Infrared Physics & Technology 90 (2018) 146–155

Contents lists available at ScienceDirect

Infrared Physics & Technology

journal homepage: www.elsevier.com/locate/infrared

Regular article

Super resolution reconstruction of infrared images based on classified dictionary learning

Fei Liu [a,b], Pingli Han [a], Yi Wang [a,c], Xuan Li [a], Lu Bai [a], Xiaopeng Shao [a,*]

[a] School of Physics and Optoelectronic Engineering, Xidian University, Xi'an 710071, China
[b] State Key Laboratory of Applied Optics, Chang Chun Institute of Optics, Fine Mechanics and Physics, Chinese Academic of Science, Changchun 130033, China
[c] Chinese Flight Test Establishment, Xi'an 710089, China

HIGHLIGHTS

- Propose an infrared image reconstruction way by training classified dictionary pairs.
- Visible-light image samples are used to compensate detail missing in infrared images.
- Each dictionary pair are trained based on an extracted feature of samples.
- High efficiency and less time cost in image reconstruction.

ARTICLE INFO

Article history:
Received 20 December 2017
Revised 28 February 2018
Accepted 5 March 2018
Available online 8 March 2018

Keywords:
Infrared imaging
Super-resolution
Dictionary learning
Classified dictionaries

ABSTRACT

Infrared images always suffer from low-resolution problems resulting from limitations of imaging devices. An economical approach to combat this problem involves reconstructing high-resolution images by reasonable methods without updating devices. Inspired by compressed sensing theory, this study presents and demonstrates a Classified Dictionary Learning method to reconstruct high-resolution infrared images. It classifies features of the samples into several reasonable clusters and trained a dictionary pair for each cluster. The optimal pair of dictionaries is chosen for each image reconstruction and therefore, more satisfactory results is achieved without the increase in computational complexity and time cost. Experiments and results demonstrated that it is a viable method for infrared images reconstruction since it improves image resolution and recovers detailed information of targets.

1. Introduction

Infrared imaging is widely applied to various fields such as remote sensing [1], medical imaging [2,3], military reconnaissance and target recognition [4,5], benefiting from abundant target information that is not available in visible-light images. However, due to its long wavelength (700–1050 nm) infrared images always suffer from serious low-resolution problems, leading to miss of details including texture, contexture, edge information, etc. This problem limits the applications of infrared imaging. Efforts have been made to meet the increasing demands for high-resolution (HR) infrared images, leading to the development of the super-resolution (SR) reconstruction technology. Super-resolution techniques such as stimulated emission depletion (STED) microscopy [6,7] and localization microscopy (e.g., PALM, STORM) [8] remove its intrinsic resolution limitation by using fluorescence label, and have been widely used in biology. Nevertheless, without employment of any fluorescence label and requirement to make changes in hardware, SR reconstruction based on signal processing is a promising technology to achieve HR images from the raw low-resolution (LR) images [9–11]. In general, one or more frames of LR images are required to reconstruct HR images.

During the evolution of SR reconstruction algorithms, two major categories, the model-based algorithm [12] and the learning-based algorithm [13], are developed. The former one is advantageous in reconstruction speed, but is sensitive to image blur, noise or image displacement. Desired results are often not achievable [14,15]. The learning-based algorithm has been widely investigated since appearance. It attempts to achieve reasonable details of the reconstructed images by searching and matching a database composed of training images [16,17]. A valuable method, denoted as sparse-coding-based SR (SCSR) algorithm, was proposed based on the compressed sensing (CS) theory [18–20]. It considerably promoted the development of learning-based SR studies. Special attention has been paid to the dictionary learning

* Corresponding author.
E-mail address: xpshao@xidian.edu.cn (X. Shao).

https://doi.org/10.1016/j.infrared.2018.03.008
1350-4495/

based methods, including single dictionary learning [21], dual-dictionary learning [19], multi-scale dictionary learning [22], and adaptive dictionary learning [23], which enables to improve the spatial resolution of the reconstructed images. These methods benefit from the ability of the SR reconstruction algorithms based on dictionary learning in combining the prior knowledge of the LR images. Efforts were made to build optimal dictionaries which eventually determine the final reconstruction results. It's not easy due to the contradiction between satisfactory reconstruction results and low time cost. For instance, better reconstruction results often need a dictionary with a huge amount of atoms, which in turn costs more time to complete the process. Besides, we find that current dictionary-learning-based SR reconstruction methods trained only one dictionary pair for all images without consideration of diverse images and differences in image contents. This damages the matching accuracy between the trained dictionary pair and LR images to be reconstructed, which further affects the quality of reconstructed images.

This study aims to reconstruct HR infrared images through a classified dictionary learning method. The features of training samples were first clustered into several groups by K-means method [24]. After being trained offline by K-means singular value decomposition (K-SVD) algorithm [25], the dictionary pair which has the highest similarity with the original LR infrared image was selected to conduct the image reconstruction online. Several groups of experiments were conducted and the results showed the images reconstructed by the proposed method were excellent on both subjective vision perception and objective evaluation value.

2. Principles of the method

2.1. SR reconstruction inspired by sparse approximations

Before introducing how to reconstruct infrared images, we first investigate how an image degrades into an LR one. In a digital imaging system, an optical low-pass filter (OLPF) is always employed between the lens and the image sensor to eliminate Moiré fringes [26]. This phenomenon originates from the frequency limit of the image sensor, called Nyquist frequency, beyond which the signal passed by the optical lens cannot be resolved. The OLPF stops the optical system from passing high-frequency information beyond the Nyquist frequency of the sensor to the sensor, thus only low-frequency information is received [16,27–29]. Fig. 1 shows the major events during imaging process of such imaging system [30]. First, a raw sample at or above the Nyquist rate is denoted with x in the following explanation. Then it was deteriorated via warping, blurring, downsampling, and noise, and finally reached the sensor. The deteriorated LR image was denoted as y in the following parts. Starting with this LR image, SR reconstruction techniques seek to restore high-frequency information of the image without updating the sensor.

