-the-art demosaicking algorithms based on multispectral filter array (MSFA) (e.g., BTES, GAP-TV, and DeSCI) both qualitatively and quantitatively.

The main contributions of the paper are as follows: (1) A novel multispectral camera array-based, spatial super-resolution imaging system is proposed. (2) A complete processing chain is provided ranging from the system design, image rectification, super-resolution reconstruction, and the final qualitative and quantitative indices. (3) The proposed Crab-HFP is used to solve the geometrical rectification problem between apertures in the form of multiple local-geometric transformations. (4) The spatial resolution of images captured by the multispectral camera array is improved based on the proposed SCSR algorithm according to the characteristics of the nonredundant spatial motion information among apertures. Furthermore, the SCSR algorithm achieves faster and higher quality reconstruction outcomes than existing methods in the spatial and spectral domains.

II. Principle and Model for Super-Resolution Imaging in a Multispectral Camera Array

A. Super Resolution Imaging in a Camera Array

The proposed CAMSRIS consists of a camera array and an array of bandpass filters and can be modeled as a planar compound-eye imaging system, as shown in III. The sampling deviation Δd between adjacent apertures can be calculated using

$$\Delta d = \frac{S \cdot f}{P \cdot R} \tag{1}$$

where f denotes the focal length, S denotes the interval between adjacent apertures, P is the pixel size of the detector, and R is the object distance. Equation (1) shows that the sampling deviation between adjacent apertures varies with the object distance. Multi-aperture super-resolution reconstruction requires that the sampling deviation is at the subpixel level to achieve nonredundancy between apertures. During camera array assembly, errors of optical axis parallelism are inevitable. This random error guarantees that the sampling deviation is a random subpixel deviation unaffected by the object distance and avoids excessive redundancy between apertures [31]. This demonstrates that the multi-aperture imaging system has practical super-resolution imaging capability when the nonredundant spatial information provided by the captured image sequence is used.

The observation model of the camera array can be expressed as,

$$\boldsymbol{I}_L^n = \boldsymbol{D}\boldsymbol{W}_n\boldsymbol{I}_H + \boldsymbol{e}_n \quad (n = 1 \ldots N) \tag{2}$$

where, $\boldsymbol{I}_L^n \in \mathrm{R}^{m\times 1}$ is a lexicographically ordered vector of the acquired image of camera n, $\boldsymbol{I}_H \in \mathrm{R}^{mN\times 1}$ is the high-resolution scene to be estimated, $\boldsymbol{W}_n$ is the warping matrix defined by the geometrical transformation relating the image recorded by camera n with the reference image, $\boldsymbol{e}_n$ is noise, $\boldsymbol{e}_n \in \mathrm{R}^{m\times 1}$ is the number of each camera, N is the number of apertures, and $\boldsymbol{D}$ is the bi-dimensional downsampling operator for the imaging focal plane arrays which captures blur due to pixel integration. Typically, Gaussian or bicubic downsampling operators are used. In this study, the implemented image super-resolution method utilized the aliased high-frequency information of the input low-resolution images. There is no usable high frequency information beyond the optical cutoff frequency due to spectral decay imposed by optical blur [32], [33]. So the observation model of the camera array based super resolution imaging system in this paper does not contain optical blur. The sensing matrix of the CAPSRIS is defined as follows:

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{D}\boldsymbol{W}_1 \\ \boldsymbol{D}\boldsymbol{W}_2 \\ \vdots \\ \boldsymbol{D}\boldsymbol{W}_n \end{bmatrix} \tag{3}$$

(2) can also be written as:

$$\boldsymbol{I}_L = \boldsymbol{M}\boldsymbol{I}_H + \boldsymbol{e} \tag{4}$$

where $\boldsymbol{I}_L \in \mathrm{R}^{mN\times 1}$ and $\boldsymbol{e} \in \mathrm{R}^{mN\times 1}$ are the concatenation of all $\boldsymbol{I}_L^n$ and $\boldsymbol{e}_n$ respectively. $\boldsymbol{M} \in \mathrm{R}^{mN\times mN}$ is a sparse matrix.

The super-resolution reconstruction process of the camera array is to reconstruct $\sqrt{N} \times \sqrt{N}$ images into a higher-resolution image [31]. The cutoff angular frequency of a single aperture in a camera array is the Nyquist angular frequency [34]. The frequency response greater than the Nyquist angular frequency is the response of the aliasing signal, which appears as an artifact on the image. The essence of super-resolution reconstruction is to disentangle the high-frequency information aliased in low resolution images; that is, to restore the frequency components greater than the Nyquist angular frequency [31], [35], [36], [37], [38]. The spatial resolution of the reconstructed image is $\sqrt{N}$ times that of a low-resolution image if all images captured by the camera array are completely non-redundant. The non-redundancy among apertures depends on the relationship between diffraction limit frequency, Nyquist frequency and the number of apertures [34].

The diffraction limit angular frequency is defined as:

$$f_{diffraction} = \frac{A}{1.22\lambda} \tag{5}$$

and the Nyquist angular frequency is defined as:

$$f_{Nyquist} = \frac{f}{2P} \tag{6}$$

Without considering factors such as registrational accuracy and optical aberration, the theoretical super-resolution magnification can be expressed as:

$$r_{SR} = \min\left(\frac{f_{diffraction}}{f_{Nyquist}}, \sqrt{N}\right) = \min\left(\frac{2AP}{1.22\lambda f}, \sqrt{N}\right) \tag{7}$$

where A is the aperture diameter of the lens, and λ is the wavelength.

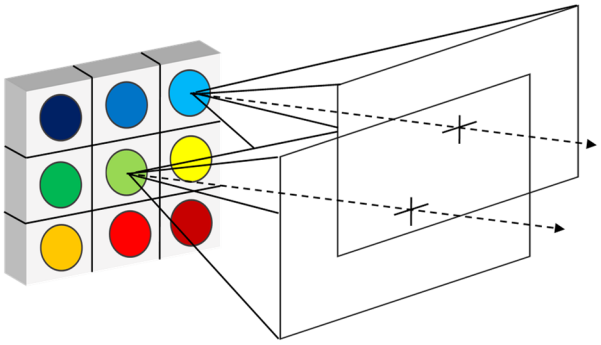


Fig. 1. Schematic of the proposed multispectral multi-aperture imaging system. The field-of-views of the images acquired by different apertures are not the same owing to the offset of each camera position, machining accuracy of the mechanical parts, and the assembly error of the optical system.

B. Observation Model for the Multispectral Camera Array

Each aperture corresponds to a unique bandpass filter in III; the observation model can be expressed as,

$$\boldsymbol{I}_L^n = \boldsymbol{D}\boldsymbol{W}_n\boldsymbol{S}_n\boldsymbol{I}_H + \boldsymbol{e}_n \quad (n = 1 \ldots N) \tag{8}$$

where $\boldsymbol{S}_n$ is a spectral operator that is related to the filter property such as the filter bandwidth and the spectral signatures of the imaged scene, so it cannot be determined *a priori* [19]. The high-resolution kth band scene to be estimated, is defined as $\boldsymbol{I}_H^k = \boldsymbol{S}_k\boldsymbol{I}_H$, $\boldsymbol{I}_H^k \in \mathrm{R}^{mN\times 1}$. The observation model of the CAMSRIS can be written as follows:

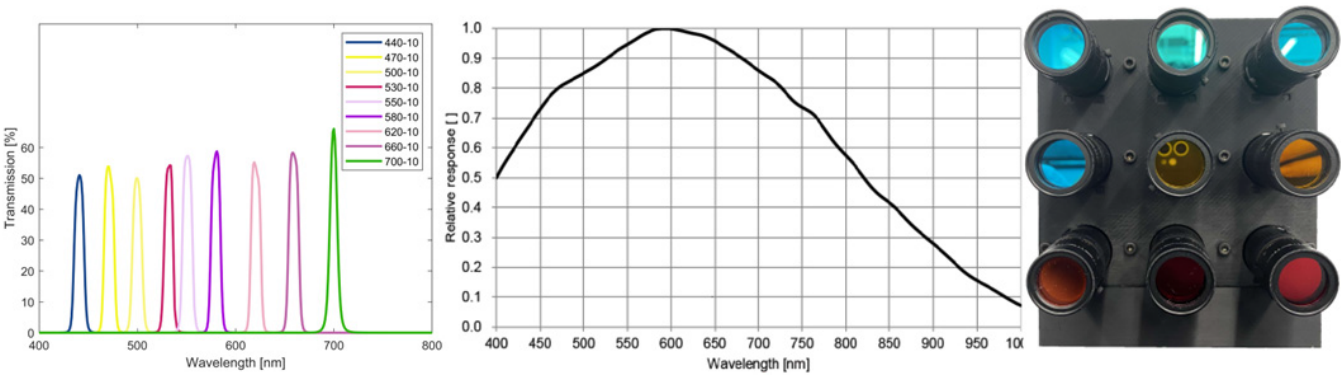
$$\boldsymbol{I}_L^{n,k} = \boldsymbol{D}\boldsymbol{W}_n\boldsymbol{I}_H^k + \boldsymbol{e}_n \quad (n = k, k = 1 \ldots N) \tag{9}$$

where k is the number of each band, and $\left\{\boldsymbol{I}_L^{n,k} \in \mathrm{R}^{m\times 1}\right\}_{n=1}^{N}$ is the low-resolution multi-aperture images of the k-th band. The image only, i.e., the case wherein $n = k$, is captured by the CAMSRIS within $\left\{\boldsymbol{I}_L^{n,k}\right\}_{n=1}^{N}$. Therefore, to create images of other apertures in the k-th band, this paper proposes SCSR. The images generated by the proposed SCSR algorithm are utilized to reconstruct the high-resolution image of the kth band. Thus, (9) can also be modified as follows:

$$\boldsymbol{I}_L^{n,k} = \boldsymbol{D}\boldsymbol{W}_n\boldsymbol{I}_H^k + \boldsymbol{e}_n \quad (n = 1 \ldots N, k = 1 \ldots N) \tag{10}$$

$$\boldsymbol{I}_L^k = \boldsymbol{M}\boldsymbol{I}_H^k + \boldsymbol{e} \quad (k = 1 \ldots N) \tag{11}$$

where $\boldsymbol{I}_L^k \in \mathrm{R}^{mN\times 1}$ are the concatenation of all $\boldsymbol{I}_L^{n,k}$ of the k-th band.