The imaging process can be modeled in mathematics and expressed in Eq. (1), where each image deterioration event is denoted by a matrix. For instance, the downsampling process is denoted as matrix S, the optical blur by matrix H, the relative motion between the target and the imaging system by B, and the noise during imaging by matrix N. Eq. (1) provides a further simplified form of the model after ignoring the relative motion between a target and the imaging system considering it is neglectable for a single-frame image.

$$y = SHBx + N = SHx + N \tag{1}$$

It's almost impossible to obtain the original HR image x by directly solving Eq. (1) which presents an ill-posed problem only with known y, i.e., the captured LR image, while S, H and N *are* kept unknown. Inspired by the CS theory, the sparse representation is imported to calculate HR image. The goal of sparse representation is to express a given signal into a linear combination of a small number of atoms where are prior- registered in a database called the dictionary [31]. In specific, an image $x \in R^n$ can be sparsely represented by an overcomplete dictionary $D = [d_1, d_2, \ldots, d_m] \in R^{n \times m}$ ($n < m$), where d_i is a vector with $i = 1, 2, \ldots, m$, as shown in Eq. (2).

$$x = D\alpha, \ \| \alpha \|_0 << n \tag{2}$$

Along with the dictionary D is the matrix of the sparse representation coefficients $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_m]^T \in R^m$, which follows the constraint $||\alpha||_0 << n$. And $||\alpha||_0$ denotes the number of nonzero elements in matrix α. Eq. (2) clearly indicates that the known dictionary D and sparse representation coefficients α enable to reconstruct a HR image, which leaves the major work in training reasonable dictionary and computing the optimal sparse representation coefficients.

From Eqs. (1) and (2), the degraded LR image can also be sparsely represented, as shown by Eq. (3),

$$y = SHD\alpha = LD\alpha \tag{3}$$

where L stands for the degradation of an HR image x into an LR image y. In fact, the product of LD can be denoted with a dictionary D_l, indicating features of LR images. Correspondingly, the dictionary D in Eq. (2) involves features of HR images, which can be further specified into D_h. Besides, $D_l = LD_h$. With this dictionary pair, both the HR image and the corresponding LR image can be sparsely represented by the same sparse representation coefficients, as shown in Eq. (4).

$$x = D_h\alpha, \quad y = D_l\alpha \tag{4}$$

Major work should be focused on training the optimal dictionary pair, the HR dictionary and the corresponding LR dictionary, to reconstruct HR images. Then the sparse representation coefficients under the given dictionary pair are achievable based on the principle of Eq. (5). Here, the symbol T controls the degree of sparse representation.

$$\alpha = \arg\min_{\alpha} \| y - D_l\alpha \|_p \ s.t. \ \| \alpha \|_0 << T \tag{5}$$

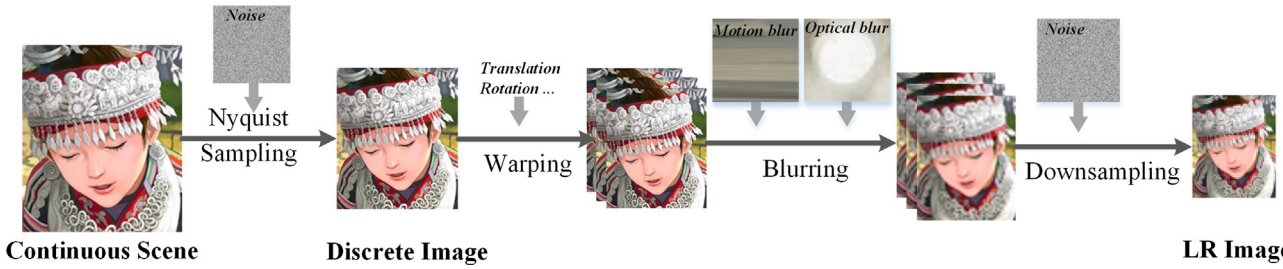


Fig. 1. General observation model of image deterioration process.

Finally, the HR image can be calculated based on the trained dictionaries and the computed sparse representation coefficients. Fig. 2 illustrates the procedures of the proposed method. Two major parts are included, the classified dictionary learning and the reconstruction algorithm, which will be detailed in the following parts.

2.2. Classified dictionary learning

As discussed before, the proposed classified dictionary learning method pursues satisfactory reconstruction results by optimizing the trained dictionaries. Although images are diverse in appearance, they can be generally characterized by several specific features, such as wavelength, phase shift, color, and so on. Inspired by this phenomenon, we make efforts to take advantage of the similarity in these features by first reasonably dividing them into several clusters, then along with each cluster, one pair of dictionaries are trained, which is composed of an HR dictionary and a corresponding LR dictionary. The reconstruction of a target image can be time efficient with employment of the optimal pair of dictionaries determined by feature extraction and membership grade calculation, since only one fraction of all the items in the trained dictionaries are involved during the reconstruction of each image patch. Besides, since the classified dictionaries introduced and utilized similarities of features in the samples, it enables more accurate reconstruction results. The detailed procedures of dictionary training are presented in Fig. 3.

(i). Samples clustering

Accurate and valid reconstruction results rely on reasonable feature clustering, which further determines the accuracy of the trained dictionaries. The dictionary training starts with partitioning each image sample into p HR patches, then all the original HR sample patches can be indicated by $\{x_i\}$, where $i = 1, 2, \ldots, p$. Gabor filter shown by Eq. (6) is used to extract features of $\{x_i\}$, including the wavelength λ, the phase offset φ, deviation σ, and the spatial aspect ratio γ.

$$\begin{cases} g(X, Y; \lambda, \theta, \varphi, \sigma, \gamma) = exp\left(-\frac{X'^2+\gamma^2Y^2}{2\sigma^2}\right)exp\left(i\left(2\pi\frac{X'}{\lambda}+\varphi\right)\right) \\ X' = X\,cos(\theta) + Y\,sin(\theta); Y' = -X\,sin(\theta) + Y\,cos(\theta) \end{cases} \tag{6}$$

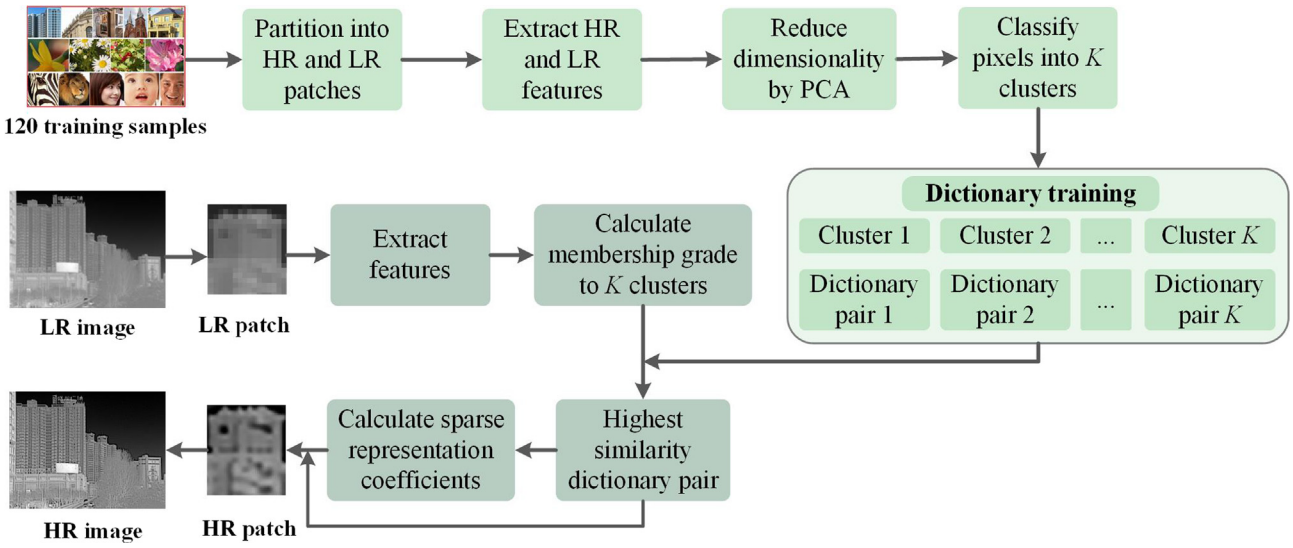


Fig. 2. Flowchart of SR reconstruction by the classified dictionary learning method.

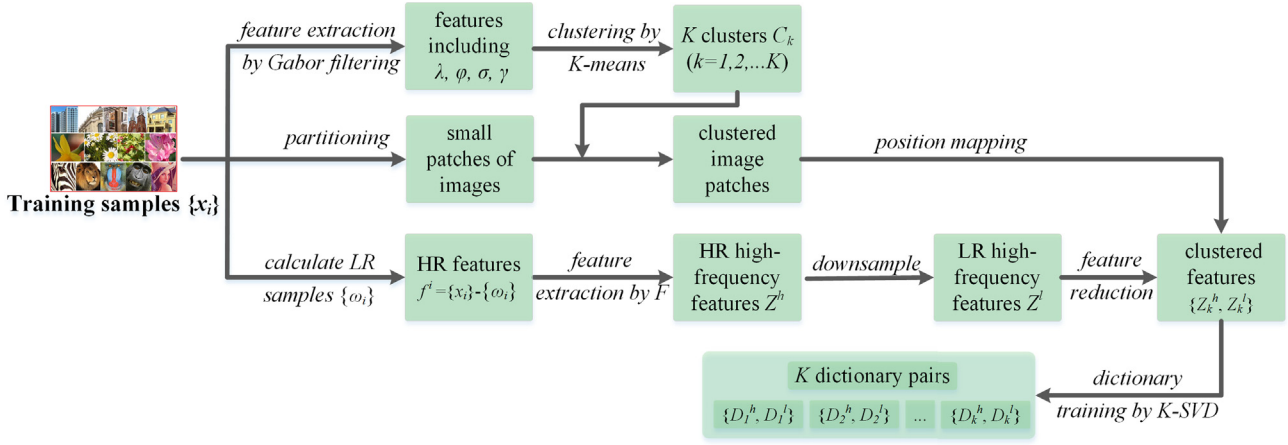


Fig. 3. Detailed process of classified dictionary training.

where (X, Y) represents the Cartesian coordinate system, and θ indicates the rotation angle between the standard Gabor filter stripes and the employed filter stripes. Fig. 4 exhibits an example of Gabor filters at diverse orientations and Fig. 5 illustrates the features extracted by these filters. In this example, only the orientation of Gabor filter was taken as a variable. Similarly, replacement of variables enables extraction of other features of the samples.

The extracted features by Gabor filter are then used for clustering. Since clustering doesn't depend on specific standards, it classifies pixels in terms of similarity in each other. It's necessary to determine the cluster number at first. For this sake, two indexes, Davies-Bouldin (DB) index and Calinski-Harabasz (CH) index in Eq. (7) are exploited [24].

$$\begin{aligned} CH(i) &= \frac{trB(i)/(i-1)}{trW(i)/(k-i)} \\ DB(k) &= \frac{1}{k}\sum_{i=1}^{k} \max_{j=1k;j=i}\left(\frac{R_i+R_j}{C_{ij}}\right) \end{aligned} \tag{7}$$

In the expression, for CH index, k is the total number of clusters; i and j *are* the index of the i *th* and j *th* cluster; $B(i)$ is the between-cluster dispersion matrix; $W(i)$ is the within-cluster dispersion matrix; tr calculates the trace of a matrix. While for DB index, R_i is the mean value of distances between data in i *th* cluster and the cluster center C_i; similarly, R_j is the mean value of distances between data in j *th* cluster and the cluster center C_j, and C_{ij} is the distance between C_i and C_j.

The higher values of CH index indicate more accurate clusters, which are characterized by high similarity in the same cluster and significant difference between clusters. In a contrary way, the lower DB values mean lower similarity between clusters, i.e., the best result is accessible when DB reaches the minimum. According to previous research, the optimal cluster number is within 2–6 [32]. From data in Table 1, we can reach the conclusion that the optimal cluster number should be 5.