(a) (b) (c)

Fig. 2. Super-resolution reconstructed results of WNNM and LR-GSC using simulated data of CAMSRIS. The size of the reconstructed image is 1500×1098. (a) Ground truth color image. (b) and (c) are the reconstructed results of color target objects (model car, flower) obtained by WNNM and LR-GSC, respectively, and (d) and (e) are the reconstructed results of colorless target objects (resolution test chart, alphabet) in the red box.

III. Construction of the Multispectral Camera Array System

As long as $f_{diffraction} > f_{Nyquist}$, the images acquired by this optical system have the potential for super-resolution processing. To ensure the requirement of nonredundancy among apertures for image super-resolution processing, we set $f_{diffraction} \geq \sqrt{N} \cdot f_{Nyquist}$. This is done to consider the influence of the optical imaging system manufacturing error, aberration, and image registration error, thus letting $f_{diffraction} / f_{Nyquist}$ be larger than $\sqrt{N}$. The CAMSRIS developed in this study consists of nine imaging apertures. According to (7), the value of $f_{diffraction} / f_{Nyquist}$ is affected by the focal length, F-number of the lens, and the pixel size of the camera. Hence, the Basler C125-2522-5M-P lens (with a fixed focal length of 25 mm, F2.2) and Hikvision MV-CA004-10UM camera (Sony IMX287, pixel size 6.9 μm $\times$ 6.9 μm) are used to build the CAMSRIS prototype, as shown in Fig. 2 (c). According to (7), at the central wavelength of the imaging sensor of the selected camera at 550 nm, we have $f_{diffraction} / f_{Nyquist} \approx 9.348 \gg 3$, hence, $r_{SR} = \sqrt{9} = 3$. At the central wavelength of the selected band, the corresponding $f_{diffraction} / f_{Nyquist}$ value is also greater than three. Therefore, the optical imaging system of the CAMSRIS prototype achieves a super-resolution factor of three. All the cameras are synchronized by software triggering to ensure that they simultaneously capture the target images. The spectral response curve of sensor is shown in Fig. 2 (b). The bandpass filters implemented in the prototype are Thorlabs bandpass filters (FB440-10, FB470-10, FB500-10, FB530-10, FB550-10, FB580-10, FB620-10, FB660-10, FB700-10), and their corresponding wavelengths are all within the spectral response region of the sensor, as shown in Fig. 2 (a). The intervals between the central wavelength of the bandpass filters are selected to be approximately equally distributed.

IV. Coarse-to-Fine Registration Algorithm Based on the Hierarchical Clustering of Feature Points

Each aperture in the camera array proposed in this study is equipped with an independent detector. Owing to the difference in the imaging perspective between the apertures, there is a geometrical transformation between the images captured by the apertures denoted by $\boldsymbol{W}_n$ in (7). The warping matrix $\boldsymbol{W}_n$ must be solved before super-resolution reconstruction. It is impossible to achieve a global registration only based on the 3×3 homograph matrix in a complex three-dimensional scene. Given these issues, the traditional method is used to divide the image into several sub-blocks, and to perform registration between the sub-blocks independently. The traditional algorithm has two fundamental problems. The first problem is that the size of the sub-block is difficult to determine [39]. The second problem is that the regional registration algorithm used by the sub-block is associated with an increased multispectral image mismatch rate. In addition, the classic stereo-matching algorithm SGM [40] and its derivatives can also achieve global registration, but they are only limited to one-dimensional translational transformation between images after camera calibration; furthermore, the offset of all pixels needs to be calculated, thus resulting in many mismatched pixels. The stereo calibration approach [41] is capable of achieving row alignment and also transforms the reference aperture image. Pinggera et al. [42] verified that the matching accuracy based on gradient features is higher than those of other registration methods such as those of the normalized cross-correlation and mutual information in multispectral images. Therefore, the coarse-to-fine registration algorithm is proposed; this is based on the hierarchical clustering of the feature points of interest and the use of the gradient feature descriptor. The specific flowchart is shown in Fig. 3. The geometrical rectification of most scenes is accomplished based on coarse registration. Only the offset of the feature points of interest is calculated in the fine registration stage, and the offset of matching pixels can be any two-dimensional translation. The registration algorithm of hierarchical clustering improves the matching accuracy of feature point pairs.

There are not only differences in illumination between images acquired through different apertures in the multispectral camera array, but also nonlinear luminance differences due to the different spectral signatures of target objects with different colors and materials [42], [43], [44], [45]. The nonlinear luminance difference affects the pixel gradient amplitude and gradient direction. Therefore, the mentioned descriptor

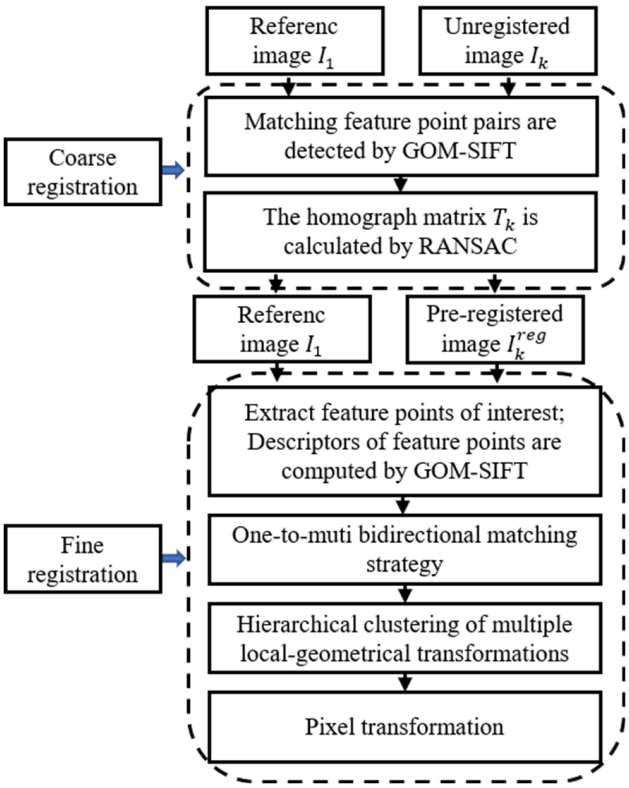


Fig. 3. Flowchart of the proposed coarse-to-fine registration algorithm based on the hierarchical clustering of feature points (Crab-HFP).

with lower computational complexity in [46] is adopted as the descriptor of feature points in the multispectral camera array. The registration algorithm that uses the SIFT feature points in conjunction with the descriptor described above is known as the GOM-SIFT.

A. Uniformly Distributed Corner and Subpixel Edge Feature Points

In coarse registration, GOM-SIFT is used to eliminate the obvious scale, rotational, and translational differences between apertures and complete rough rectification. Based on coarse registration, fine registration is adopted to achieve global and accurate rectification. The feature points of interest are detected first during the process of fine registration. The Harris corner detector is selected because its performance is more stable than other feature-point detectors [47]. The traditional corner extraction method is performed by setting a threshold for the entire image. However, the extracted corners are unevenly distributed. To achieve global image registration, it is necessary to extract image corners that are evenly distributed throughout the image. Moreover, because the thresholds are different for different band images and different scenes, we adopt a divided block and threshold-free method; specifically, we divide the image into $M \times M$ sub-regions, and extracted the top k points with the highest response values from each sub-region [48].

The location at which the local-geometrical transformation changes is at the object's edge. Therefore, edge points also serve as feature points of interest in fine registration. The Canny algorithm requires two thresholds that can be acquired directly from the image information without being set. In this way, edge points can be detected [49]. If the edge detection is performed directly on the entire image, the detected edge feature points will be unevenly distributed. Divided block detection, like the Harris corner detection, is adopted to obtain uniformly distributed edge-feature points. Because the edge-feature points are dense-feature points, and the edge-feature points located on the same edge (usually in local regions) have similar geometrical transformational properties, local nonmaximum suppression is performed to reduce the calculation amount. The super-resolution reconstruction requires the registration algorithm to achieve subpixel accuracy; therefore, after the extraction of the edge-feature points of the entire pixel, the Zernike orthogonal moments are calculated for the edge-feature points to obtain the sub-pixel location [50]

B. A One-to-Many Bidirectional Matching Strategy

After the detection of the feature points of interest, the descriptor of the extracted corner and edge feature points is constructed based on the descriptor of GOM-SIFT. Although GOM-SIFT has been improved by modifying the gradient direction, there are inconsistencies in texture information due to the spectral signatures differences in different spectral images. This results in lower descriptor distinguishability for some feature points. If the matching strategy in SIFT is adopted, the correct matching point pairs will be erroneously removed. Therefore, a one-to-many bidirectional matching strategy is adopted in which a feature point selects the k-matching feature points with the smallest distance metric, and the matching feature point pairs are screened out by the bidirectional matching method [51]. Assuming that the descriptors of feature points in the reference image are $P_q^i, i = 1 \ldots N_q$, descriptors of feature points in the preregistered image are $P_t^j, j = 1 \ldots N_t$, set of k feature points in the preregistered image closest to the reference image feature point is C_1 [as shown in (12)], and the set of k feature points in the reference image closest to the feature point of the preregistered image is C_2 [as shown in (13)], then the final matching result is $C_1 \cap C_2$. When multiple feature points match the same feature point after screening, the matching feature point pair with the smallest distance metric is selected. The mink of (12) and (13) returns the smallest k elements from the set.

$$\boldsymbol{P}_t^j \in \boldsymbol{C}_1 = \text{mink}\left(\left\|\boldsymbol{P}_q^i - \boldsymbol{P}_t^m\right\|\right) \quad (m = 1 \ldots N_t) \tag{12}$$

$$\boldsymbol{P}_q^i \in \boldsymbol{C}_2 = \text{mink}\left(\left\|\boldsymbol{P}_t^j - \boldsymbol{P}_q^n\right\|\right) \quad (n = 1 \ldots N_q) \tag{13}$$

C. Hierarchical Clustering of Multiple Local-Geometrical Transformations

The homograph matrix automatically selects the matrix that satisfies most of the image information. However, the geometrical rectification of the homograph matrix is limited to a plane [52]. The term "plane" is not a plane object in the actual scene, but an approximate plane composed of multiple targets. For the complex three-dimensional scene,

a single homograph matrix cannot satisfy the global rectification requirements. Therefore, this study adopts the multiple transformation matrices mentioned in [53] to achieve global rectification. The pixel offset of the target object between the apertures varies as a function of the imaging distances owing to the foreshortening effects of the lens. Translational transformation is the main transformation in fine registration. However, considering that there are still a few other types of transformations, the local transformation in fine registration adopts the similarity transformation. In Section IV-B, the matching relationship between the feature points of the preregistered and reference images is established. Based on the constraints of the feature space and the local-geometrical transformation, the initial matching feature point pairs are divided into many clusters; each cluster corresponded to a unique similarity transformation. The hierarchical K-means method is used to cluster matching feature point pairs [54] iteratively.

A feature space, which contains the position of the reference image's feature points and the translation of the associated matching feature point pair, denoted as $\left[x^q, y^q, \delta x, \delta y\right]$, is established. As stated, translational transformation is the main transformation in fine registration, but clustering based only on the translation of matching feature point pairs may lead to the intersection of clusters. Because adjacent pixels have similar transformational characteristics, the positional parameters are included to assure increased clustering accuracy and avoid clustering intersection. To eliminate mismatched feature point pairs, a four-dimensional histogram is established. The four dimensions are $\left[x^q, y^q, \delta x, \delta y\right]$, and the unit width of the histogram is $\left[col/4, row/4, col/4, row/4\right]$, where col and row are the columns and the rows of the image, respectively. The feature points are assigned to the corresponding histogram bins. The histogram is counted, and the feature point pairs under the histogram bins with three or fewer votes are deleted [55]. Matching feature-point pairs that have an offset smaller than the threshold are removed. They are not needed for subsequent hierarchical clustering, and they are removed by similarity space clustering. This is performed to reduce the calculation load and to not destroy the registered areas in coarse registration.

According to the established feature space, the K-means algorithm is used to divide the matching point pairs within the node into k clusters. In different depth layers of hierarchical clustering, the contributions of parameters in the feature space are different. Therefore, the weight in [53] is adopted.

RANSAC [56] is used to estimate the similarity transformation of each node. A nonparametric outlier detection technique is used to delete feature point pairs with a large backprojection error. These feature point pairs are mismatched or incorrectly clustered. The statistical distribution of backprojection errors is used to set the threshold to $\tau_\gamma = q_3 + \gamma (q_3 - q_1)$, where q_i denotes the ith quartile of back projection errors of feature point pairs. The feature point pairs larger than τ_γ are then deleted. By adjusting γ, deleting incorrectly clustered but correctly matched feature-point pairs can be avoided. As the number of layers increases, the number of clusters also increases, and the probability that a high-backprojection error causes incorrect clustering decreases; thus, γ can also be set to smaller values.

$$\gamma = \max\left(\gamma_0 - \varepsilon\left(tree\_depth - 3\right), 0\right) \tag{14}$$

where γ_0 is the initial value, $tree\_depth$ is the layer's serial number of the tree structure where the node is found, and ε is the step length. The serial number of the layer at the beginning of clustering is the third layer.

The ratio of inliers and the number of outliers is used to determine whether a node is a leaf node. When the ratio of inliers is above a threshold, i.e., $\rho_i = M_i / (M_i + O_i) > \tau_\rho$, and when the number of outliers is below a threshold $O_i < \tau_o$, the node is deemed a leaf node. Herein, M_i is the number of inliers, O_i is the number of outliers, and τ_ρ and τ_o are the threshold constants.

In the correction phase, linear discriminant analysis is used to reassign outliers to other clusters. Clustering is repeated, the similarity transformation is estimated, and corrections are made until all nodes are leaf nodes.

Counting the inliers in all the leaf nodes, the offsets of other pixels are obtained by linearly interpolating the inlier offsets. Multiple mesh-to-grid resampling steps degrade image quality. Therefore, although a double-registration algorithm is adopted, only one resampling procedure is conducted [17].

V. Spectral Clustering Super-Resolution Reconstruction Algorithm

The target object shows different spectral signatures in different bands, affected by color, material, and other factors. The difference between spectral signatures is manifested in two ways: first, although the texture structural information is consistent, the gray values are different. This emerges as a linear luminance difference [21]. Second, the texture structure information is inconsistent or even vanished, and this manifests itself as a nonlinear luminance difference [24]. As described in Section II-B, every aperture needs to provide a multispectral data cube to achieve multispectral camera array super-resolution. However, in actual imaging, each aperture captures only a single-band image. The spectral information of eight band images of other apertures is exploited to simulate the corresponding band images subject to the imaging perspective of the aperture in question. The proposed SCSR algorithm maintains the spectral signatures of a simulated band and the nonredundant spatial information of the original image during the process of simulation. The specific flow chart is shown in Fig. 4. "Spectral clustering" is a new definition proposed in this study for simplicity compared with its meaning in the traditional sense [57]; it means that pixels with the same or similar spectral signatures are grouped in this study.

A. Spectral Clustering

For the luminance difference only in the first listed case above, we can directly classify the pixels with the same or similar spectral signatures into a cluster by K-means [58], [59], [60]. However, there is inconsistency in texture structure information between different band images during imaging [24]. K-means must give a certain number of

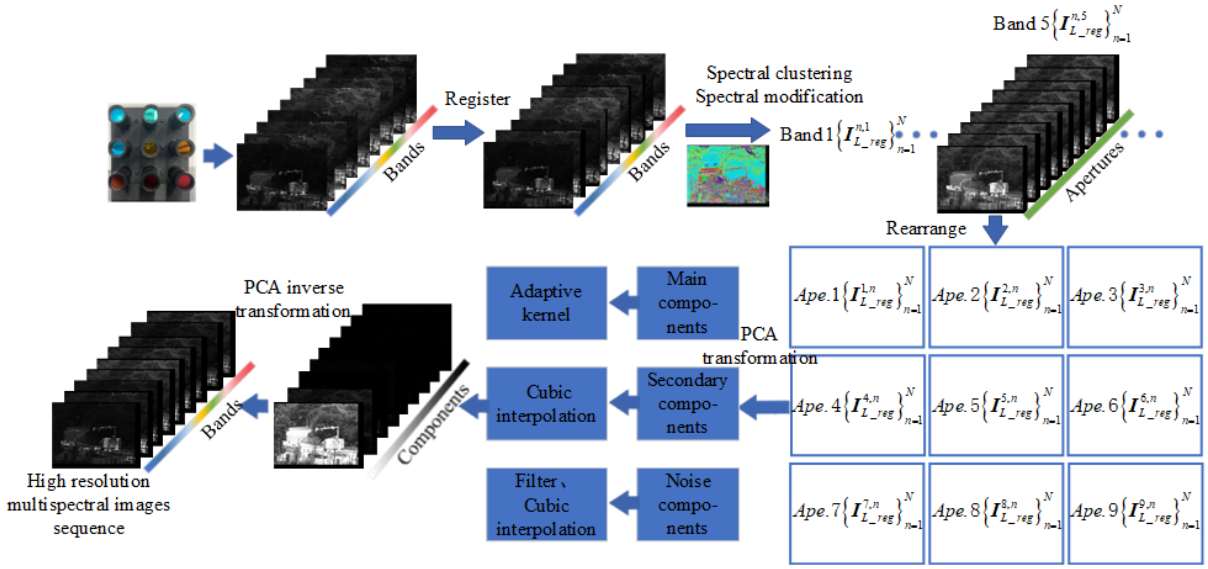


Fig. 4. Scheme of the proposed spectral clustering based super-resolution reconstruction algorithm (SCSR) for the proposed camera array based multispectral super resolution imaging system (CAMSRIS). Low-resolution images are acquired using the CAMSRIS prototype. The fields-of-view of the nine monochromatic images collected in different spectral bands are not consistent. The proposed Crab-HFP algorithm is used to register and align images in the different spectral bands. Spectral clustering and spectral modification are performed on the aligned multispectral data cube; therefore, multispectral data cube for each aperture can be constructed after rearrangement. PCA transform is used to obtain the main, secondary, and noise components of each aperture. The adaptive kernel maximum likelihood estimation algorithm is used for the super-resolution reconstruction of the main components. Finally, high-resolution multispectral images are generated using PCA inverse transformation.

clusters, while the inconsistent texture structure information is usually only a tiny local area, and it is difficult to count its number. Therefore, the ISODATA algorithm [61] is added to K-means.

With the fifth aperture as the reference image, other aperture images are geometrically rectified. The multispectral data cube is constructed from the rectified multi-aperture images, and each sample represents the gray value of nine bands at the same spatial position. The K-means ++ algorithm [62] selects k clustering centroids as the initial clustering centroids of K-means. Based on the K-means clustering results, the ISODATA algorithm is adopted. To identify localized small region clusters with large variance, a novel split metric method is proposed to replace the variance metric in the ISODATA algorithm, as follows:

(1) Delete the clusters whose sample number is less than the threshold and reassign the samples of these clusters to the remaining clusters.

(2) The samples of cluster i are denoted as $\boldsymbol{X}_i$. Screen out the sample sets $\left\{\boldsymbol{X}_i^1, \cdots, \boldsymbol{X}_i^k, \cdots, \boldsymbol{X}_i^N\right\}$ in N dimensions where the samples of each set are the distances to the centroid of cluster i in the k-th dimension that are greater than the standard deviation threshold τ_σ. Count the number of samples in $\left\{\boldsymbol{X}_i^1, \cdots, \boldsymbol{X}_i^k, \cdots, \boldsymbol{X}_i^N\right\}$ and select $\boldsymbol{X}_i^k$ according to the maximum number of samples $n_{\max}$. If $n_{\max} \geq N_{\min}$, then split cluster i into two clusters. The centroids of the two clusters are respectively $\boldsymbol{m}_i^{(+)}(k) = \boldsymbol{X}_i^k(:, k)$ and $\boldsymbol{m}_i^{(-)}(k) = \mathbf{X}_i \backslash \boldsymbol{X}_i^k(:, k)$. This operation traverses all clusters. If splitting is performed, the samples are reassigned according to the current centroids.

(3) Calculate the distance between the centroids of two clusters and merge the clusters where the distance is less than the threshold. The combined centroids are calculated according to [61]. If merging is performed, the samples are reassigned according to the current cluster centroids.

(4) Repeat steps (2) to (3) until the centroids of clusters remain unchanged or reach the maximum number of iterations.

For simplicity of description, we refer to this algorithm as K-means+ISODATA. When there is a large number of clusters, calculating the distance between each sample and the centroids of all clusters and then picking out the cluster with the shortest distance, requires numerous calculations. The hierarchical tree structure in Section IV-C is adopted to reduce the calculation amount so that each sample point only needs to calculate the distance to the centroids of clusters within its affiliated node instead of calculating the distance to all the cluster centroids. This process is described as follows: First, the initial clusters are obtained by K-means. Second, whether or not the node needs to be split is decided by two conditions pertaining to the splitting criteria of (2) in the K-means+ISODATA algorithm; furthermore, it is also decided whether the number of samples within a node is more than $2N_{\min}$. When the two conditions are true, the K-means+ISODATA algorithm is applied, where K = 2. Otherwise, the node is a leaf node. Third, repeat step 2 until all nodes are leaf nodes or until they reach the maximum number of iterations.