The practical clustering procedure is completed by K-means cluster algorithm [33,34]. It first randomly selects K ($K = 5$ in this

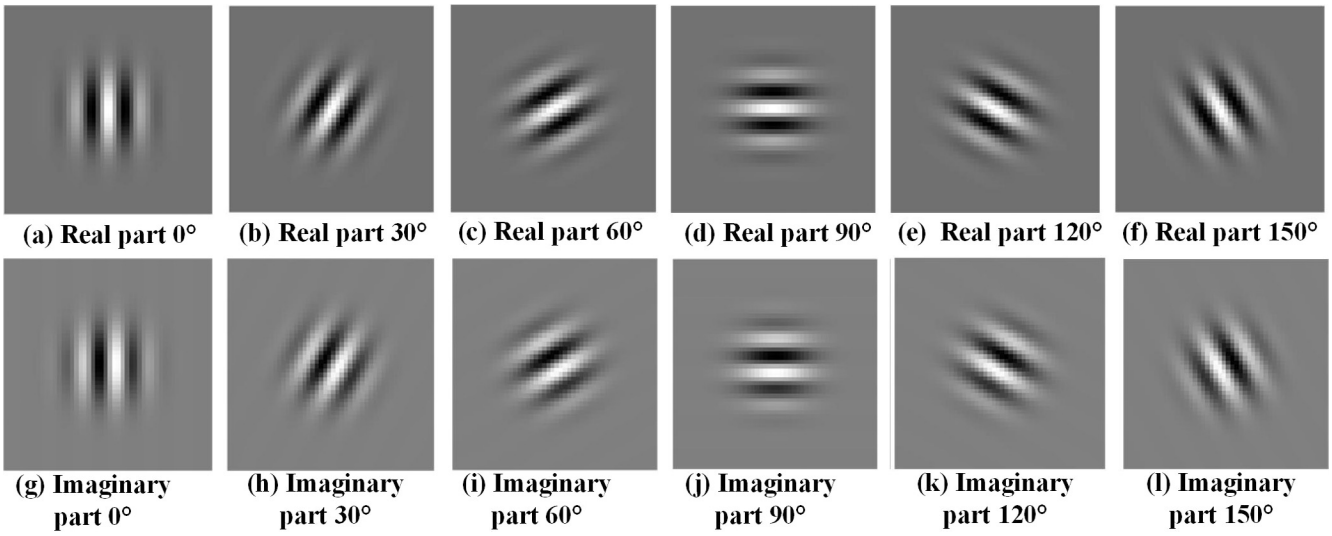


Fig. 4. Gabor function at different orientations.

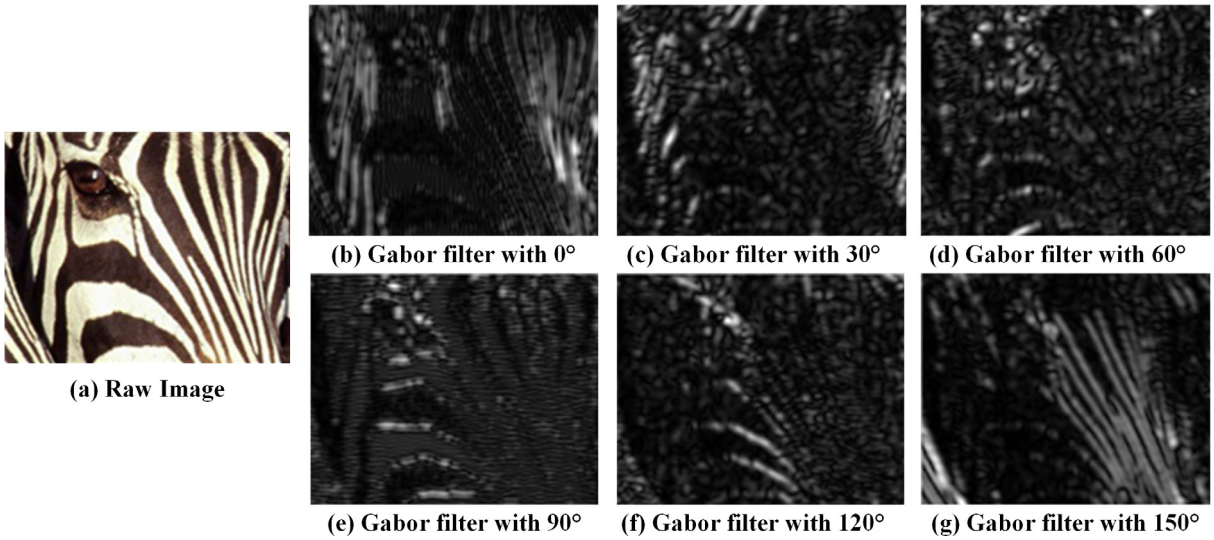


Fig. 5. Filtering results with Gabor functions at different orientations.

Table 1
DB and CH index in this study.

K-Value	1	2	3	4	**5**	6
DB index	–	2.3486	2.7616	2.4465	**2.3138**	3.2760
CH index	–	543.27	420.13	507.85	**574.93**	359.44

study as discussed above) pixels from the training samples as initial clustering centers C_k (k = 1, 2, …, k) for the designated K clusters. Afterwards, the Euclidean distances from every pixel to the clustering centers C_k are computed [35]. Place one pixel into the cluster whose center is the nearest compared to the others. Then the clustering centers C_k require updates for more accurate clusters based on Eq. (8),

$$\overline{D}_i = \frac{1}{n-1} \times \sum_{i \neq j, j=1}^{n} d_{ij},\ i = 1, 2, ..., n \tag{8}$$

where $\overline{D}_i$indicates the average distance of pixels between each other in one cluster, n refers to the total number of pixels in the cluster, and d_{ij} means the Euclidean distance between x_i and x_j. For the k *th* cluster, when $\overline{D}_i$reaches the minimum, the corresponding i *th* pixel is then updated as the new cluster center. All the K cluster centers are updated by this way. Check the difference between the updated cluster center and the replaced cluster center, if it's small enough, the clustering results are feasible for training classified dictionary pairs; if not, then clustering should be restarted by calculating the Euclidean distances.