B. Spectral Modification

Histogram matching is used to process the image with the specified histogram. The histogram shape of the processed image is similar to the specified histogram shape but retains the original image's spatial information. Therefore, histogram matching is used to rectify spectral differences. Spectral

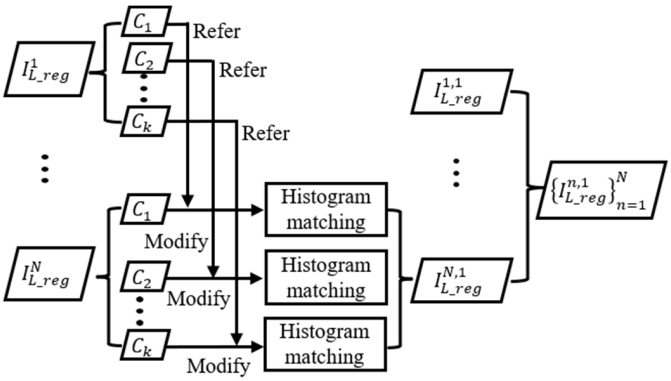


Fig. 5. Schematic of the spectral modification process for nine aligned, single-band, monochromatic spectral images. Using the image of the first band as the reference image and traversing all the clusters to correct image spectral differences in the other bands enables the acquisition of the first-band images by other N - 1 cameras with different aperture positions. Note that the images of the first band collected by the remaining N - 1 apertures are obtained by calculation and not by actual shooting.

clustering is performed on the multispectral data cube, and c clusters are obtained. Select the kth band image as the reference band image (k = 1 . . . N), and repeat these steps:

(1) The histogram matching algorithm is adopted to modify the spectral signatures of images with other apertures captured. The histogram shape of the reference band image is used as the specified histogram in cluster i so that the gray value of the modified image tends to be consistent or close to that of the reference band image.

(2) Repeat step 1 to traverse all the clusters and reassemble all the modified clusters into the kth band multi-aperture images sequence $\left\{\boldsymbol{I}_{L\_reg}^{n,k}\right\}_{n=1}^{N}$. $\left\{\boldsymbol{I}_{L}^{n,k}\right\}_{n=1}^{N}$ can be obtained by the inverse transformation of W_n.

The process of spectral modification and reconstruction of other band images in the first band image (used as the reference band image) is described in Fig. 5.

C. Adaptive Kernel Maximum Likelihood Estimation

After $\left\{\boldsymbol{I}_{L}^{n,k}\right\}_{n=1}^{N}$ is established, the high-resolution image of the kth band can be reconstructed according to (9). If the super-resolution reconstructions of the nine bands are performed independently, the computational load will be heavy. Therefore, the PCA algorithm is adopted to obtain the main, secondary, and noise components of the multispectral image [63]. Therefore, super-resolution reconstruction of the main components can thus be performed.

The objective function of the super-resolution reconstruction of the main components is described as follows:

$$\boldsymbol{C}_H^k = \arg\min_{\boldsymbol{C}_H^k} \left\| \boldsymbol{M}\boldsymbol{C}_H^k - \boldsymbol{C}_L^k \right\|_2^2 \quad (k = 1 \ldots P) \tag{15}$$

where $\boldsymbol{C}_L^k \in \mathrm{R}^{mN\times 1}$ is the concatenation of all $\boldsymbol{C}_L^{n,k} \in \mathrm{R}^{m\times 1}$ of the k-th main component, P is the number of main components, and $\boldsymbol{C}_H^k \in \mathrm{R}^{mN\times 1}$ is the k-th high-resolution main component to be estimated.

The adaptive kernel maximum likelihood estimation algorithm is proposed for the super-resolution reconstruction of the main components. The reasons for selecting the maximum likelihood estimation are that a) thanks to the layout of the camera array, the offsets among all apertures are uniformly distributed in a disk, which makes the super-resolution reconstruction problem of the camera array well-posed [32]. b) the super-resolution reconstruction performed based on the camera array aims to recover the high-frequency information from the image; moreover, there may be a small amount of noise within the acceptable degree of human vision. c) Although the the state-of-the-art denoising prior model has a considerable effect on denoising and can preserve the edge and suppress the noise, the calculation time is too long and the prior terms and data guarantee terms need to be balanced. The denoising prior still has certain smoothing side effects on the edge and in the texture region [27], [28], [29], [30], [31], [64]. d) The maximum likelihood estimation algorithm is based on the nonprior Bayesian model associated with increased calculation speed and is similar to a Lucy–Richardson iterative deconvolution. Considering that the proposed system is a bandpass multispectral imaging system, wherein the signal-to-noise ratio (SNR) of the image is low and the noise is amplified by the iteration process, an adaptive kernel is used to establish the sensing matrix $\boldsymbol{M}$; that is, the kernel is adaptively selected according to the characteristics of the image to reduce the noise of the reconstructed image and to ensure the sharpness of the edge [31].

The objective function (15) is iteratively solved based on the maximum likelihood estimation algorithm according to

$$\boldsymbol{C}_H^{k,i+1} = \mathrm{diag}\left(\boldsymbol{C}_H^{k,i}\right)\boldsymbol{M}^T\left(\mathrm{diag}\left(\boldsymbol{M}\boldsymbol{C}_H^{k,i}\right)\right)^{-1}\boldsymbol{C}_L^k \tag{16}$$

where $\boldsymbol{C}_H^{k,i}$ is the i-th approximation result of the k-th main component of the high-resolution scene to be estimated, and $\mathrm{diag}\left(\boldsymbol{C}_H^k\right)$ is the diagonal matrix with elements of vector $\boldsymbol{C}_H^k$. For initialization, $\boldsymbol{C}_H^{k,0} = \boldsymbol{M}^T\boldsymbol{C}_L^k$.

To utilize the subpixel information of spatial motion more effectively, forward warping is used to construct the sensing matrix. That is, the pixels of images captured by the other apertures are mapped onto the high-resolution mesh of the reference image and then interpolated to the grid by the adaptive spatial support of the sampling kernel. Subsequently, the pixel values are aggregated on the grid, as depicted in Fig. 6. The large sparse matrix constructed based on the above-mentioned steps is the transpose of the sensing matrix $\boldsymbol{M}$.

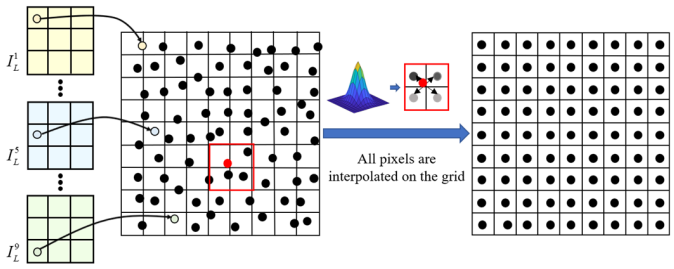


Fig. 6. Schematic of the forward warping process. The pixels of images captured by the other apertures are mapped onto the high-resolution mesh of the reference image, and interpolated to the grid by the adaptive spatial support of the sampling kernel. Then the pixel values are aggregated on the grid.

The type of chosen kernel function has little influence on the reconstruction result [65]; therefore, the Gaussian function with lower computational complexity is selected as the kernel function. However, if the same Gaussian function is utilized for all samples, the denoising effect of the small kernel space support is insignificant, while the large kernel space support will cause blurred edges. Based on the above problems, the anisotropy Gaussian kernel [66] is employed to adjust the sampling kernel shape according to the gradient structure tensor of the sample.

Given that the fuzzy edge of the low-resolution image reduces the detection accuracy of the edge direction, the isotropic Gaussian kernel with a fixed size can be used to reconstruct the higher-resolution image first to detect the edge direction more accurately, and the adaptive kernel of all pixels in the lower resolution image sequence is then calculated from the higher-resolution image.

VI. Results and Analysis

A. Parameter Setting

In the spectral clustering stage, the initial standard deviation threshold τ_σ is set to 50; its value decreases layer-by-layer according to the step size, which is set to 10. To avoid excessively small τ_σ values at deep levels, the minimum τ_σ values are set within the range of 10–25 as the merging threshold $d_{\min}$. $N_{\min}$ depends on the image size. The maximum number of layers in the tree structure is set to 10. The maximum number of iterations for the K-means+ISODATA method is set to 15. The number of iterations of the adaptive kernel maximum likelihood estimation is 5. The number of image blocks of the quality index Q [67] in the spectral distortion index [68] is set to 3×3. The aforementioned parameters are configured for 2^8 bit-depth images.

B. Simulation Results and Analysis

To evaluate CAMSRIS, we apply it to the CAVE [69] and CAMSI [23] datasets to validate its feasibility in the cases of visible and visible and near-infrared spectra. The CAVE dataset has a band in the range of 400–700 nm and contains 31 band images at 10 nm intervals (image size = 512×512). Take a sample every three bands from the CAVE dataset to simulate a nine-aperture multispectral camera array. The selected bands are 430, 460, 490, 520, 550, 580, 610, 640, and 670 nm. The CAMSI dataset contains the nine-band images with central wavelengths at 400, 450, 470, 500, 524, 660, 750, 850, and 950 nm (image size = 1500×1100). The proposed SCSR algorithm is compared with state-of-art multiple image super-resolution reconstruction algorithms, namely WNNM [29], [64] and LR_GSC [30], to highlight the superiority of SCSR. WNNM is a maximum posterior probability algorithm based on low-rank attributes of images, while the LR_GSC is a maximum posterior probability algorithm combining low-rank and sparse image attributes. CAMSRIS is also compared with the current advanced multispectral imaging technology MSFA to analyze its performance. To make the comparison effective, it is required that the same number of bands and total number of pixels are used. The current multispectral demosaicking algorithms are primarily based on filter arrays with 4- or 5-band arrays [21], [22], [23], [24], which are not suitable for 9-band reconstruction. Therefore, the demosaicking algorithms adopt more general methods, such as the BTES [20] and the compressed sensing algorithms, which are GAP-TV [25] and DeSCI [26]. According to [70], the reconstruction results of the BTES and the compressive sensing algorithms based on the binary tree filter pattern are better than the random and regular models. The optimal model for MSFA is adopted to emphasize the benefits of CAMSRIS.

The low-resolution multispectral image sequence is simulated with the ground truth multispectral dataset based on the observation model of the multispectral camera array described in Section II-B. The ground truth multispectral images are degraded into low-resolution multispectral images after geometrical transformation (random offset, rotation, and scale transformation), downsampling, and noise interference. The downsampling factor is equal to three, and the Gaussian weight function is used during the downsampling operation. The noise interference is Gaussian white noise. For MSFA, the ground truth multispectral dataset is sampled based on binary tree filter patterns to obtain a multispectral mosaic image with the same noise as CAMSRIS. The reconstruction results are compared qualitatively and quantitatively. The quantitative methods include PSNR, MultiSSIM [71], and BRISQUE [72]. A smaller BRISQUE indicates a better image quality. Considering the practicability and feasibility of the system, we evaluate the execution time of the reconstruction algorithms of the two imaging systems in MATLAB (version xx, MathWorks, Natick, MA, USA) based on an Intel I5-9500 processor (3.00 GHz) and 16 GB of memory. DeSCI imposes heavy requirements on running memory; thus, the 16 GB of memory is insufficient to meet its requirements. Therefore, the DeSCI is executed on an Intel Xeon(R) Gold 5220R processor (2.20 GHz $\times$ 96) and with 125.5 GB of memory; all other experiments are executed based on the configuration discussed earlier.

A high-resolution image is reconstructed using the multiframe super-resolution reconstruction algorithms. They are applied to the degraded multispectral images of the CAMSI dataset; these images are generated according to the image formation model of the multispectral camera array as described previously. The multiframe super-resolution reconstruction algorithms are not appropriate for the multispectral camera array, as illustrated in Fig. 7. Because of the luminance difference between the images of different bands, reconstructed images not only lose the spectral signatures but also have an obvious grid and spots for color or colorless target objects.

CAMSRIS and MSFA are compared based on the CAMSI and CAVE datasets. The reconstructed results of the CAMSI dataset are shown in Figs. 8 and **9**; as shown, GAP-TV and DeSCI produce spatial and spectral distortions. As shown in Fig. 9, GAP-TV and DeSCI produce severe spectral distortions for the 400 nm band in both the background and flower regions. As shown in Fig. 8, BTES, GAP-TV, and DeSCI produce unexpected artifacts and deformation in the spatial domain. These are manifestations of the disadvantages of MSFA that become more obvious with more bands.

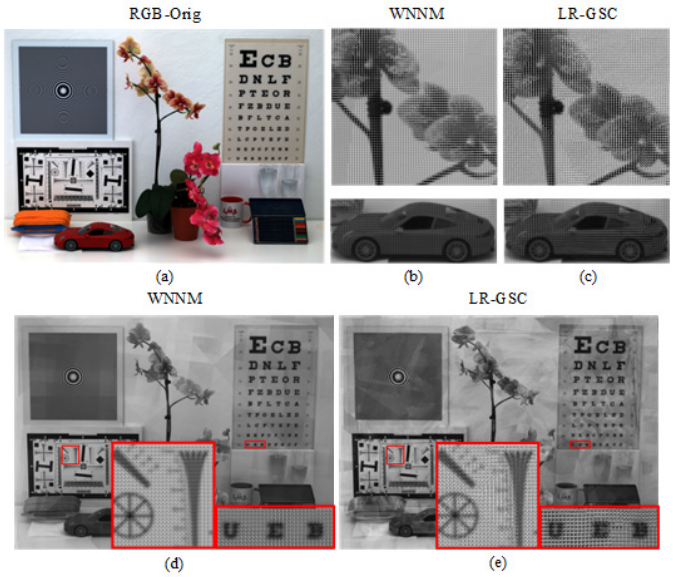


Fig. 7. Super-resolution reconstructed results of WNNM and LR-GSC using simulated data of CAMSRIS. The size of the reconstructed image is 1500 × 1098. (a) Ground truth color image. (b) and (c) are the reconstructed results of color target objects (model car, flower) obtained by WNNM and LR-GSC, respectively, and (d) and (e) are the reconstructed results of colorless target objects (resolution test chart, alphabet) in the red box.

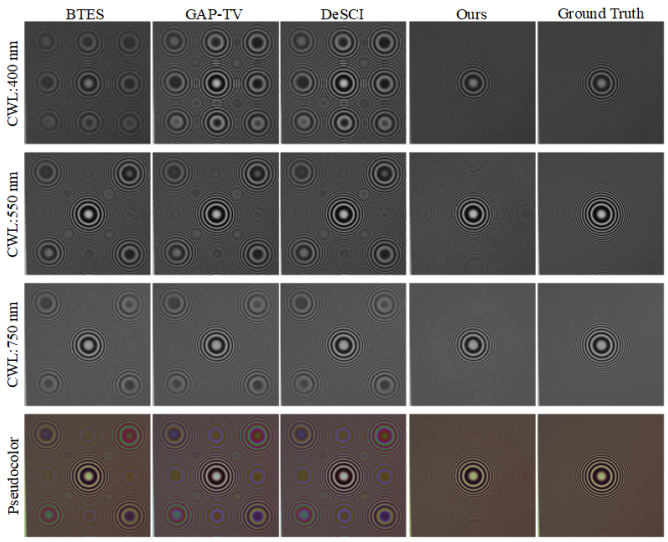


Fig. 8. Comparison of the multispectral images reconstructed by BTES, GAP-TV, DeSCI, and the method proposed in this study. The CAMSI dataset is used for simulations (concentric circles). The images of three bands (400, 550, and 750 nm) in the reconstructed 9-band multispectral images are selected for display. The selected images at these three different bands are used to reconstruct pseudocolor images for comparison. In the simulation of the three methods of BTES, GAP-TV, and DeSCI, the data collected by the 9-band snapshot multispectral imaging sensor is simulated first, and the three methods are then used for demosaicking to restore the 9-band full-resolution multispectral image.

In contrast, as the number of bands increases, the number of apertures also increases and the spatial resolution can thus be improved within the diffraction limit for CAMSRIS. Fig. 9 shows the reconstructed multispectral images of CAMSRIS and MSFA in the chart and the stuffed toy of the CAVE dataset. It is clear from the concentric circle of Fig. 8 and the resolution test chart of Fig. 10 that our proposed algorithm

TABLE I

AVERAGE PSNR, MULTISSIM, BRISQUE, AND CALCULATION TIME CONSUMPTION OF THE MULTISPECTRAL IMAGE RECONSTRUCTION RESULTS TO THE CAMSI AND CAVE DATASETS

	Algorithm	BTES	GAP-TV	DeSCI	Ours
CAMSI	PSNR/dB	26.85	27.59	27.69	**29.62**
	MultiSSIM	0.9080	0.9188	0.9214	**0.9491**
	BRISQUE	43.73	45.37	45.19	**24.64**
	Time/s	194.13	120.35	9885.99	**65.80**
CAVE	PSNR/dB	29.37	30.33	**30.78**	30.54
	MultiSSIM	0.9709	0.9711	0.9712	**0.9785**
	BRISQUE	44.24	47.56	40.95	**34.85**
	Time/s	39.45	18.66	3426.08	**15.80**

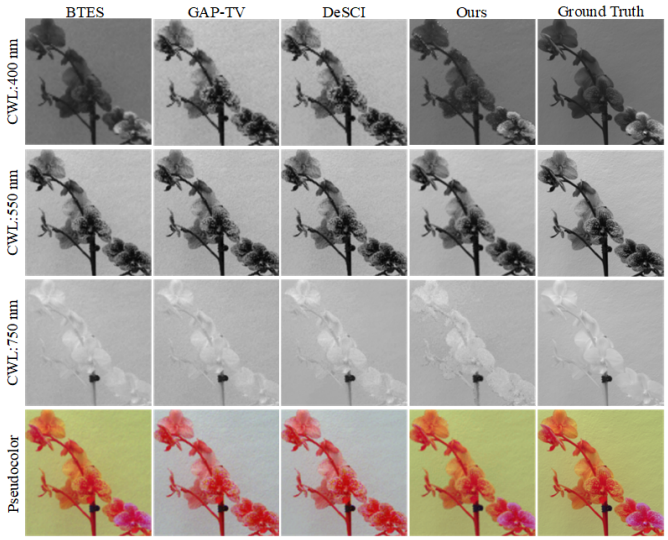


Fig. 9. Comparison of the spectral profiles/signatures reconstructed by BTES, GAP-TV, DeSCI, and the method proposed in this study. The CAMSI dataset (artificial flower target) is used for simulations. The images of the three bands (400, 550, and 750 nm) in the reconstructed 9-band multispectral images are selected for display. The selected images at these three different bands are used to reconstruct a pseudo-color image for comparison. In the simulations of the three methods of BTES, GAP-TV, and DeSCI, the data collected by the 9-band snapshot multispectral imaging sensor are simulated first, and the three methods are then used for demosaicking to restore the 9-band, full-resolution, multispectral image.

can recover small-scale details, while the BTES, GAP-TV, and DeSCI not only smooth out small-scale details but also produce distortion and artifacts in large-scale contours. By observing the spectral curves and spectral angle mapper in Fig. 11, it can be inferred that the proposed algorithm has higher spectral reconstruction accuracy.

The listings in Table I show that the PSNR and MultiSSIM of our proposed algorithm perform better, wherein the PSNR is 2.0 dB higher than the recently published DeSCI algorithm on the CAMSI dataset, while BRISQUE and the execution time are the lowest among all the algorithms. CAMSRIS is superior to MSFA for all demosaicking algorithms and indices in the CAMSI dataset. Considering the execution time, our algorithm focuses on recovering high-frequency information. The adaptive kernel has a good denoising effect, but it is not as good as the low-rank regularization of DeSCI. Therefore, our PSNR is slightly lower than for the DeSCI algorithm, as shown in Table I.