(ii). Dictionary training

To extract the high-frequency features in HR patches $\{x_i\}$, image downsampling to $\{x_i\}$ is conducted employing the bicubic method by a factor of 3 in each dimension, followed with upsampling in the same method which restores them to the original size but with reduced resolution, denote the results with $\{\omega_i\}$, i.e., the corresponding LR training patches. A subtraction of ω_i from x_i gets to the HR features, f^i, as shown in Eq. (9).

$$f^i = \{x_i\} - \{\omega_i\} \tag{9}$$

A 2 × 2 filter bank F composed of four different sub-operators in Eq. (10) is used to extract the high-frequency features in f^i.

$$F = [f_1, f_2, f_3, f_4] \tag{10}$$

The four sub-operators are listed in Eq. (11), where LoG represents a 5 × 5 Gaussian-Laplacian operator.

$$f_1 = [1, -1],\ \ f_2 = f_1^T,\ f_3 = LoG,\ f_4 = f_3^T \tag{11}$$

After filtered by F, high-frequency features of the training samples can be extracted and expressed into a feature matrix $Z = [Fx_1, Fx_2, \ldots, Fx_p]$. According to the clustering results obtained by K-means strategy, partition Z into small grids and map the pixels in Z to the corresponding K clusters. Each cluster is accompanied with a pair of feature matrices $\{Z_k^h, Z_k^l\}$, where $Z_k^h = [Fx_1^k, Fx_2^k, \ldots, Fx_p^k]$ (k = 1,2,…K) are K high-frequency feature matrices of the HR patches, and Z_k^l (k = 1,2,…, K) are K corresponding high-frequency feature matrices of the LR patches derived by downsampling Z_k^h. To reduce the time cost of dictionary training, Principle Components Analysis (PCA) algorithm is applied to reduce the dimensionality of Z_k^l [36]. Finally, K pairs of dictionaries $\{D_k^h, D_k^l\}$ can be trained by Eq. (12), corresponding to each cluster.

$$\min_{D^k, \alpha_i} \{\| Z_k - FD_k\alpha_i^k \|\},\ \text{s.t.}\ \| \alpha_i^k \|_0 \leqslant T_k \tag{12}$$

In Eq. (12), D_k=$\{D_k^h, D_k^l\}$ is the dictionary pair corresponding to the k *th* cluster, which is the key to perform the super-resolution image reconstruction. Here, D_k^h is the HR dictionary and D_k^l the LR dictionary; α_i^k represents the sparse representation coefficients of the k *th* dictionary pair and T_k is constant that controls the degree of sparse representation. The problem in Eq. (12) can be solved by the K-SVD strategy [25], which finally completes the dictionary training process.

2.3. SR reconstruction process

The learning process in Section 2.2 trained K pairs of dictionaries. An LR image can be reconstructed into an HR one by selecting the optimal pair of the dictionaries and applying a reconstruction algorithm.

Fig. 6 schematically illustrates the reconstruction process: An LR image is firstly partitioned into m small patches y_i (i = 1, 2, …, m) with a certain size of $Q \times Q$ pixels. According to the optical

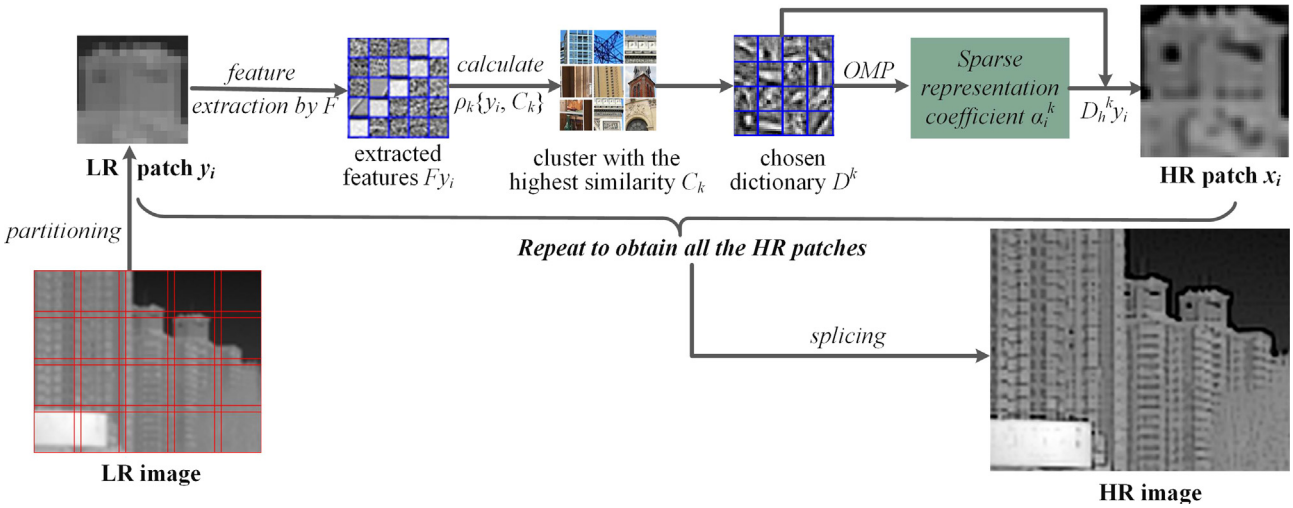


Fig. 6. SR reconstruction flow by the classified dictionary learning.

observation model discussed in Section 2.1, each LR patch y_i is further expressed by Eq. (13):

$$y_i = SH\widehat{y}_i + N \tag{13}$$

$\widehat{y}_i$ is the ideal HR patch we attempt to reconstruct. For each LR patch, the most appropriate pair of dictionaries should be selected. It's completed based on the similarity between the LR patch y_i and the K clusters. For i *th* LR patch y_i, it's firstly enlarged to three times its original size by upsampling. Then the membership grades of y_i to each cluster are computed with known Euclidean distances between y_i and the cluster centers C_k (k = 1,2,...,K). Eq. (14) shows how to determine the optimal dictionary pair $D^k = \{D_h^k, D_l^k\}$ for ith LR patch, where $\rho_k\{y_i, C_k\}$ denotes the membership grade.