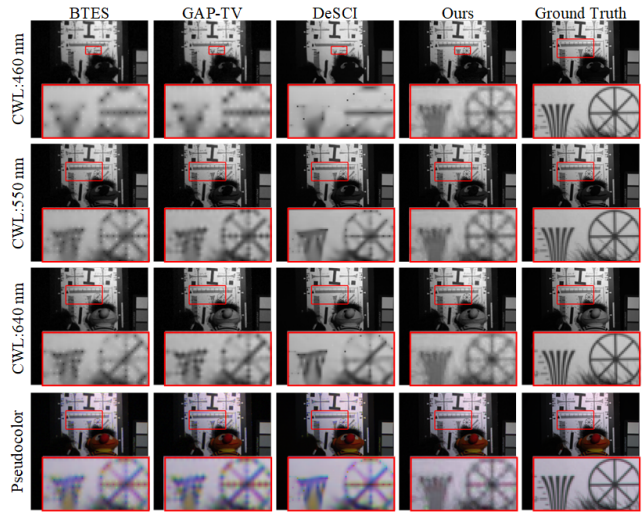


Fig. 10. Comparison of the multispectral images reconstructed by BTES, GAP-TV, DeSCI, and the method proposed in this study. The CAVE dataset (chart and stuffed toy) is used for simulations. The images of the tree bands (400, 550, and 750 nm) in the reconstructed 9-band multispectral images are selected for display. The selected images at these three different bands are used to reconstruct a pseudocolor image for comparison. In the simulations of the three methods of BTES, GAP-TV, and DeSCI, the data collected by the 9-band snapshot multispectral imaging sensor are simulated first, and the three methods are then used for demosaicking to restore the 9-band, full-resolution multispectral image.

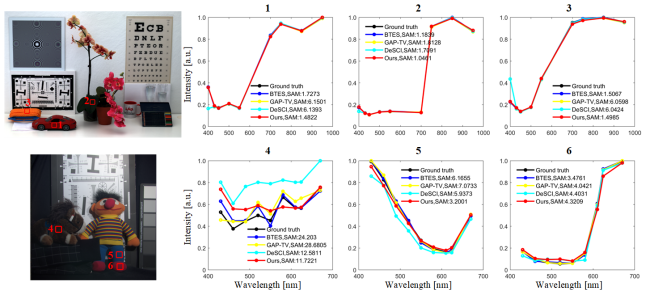


Fig. 11. Reconstructed spectra of indoor scenes from the CAMSI dataset and chart and stuffed toy from the CAVE dataset. (Left) Color images of the scenes. The spectra of the three-color regions are shown on the right, where the ordinate is normalized in all the cases (divided by the maximum value of the samples in the selected region). The spectral angle mapper (SAM) of the reconstructed spectral images calculated using BTES, GAP-TV, DeSCI, and our methods are shown in the legends.

C. Real Data Results and Analysis of Super-Resolution Reconstruction

As the ordinary indoor light sources, such as incandescent lamps and light-emitting diode (LED) lamps, cannot provide reliable and stable lighting in the visible spectrum, six spectral tunable light sources (Thouslite, LEDCube) are used as light sources to illuminate the indoor scenes. For the field tests, the images are acquired on sunny days.

LSS [73], GOM-SIFT [46], and Crab-HFP are compared. LSS and GOM-SIFT are the same except for the descriptor. They both adopt the SIFT algorithm to extract feature points and RANSAC to solve the transform matrix. Owing to the nonlinear luminance differences between band images, it is not suitable to use the difference between the reference and the band image after geometric rectification to assess the matching effect. Although all band images can be collected

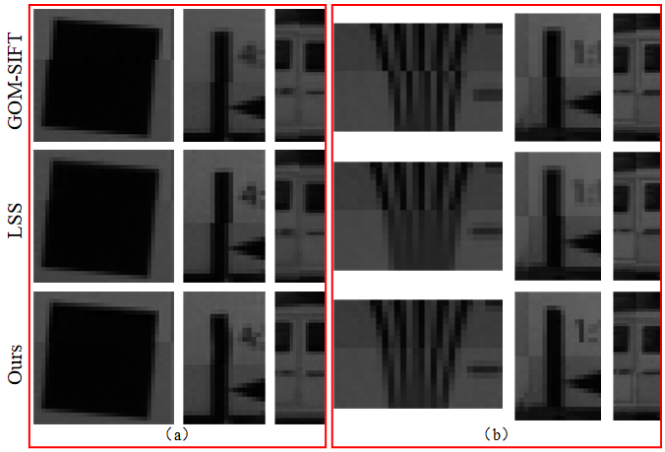


Fig. 12. (a) The local results of checkerboard stitching in 440 nm band images after geometric rectification and the 550 nm reference band image (b) The local results of checkerboard stitching in 700 nm band images after geometric rectification and the 550 nm reference band image.

TABLE II
BRISQUE OF DIFFERENT ALGORITHMS IN DIFFERENT SCENARIOS

Algorithm	WNNM	LR-GSC	Tric	Ours
indoor scene	32.18	35.44	46.08	**27.33**
traffic	42.66	43.53	51.46	**29.02**
building	43.46	43.45	49.02	**26.75**

at the reference aperture position by changing the filter for the static scene; the camera array will shift slightly during the change of the filter. Therefore, a checkerboard stitching method is adopted. The reference band image and the rectified band image are divided into 16 × 16 blocks, and their blocks are spliced alternately. Fig. 12 shows the local results of checkerboard stitching in the 440 nm and 700 nm band images after geometric rectification and the 550 nm reference band image. Fig. 10 shows that the LSS and GM-SIFT outcomes are misaligned, while the proposed registration algorithm can align the image to be registered with the reference image.

The feasibility of CAMSRIS in practical applications is further verified by using actual data. We cannot compare with MSFA (real data cases) because of the lack of ground truth multispectral images. Therefore, this section will focus on the comparison of the multispectral images obtained via triangulation-based cubic interpolation (Tric) [74] and SCSR with those acquired using the camera array. In real data cases lacking ground truth multispectral images, three indices are used to evaluate image quality. The first is BRISQUE. The second is the super-resolution multiplication index (see Section II-A), and the test card adopted sinusoidal Siemens stars. The third is the spectral distortion index, which is evaluated by calculating the difference between the inter-band correlation of multispectral images before and after reconstruction.

Three representative scenes were captured, namely the indoor scene, the static building, and fast-moving traffic. As shown in Figs. 13–16, block artifacts are generated in the multispectral images captured by the apertures, while blurred artifacts exist in the multispectral images after cubic interpolation. The images after reconstruction by WNNM and LR-GSC

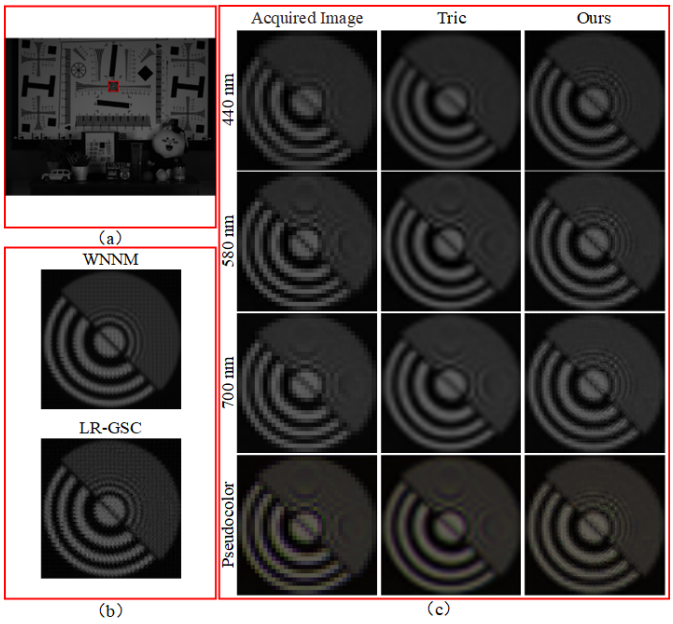


Fig. 13. Comparison of the reconstruction results of the real indoor scene using the constructed CAMSRIS described in Section III. (a) Global image. (b) Reconstruction results of WNNM and LR-GSC related to the information in the red box in (a). (c) Multispectral images captured by the apertures, multispectral images after cubic interpolation, and after SCSR in the red boxes of (a) for 440, 580, and 700 nm bands and pseudocolor images synthesized by these three bands. The size of the reconstructed multispectral images after SCSR is $2160 \times 1620 \times 9$.

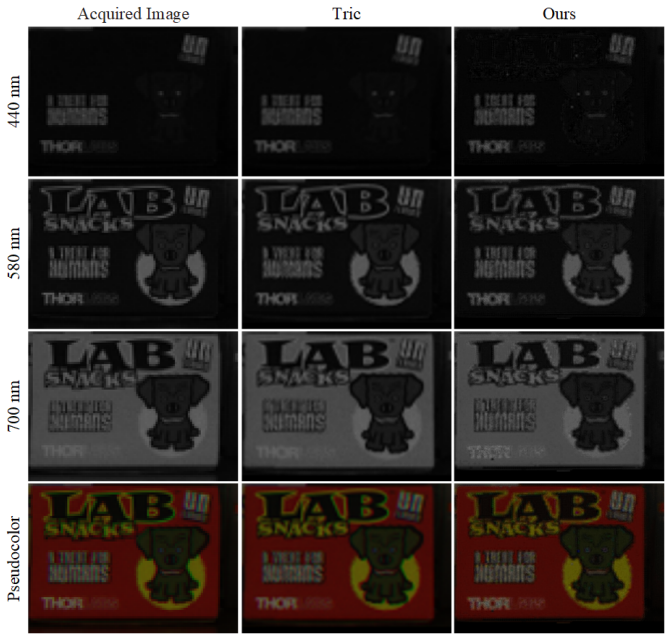


Fig. 14. Spectral profiles/signatures of the images acquired in an indoor environment using the constructed CAMSRIS described in Section III (first column), those processed using Tric (second column), and those reconstructed by the proposed method (third column). The images in the three bands (440, 580, and 750 nm) from the reconstructed nine-band multispectral images are selected for display. The selected images at these three different bands are used to reconstruct a pseudocolor image for comparison.

not only lost spectral signatures but also produced obvious grids and spots. The proposed algorithm could produce sharp edges and eliminate image noise while preserving the exact spectral information without introducing false information.

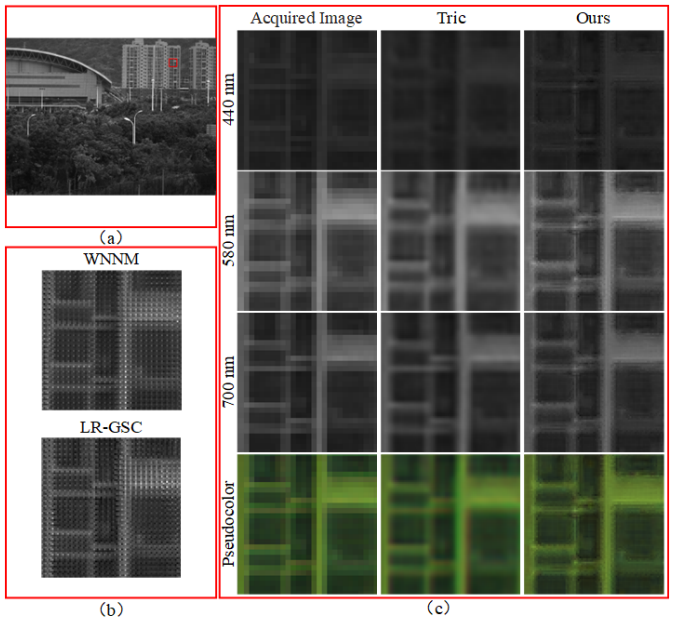


Fig. 15. Comparison of the reconstruction results of an real architectural scene using the constructed CAMSRIS described in Sect. III.