$$D^k = max(\rho_k\{y_i, C_k\}),\ k = 1, 2, ..., K,\ i = 1, 2, ..., m \tag{14}$$

After the dictionary pair D^k for reconstruction of y_i is determined, the crucial problem left is to compute the corresponding sparse representation coefficients α_i^k. The filter operator F explained in Eqs. (10) and (11) is adapted to extract its high-frequency features $\{Fy_i\}$. Subsequently, dimensionality reduction to $\{Fy_i\}$ is also completed by PCA algorithm for less time cost. The sparse representation coefficients α_i^k are computed by Eq. (15) based on the Orthogonal Matching Pursuit (OMP) method [37,38]. And finally, the HR patch $\widehat{y}_i$is then achievable through multiplying D_h^i by α_i^k.

$$\alpha_i^k = \arg\min_{\alpha_i^k} \| Fy_i - FSHD_l^k\alpha_i^k \|_2^2,\ \text{s.t.} \| \alpha_i^k \| \leqslant T_0 \tag{15}$$

$$\widehat{y}_i^k = D_h^k\alpha_i^k \tag{16}$$

Eq. (16) implies that each of the LR patches can be processed by the above procedures for the HR patches $\widehat{y}_i$(i = 1,2,...,m), and consequently a whole HR image is obtained after splicing all the HR patches together.

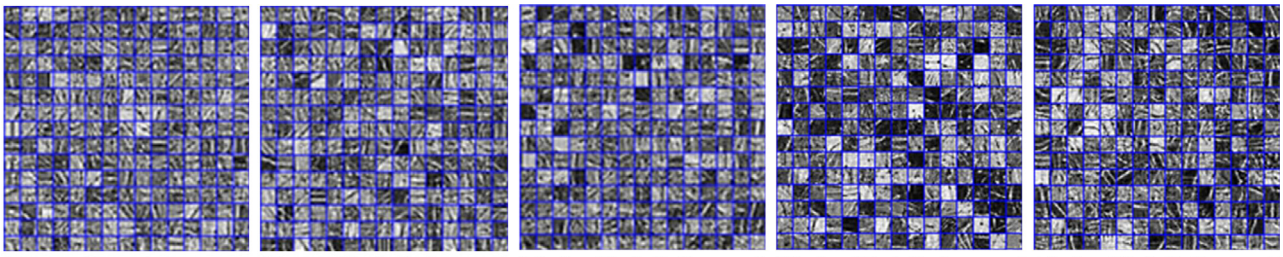
Fig. 7. Partial training samples.

3. Experiments and results

To verify the effects of the proposed classified dictionary learning method, 120 visible-light images are employed as training samples to train the dictionary pairs, of which 60 images are from the internet, and the rest are taken by the authors. Fig. 7 lists some examples of the training samples. They are characterized by broad frequency ranges and abundant contents involving architectures, plants, animals, human faces and so on. Trained dictionaries from these samples can make up the detail deficiency in infrared images. Each sample image is first processed by feature extraction and then partitioned into small grids with a size of 4 × 4 pixels. 66,119 grids are totally generated by the 120 training samples. The grids are then divided into 5 clusters, each of which is composed of 11,714, 13,460, 13,909, 13,049 and 13,987 grids, respectively. Therefore, 5 dictionary pairs are trained. Each dictionary contains 1000 atoms for both effectiveness and efficiency. Partial atoms in the 5 HR dictionaries are exhibited in Fig. 8.

In the experiment, infrared images of 2 diverse scenes were first collected by a *FLIR* infrared camera for testing, which had a resolution of 692 × 520 shown in Fig. 9(a) and (b). LR images of the two scenes were created by downsampling to reduce the pixels in each dimension into 1/3 of the original number. This is a good choice for demonstration of reconstruction results since the original captured images make a good comparison.

As refered before, Fig. 9 (a) and (b) are the original and downsampled pictures of a distant street scene. In comparison, Fig. 9 (c) and (d) present the original and downsampled images of a close building. These images are 'veiled' due to the limitations of infrared imaging. Accordingly, before the reconstruction step, preprocess of the LR image was performed with regard to the loss of details in the infrared images. By applying Histogram Double Equalization (HDE) and edge enhancement algorithm, the LR image tends to present much more available details. Afterwards, the preprocessed LR images were reconstructed by the Bicubic method, the Zeyde's method and the proposed method, respectively. The corresponding results to Fig. 9(b) were listed in Fig. 10(a–c), and those to Fig. 9(d) in Fig. 10(d–f), respectively. Visual feeling of these reconstructed images is better than that of Fig. 9. Of note, the images in Fig. 10 present quite clear appearances and dramatically enhanced contrast without the 'veil' in Fig. 9 and the original HR image. Besides, rich details are appreciable in Fig. 10, for example, the group of buildings in the distant street scene with clear texture. The top of one building marked by a small red rectangle is enlarged for clearer demonstrations. In Fig. 10(a), the result reconstructed by the Bicubic method, much more details including the edges and windows are visualized in contrast with the original HR image. We considered these results encouraging but yet can be improved. Fig. 10(b) shows an improved result after reconstruction by Zeyde's method, which effectively handled the blurring problem in Fig. 10 (a). Even though, it still suffers from background noise and there're some stripes in the margin area. The yielded image by the



(a) classified dictionary 1 (b) classified dictionary 2 (c) classified dictionary 3 (d) classified dictionary 4 (e) classified dictionary 5

Fig. 8. Partial atoms of the 5 trained HR dictionaries.