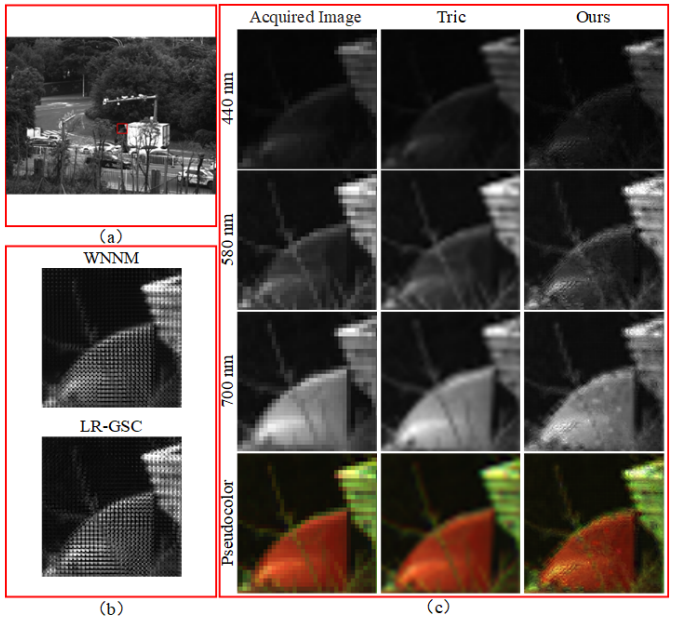


Fig. 16. Comparison of the reconstruction results of real traffic scene using the constructed CAMSRIS described in Sect. III. The frame rate is 30 frames per second.

According to the BRISQUE in Table II, the reconstruction results of the proposed algorithm have the best spatial quality for the three mentioned scenes. Sinusoidal Siemens stars is used to obtain the spatial frequency response curves of the multispectral images captured by the apertures. As shown in Figs. 17 and 18 our algorithm can recover high-frequency information on a smaller scale, eliminate artifacts, and improve limit resolution.

The limiting frequency of a single aperture is the Nyquist frequency equal to 0.5 cycles/pixel. The limit resolution of the reconstructed image is the frequency of the intersection

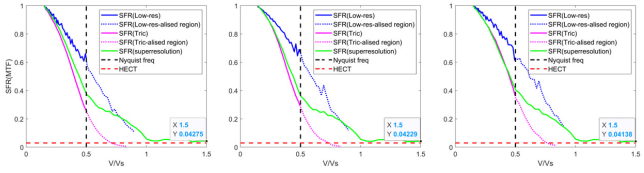


Fig. 17. Spatial frequency response curves calculated from Sinusoidal Siemens Stars images acquired by the constructed CAMSRIS described in Sect. III, those processed using Tric, and those reconstructed by the proposed method in Fig. 18.

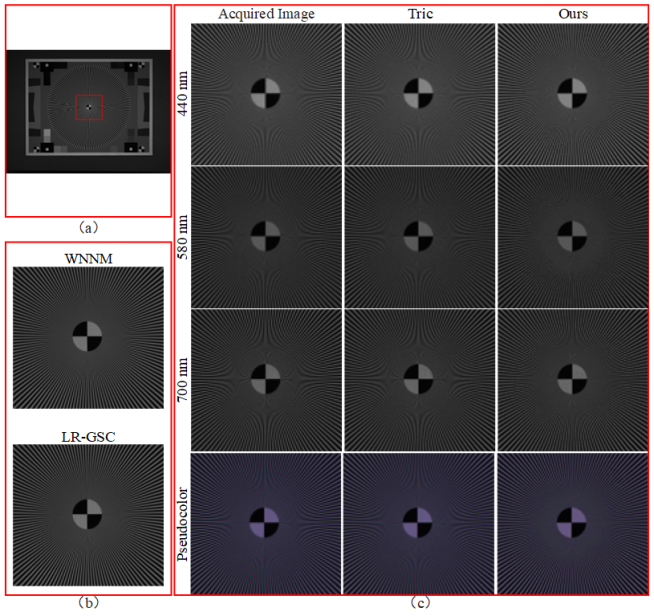


Fig. 18. Comparison of the reconstruction results of real sinusoidal Siemens stars using the constructed CAMSRIS described in Sect. III. The central high-frequency region is that highlighted the red box in (a).

TABLE III
SPECTRAL DISTORTION AND THE TIME CONSUMPTION OF SPECTRAL IMAGE REGISTRATION AND RECONSTRUCTION TO THE IMAGES ACQUIRED BY THE PROPOSED CAMSRIS

Scene	Indoor	Traffic	Building
spectral distortion	0.014	0.0376	0.0548
registration	8.21 s	7.30 s	6.68 s
reconstruction	99.36 s	138.89 s	104.80 s

between its spatial frequency response curve and the human eye contrast threshold (HECT); its value is approximately equal to 0.03. The limit resolution of the reconstructed image after SCSR is 1.5 cycles/pixel, as shown in Fig. 17; thus the super-resolution multiplication value is equal to three. A comparison of the images captured by the aperture and after cubic interpolation in the same band in Fig. 18 shows that the latter does not increase the image information nor does it eliminate image artifacts; thus, its limit frequency is still the Nyquist frequency (equal to 0.5 cycle/pixel). Table III shows that the spectral distortion of the reconstructed results of our algorithm is small, thus ensuring the reconstructed spectral accuracy.

VII. CONCLUSION

Existing multispectral imaging technology cannot simultaneously achieve multispectral imaging with high spatial and temporal resolutions. A new multispectral imaging system known as CAMSRIS is proposed. The proposed multispectral camera array acquires low-resolution multispectral images, thus reducing cost and transmission bandwidth. These features are beneficial to the imaging of fast or even high-speed moving objects. The proposed system exploits the nonredundant spatial motion information between the apertures to perform super-resolution reconstruction, thus improving the spatial resolution of the captured images. Crab-HFP is used to align multispectral images from different imaging perspectives, and the SCSR algorithm is then used to reconstruct high-resolution multispectral images. CAMSRIS and MSFA qualitatively and quantitatively were analyzed based on different multispectral datasets. The results showed that the proposed system and the algorithm have better reconstruction quality than MSFA, a much shorter execution time than the demosaicking algorithms, and a more flexible band selection. Owing to the limitation of the adaptive kernel, the reconstructed results still suffer from jagged edges; thus, additional improvements are needed. In the future, we will extend the multispectral camera array to hyperspectral imaging. The proposed algorithm should also be accelerated to achieve real-time reconstruction and multispectral imaging at video frame rates.

ACKNOWLEDGMENT

The authors would like to thank Associate Professor Jing Wu, Weilan Huang, and Jianliang Chen for their assistance.

REFERENCES

[1] W. Cho, J. Jang, A. Koschan, M. A. Abidi, and J. Paik, "Hyperspectral face recognition using improved inter-channel alignment based on qualitative prediction models," *Opt. Exp.*, vol. 24, no. 24, pp. 27637–27662, 2016.
[2] R. T. Kester, N. Bedard, L. Gao, and T. S. Tkaczyk, "Real-time snapshot hyperspectral imaging endoscope," *J. Biomed. Opt.*, vol. 16, no. 5, May 2011, Art. no. 056005.
[3] B. Garavelli, A. Mencarelli, and L. Zanotti, "XSpectra: The most advanced real time food contaminants detector," in *Proc. IEEE Conf. Biomed. Circuits Syst.*, Oct. 2017, pp. 1–4.
[4] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for Earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.
[5] P. Colantoni, R. Pillay, C. Lahanier, and D. Pitzalis, "Analysis of multispectral images of paintings," in *Proc. 14th Eur. Conf. Signal Process.*, 2006, pp. 1–5.
[6] X. Briottet et al., "Military applications of hyperspectral imagery," in *Proc. SPIE*, vol. 6239, 2006, Art. no. 62390B.
[7] D. Manolakis, "Signal processing algorithms for hyperspectral remote sensing of chemical plumes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2008, pp. 1857–1860.
[8] D. N. Stratis, K. L. Eland, J. C. Carter, S. J. Tomlinson, and S. M. Angel, "Comparison of acousto-optic and liquid crystal tunable filters for laser-induced breakdown spectroscopy," *Appl. Spectrosc.*, vol. 55, no. 8, pp. 999–1004, 2001.
[9] N. Tack, A. Lambrechts, P. Soussan, and L. Haspeslagh, "A compact, high-speed, and low-cost hyperspectral imager," in *Proc. SPIE*, vol. 8266, 2012, p. 82660Q.
[10] O. Pust, "Innovative filter solutions for hyperspectral imaging: Linear variable bandpass filters specifically designed for hyperspectral imaging," *Optik Photonik*, vol. 11, no. 3, pp. 24–27, 2016.
[11] Y. Xu, J. Li, C. Bai, H. Yuan, and J. Liu, "Ultra-compact Fourier transform imaging spectrometer using a focal plane birefringent interferometer," *Opt. Lett.*, vol. 43, no. 17, pp. 4081–4084, 2018.
[12] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, 2008.

[13] H. Arguello and G. R. Arce, "Colored coded aperture design by concentration of measure in compressive spectral imaging," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1896–1908, Apr. 2014.

[14] A. Parada-Mayorga and G. R. Arce, "Colored coded aperture design in compressive spectral imaging via minimum coherence," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 202–216, Jun. 2017.

[15] X. Wang, Y. Zhang, X. Ma, T. Xu, and G. R. Arce, "Compressive spectral imaging system based on liquid crystal tunable filter," *Opt. Exp.*, vol. 26, no. 19, pp. 25226–25243, 2018.

[16] M. Hubold, E. Montag, R. Berlich, R. Brunner, and R. Brüning, "Multi-aperture system approach for snapshot multispectral imaging applications," *Opt. Exp.*, vol. 29, no. 5, pp. 7361–7378, 2021.