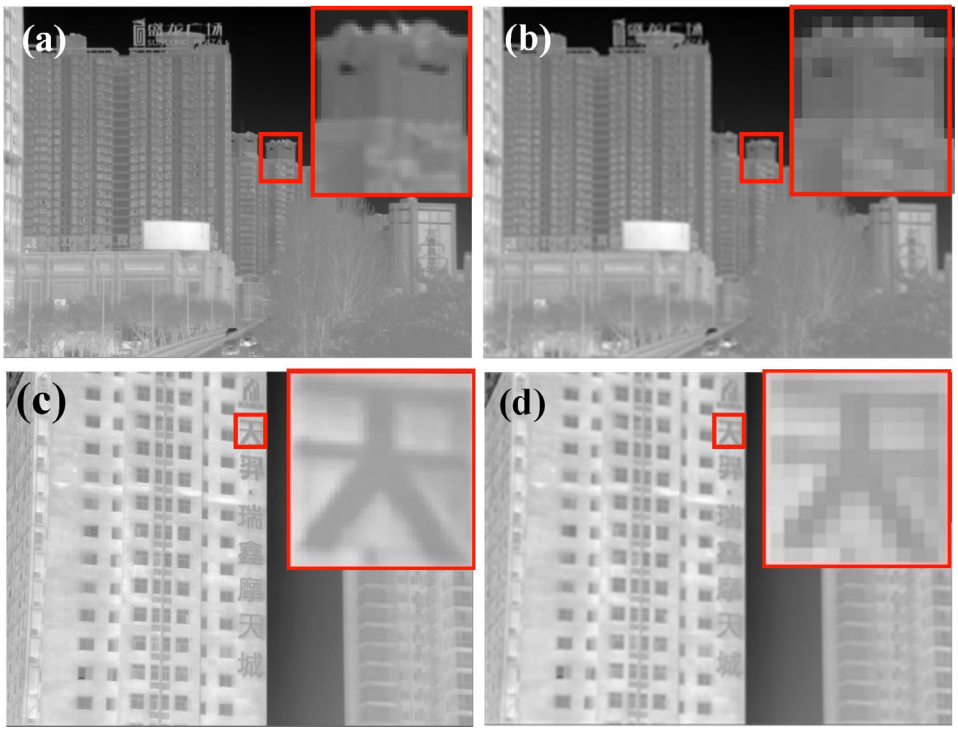


Fig. 9. Captured raw HR images and downsampled LR images. (a) Raw image of a distant street scene; (b) Downsampled LR image of (a); (c) Raw image of a close building; (d) Downsampled LR image of (c).

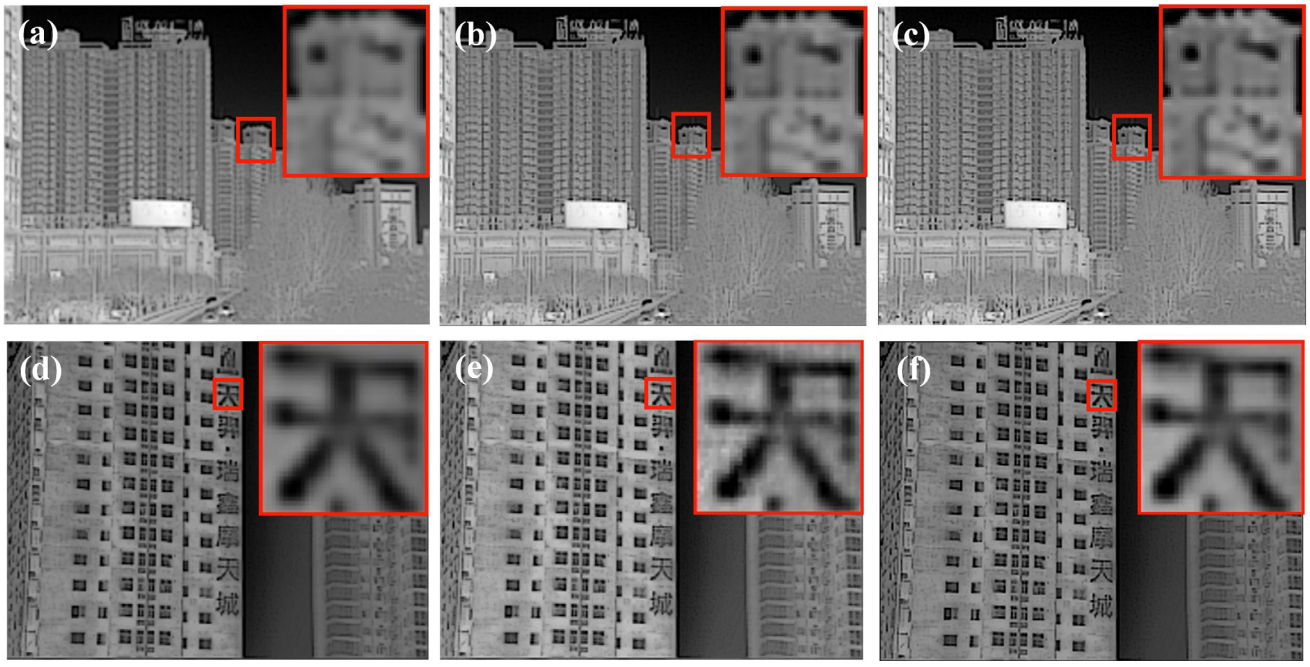


Fig. 10. (Upper row) Reconstructed HR images of Fig. 9(b) by Bicubic, Zeyde's and the proposed method; (Bottom row) Reconstructed HR images of Fig. 9(d) by Bicubic, Zeyde's and the proposed method.

proposed method is freed from these troubles as Fig. 10(c) shows. The example of the close building makes a more convincing comparison of the methods. All the reconstructed images show a clear appearance, but the result from the Bicubic method is still blurred and that from the Zeyde's method gets serious noise in it.

Numerical analysis is implemented to evaluate the reconstruction by calculating the following indexes: peak signal to noise ratio (PSNR) and Mean Structural Similarity (MSSIM). By PSNR, the gray similarity between two images is measured; and MSSIM presents the content structural similarity between two images. Employment of these two parameters makes an evaluation of two images from both the gray value and structure. In Table 2, the listed data are calculated from the six reconstructed images in Fig. 10. Test 1 indicates the distant street scene and test 2 the close building

Table 2
Numerical comparisons of reconstruction results.