[17] N. Genser, J. Seiler, and A. Kaup, "Camera array for multi-spectral imaging," *IEEE Trans. Image Process.*, vol. 29, pp. 9234–9249, 2020.

[18] B. Geelen, N. Tack, and A. Lambrechts, "A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic," in *Proc. SPIE*, vol. 8974, 2014, p. 89740L.

[19] A. V. Kanaev, M. R. Kutteruf, M. K. Yetzbacher, M. J. Deprenger, and K. M. Novak, "Imaging with multi-spectral mosaic-array cameras," *Appl. Opt.*, vol. 54, no. 31, p. F149, 2015.

[20] L. Miao, H. Qi, R. Ramanath, and W. E. Snyder, "Binary tree-based generic demosaicking algorithm for multispectral filter arrays," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3550–3558, Nov. 2006.

[21] Y. Monno, M. Tanaka, and M. Okutomi, "Multispectral demosaicking using guided filter," in *Proc. SPIE*, Jan. 2012, p. 82990.

[22] Y. Monno, S. Kikuchi, M. Tanaka, and M. Okutomi, "A practical one-shot multispectral imaging system using a single image sensor," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3048–3059, Oct. 2015.

[23] Y. Monno, D. Kiku, M. Tanaka, and M. Okutomi, "Adaptive residual interpolation for color and multispectral image demosaicking," *Sensors*, vol. 17, no. 12, p. 2787, Dec. 2017.

[24] S. P. Jaiswal, L. Fang, V. Jakhetiya, J. Pang, K. Mueller, and O. C. Au, "Adaptive multispectral demosaicking based on frequency-domain analysis of spectral correlation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 953–968, Feb. 2017.

[25] X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2539–2543.

[26] Y. Liu, X. Yuan, J. Suo, D. J. Brady, and Q. Dai, "Rank minimization for snapshot compressive imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 2990–3006, Dec. 2018.

[27] A. Marquina and S. J. Osher, "Image super-resolution by TV-regularization and Bregman iteration," *J. Sci. Comput.*, vol. 37, no. 3, pp. 367–382, Dec. 2008.

[28] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, Aug. 2014.

[29] Y. G. Peng, J. L. Suo, Q. H. Dai, and W. L. Xu, "Reweighted low-rank matrix recovery and its application in image restoration," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2418–2430, Dec. 2014.

[30] Z. Zha, B. Wen, X. Yuan, J. Zhou, and C. Zhu, "Image restoration via reconciliation of group sparsity and low-rank models," *IEEE Trans. Image Process.*, vol. 30, pp. 5223–5238, 2021.

[31] G. Carles, J. Downing, and A. R. Harvey, "Super-resolution imaging using a camera array," *Opt. Lett.*, vol. 39, no. 7, pp. 1889–1892, 2014.

[32] N. L. Nguyen, J. Anger, A. Davy, P. Arias, and G. Facciolo, "Self-supervised multi-image super-resolution for push-frame satellite images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 1121–1131.

[33] B. Lecouat, J. Ponce, and J. Mairal, "Lucas–Kanade reloaded: End-to-end super-resolution from raw image bursts," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2370–2379.

[34] G. Carles, G. Muyo, N. Bustin, A. Wood, and A. R. Harvey, "Compact multi-aperture imaging with high angular resolution," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 3, pp. 9–411, Mar 1, 2015.

[35] J. Tanida et al., "Thin observation module by bound optics (TOMBO): Concept and experimental verification," *Appl. Opt.*, vol. 40, no. 11, pp. 1806–1813, 2001.

[36] M. W. Haney, M. P. Christensen, D. Rajan, S. C. Douglas, and S. L. Wood, "Adaptive flat micro-mirror array-based computational imaging architecture," in *Proc. Comput. Opt. Sens. Imag.*, 2005, p. CMB5.

[37] V. R. Bhakta, M. Somayaji, S. C. Douglas, and M. P. Christensen, "Experimentally validated high-resolution imaging with adaptive multi-aperture folded architecture," in *Proc. Comput. Opt. Sens. Imag.*, 2009, p. CWB5.

[38] M. Somayaji et al., "Field test of PANOPTES-based adaptive computational imaging system prototype," in *Proc. Comput. Opt. Sens. Imag.*, 2011, p. CPDP3.

[39] A. Andronache, M. von Siebenthal, G. Székely, and P. Cattin, "Non-rigid registration of multi-modal images using both mutual information and cross-correlation," *Med. Image Anal.*, vol. 12, no. 1, pp. 3–15, 2008.

[40] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[41] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision With the OpenCV Library*. Sebastopol, CA, USA: O'Reilly Media, 2008.

[42] P. Pinggera, T. Breckon, and H. Bischof, "On cross-spectral stereo matching using dense gradient features," in *Proc. Brit. Mach. Vis. Conf.*, vol. 2, 2012, p. 3.

[43] J. Holloway, K. Mitra, S. J. Koppal, and A. N. Veeraraghavan, "Generalized assorted camera arrays: Robust cross-channel registration and applications," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 823–835, Mar. 2015.

[44] S. Kim, S. Ryu, B. Ham, J. Kim, and K. Sohn, "Local self-similarity frequency descriptor for multispectral feature matching," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5746–5750.

[45] C. A. Aguilera, A. D. Sappa, and R. Toledo, "LGHD: A feature descriptor for matching across non-linear intensity variations," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 178–181.

[46] Z. Yi, C. Zhiguo, and X. Yang, "Multi-spectral remote image registration based on SIFT," *Electron. Lett.*, vol. 44, no. 2, pp. 107–108, Jan. 2008.

[47] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vis.*, vol. 37, no. 2, pp. 151–172, Jun. 2000.

[48] Y. Ye and J. Shan, "A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences," *ISPRS J. Photogramm. Remote Sens.*, vol. 90, pp. 83–95, Apr. 2014.

[49] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[50] Q. Ying-Dong, C. Cheng-Song, C. San-Ben, and L. Jin-Quan, "A fast subpixel edge detection method using Sobel–Zernike moments operator," *Image Vis. Comput.*, vol. 23, no. 1, pp. 11–17, Jan. 2005.

[51] L. D. Stefano, M. Marchionni, and S. Mattoccia, "A fast area-based stereo matching algorithm," *Image Vis. Comput.*, vol. 22, no. 12, pp. 983–1005, Oct. 2004.

[52] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[53] G. A. Puerto-Souza and G. L. Mariottini, "Hierarchical multi-affine (HMA) algorithm for fast and accurate feature matching in minimally-invasive surgical images," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 2007–2012.

[54] J. Franklin, "The elements of statistical learning: Data mining, inference and prediction," *Math. Intell.*, vol. 27, no. 2, pp. 83–85, 2005.

[55] W. E. L. Grimson and D. P. Huttenlocher, "On the sensitivity of the Hough transform for object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 3, pp. 255–274, Mar. 1990.

[56] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[57] U. Von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[58] J. P. Theiler and G. Gisler, "Contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation," in *Proc. SPIE*, vol. 3159, 1997, pp. 108–118.

[59] T. N. Tran, R. Wehrens, and L. M. C. Buydens, "SpaRef: A clustering algorithm for multispectral images," *Analytica Chim. Acta*, vol. 490, nos. 1–2, pp. 303–312, Aug. 2003.

[60] X. Wang and M. Leeser, "K-means clustering for multispectral images using floating-point divide," in *Proc. 15th Annu. IEEE Symp. Field-Program. Custom Comput. Mach. (FCCM)*, Apr. 2007, pp. 151–162.

[61] G. H. Ball and D. J. Hall, "ISODATA, a novel method of data analysis and pattern classification," Stanford Res. Inst., Menlo Park, CA, USA, 1965.

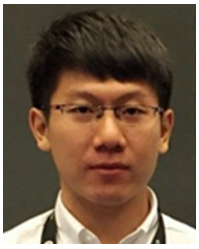
[62] S. Vassilvitskii and D. Arthur, "k-means++: The advantages of careful seeding," in *Proc. Annu. ACM-SIAM Symp. Discrete Algorithms*, 2006, pp. 1027–1035.

[63] H. Zhang, H. Shen, and L. Zhang, "A super-resolution reconstruction algorithm for hyperspectral images," *Signal Process.*, vol. 92, no. 9, pp. 2082–2096, Sep. 2012.
[64] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.
[65] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Evanston, IL, USA: Routledge, 2018.
[66] B. Wronski et al., "Handheld multi-frame super-resolution," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–18, Jul. 2019.
[67] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[68] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.
[69] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
[70] L. Bian, Y. Wang, and J. Zhang, "Generalized MSFA engineering with structural and adaptive nonlocal demosaicing," *IEEE Trans. Image Process.*, vol. 30, pp. 7867–7877, 2021.
[71] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402.
[72] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[73] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[74] I. Amidror, "Scattered data interpolation methods for electronic imaging systems: A survey," *J. Electron. Imag.*, vol. 11, no. 2, pp. 157–176, 2002.

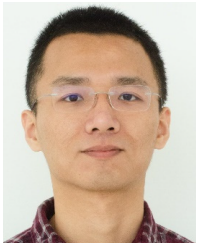
Feng Huang was born in Fuzhou, China, in 1979. He is currently with the School of Mechanical Engineering and Automation, Fuzhou University, China. His research interests include opto-electronic imaging.

Yating Chen received the B.S. degree from Hainan University, Haikou, in 2017. She is currently pursuing the Ph.D. degree with the School of Mechanical Engineering and Automation, Fuzhou University, China. Her research interests include computing imaging and computer vision.

Xuesong Wang received the B.S. degree from Guangdong Ocean University, Zhanjiang, in 2019. He is currently pursuing the Ph.D. degree with the School of Mechanical Engineering and Automation, Fuzhou University, China. His research interests include image super-resolution reconstruction and multimodal image fusion.



Shu Wang was born in Heilongjiang, China, in 1990. He received the Ph.D. degree in optics from the Key Laboratory of Optoelectronic Science and Technology for Medicine, Ministry of Education, Fujian Normal University, Fuzhou, China. He is currently with the College of Mechanical Engineering and Automation, Fuzhou University, Fuzhou. His main research interests include the polarization/spectral imaging techniques and the applications of nonlinear optical microscopy in biomedical research.



Xianyu Wu received the Ph.D. degree in mechanical engineering from North Carolina State University. His laboratory works on the fundamental understanding of optical sensing and imaging in precision measurement systems, and the application of this knowledge toward problems in non-destructive testing, structural health monitoring, and rapid diagnostics for mechanical systems. His research team is particularly interested in developing next-generation optical measurement and imaging systems that combine advances in hardware and software.