Image	Parameter	Bicubic	Zeyde	Proposed method
Test 1	PSNR	31.529	38.033	39.396
	MSSIM	0.790	0.795	0.831
Test 2	PSNR	31.790	35.086	35.723
	MSSIM	0.823	0.833	0.859

scene. According to the definitions of the indexes, higher values of PSNR and MSSIM indicate higher similarities between two images. The data in Table 2 show accordant tendencies for PSNR and MSSIM that the proposed method provides reconstruction results closer to the original HR images.

Till now, the analyses above are all based on the LR images downsampled from HR images. Therefore, in the following parts, analyses based on captured LR infrared images are completed. We chose two different scenes shown as Fig. 11(a) and (e) in experiments. Taking Fig. 11(a) (captured by the *FLIR* camera) as an example, before reconstruction it was also preprocessed by HDE and edge enhancement algorithm. The reconstructed images in Fig. 11(b)–(d) were from the Bicubic method, Zeyde's method and the proposed method. Details regarding a lightning rod on top of a building were enlarged for comparison, which present sim-

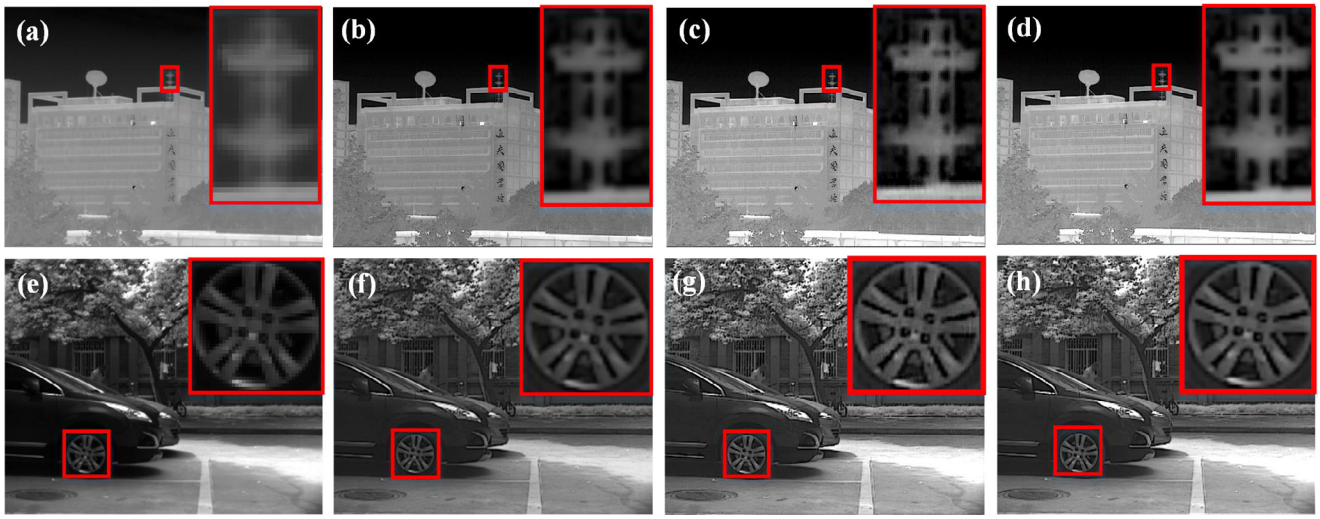


Fig. 11. Two different examples of infrared image reconstruction. (Upper row) (a) A raw LR image of a distant building; (b) (c) and (d) Reconstructed results of (a) by the Bicubic, Zeyde's and the proposed method; (Bottom row) (e) A raw LR image of close scene; (f) (g) and (h) Reconstructed results of (e) by the Bicubic, Zeyde's and the proposed method.

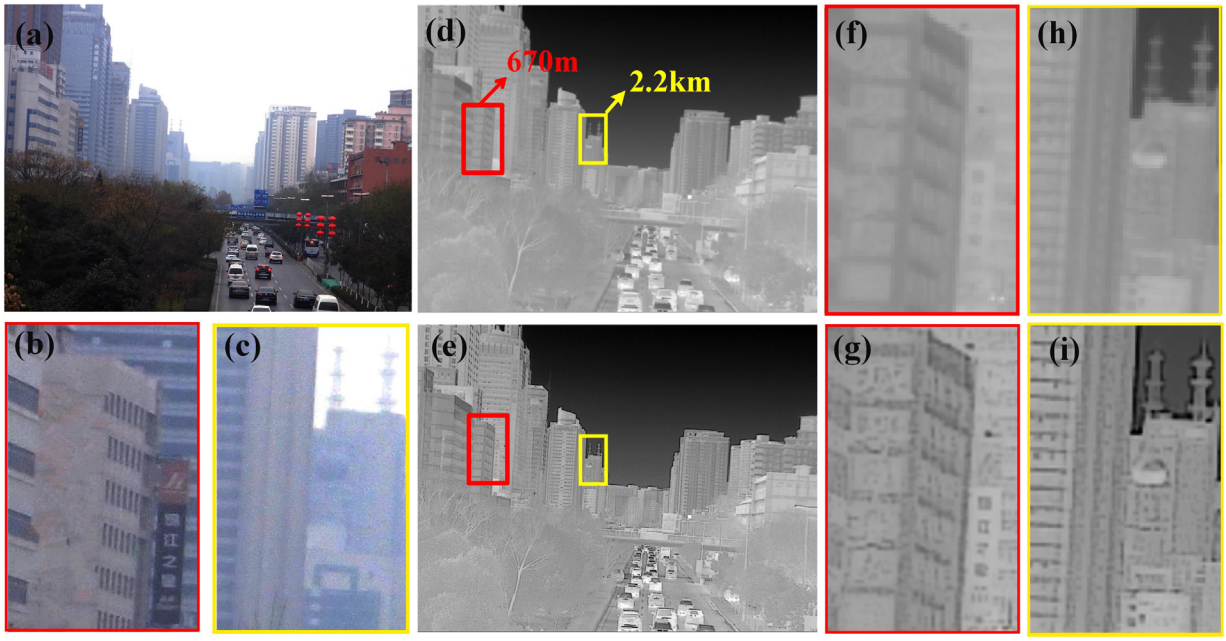


Fig. 12. (a) A raw visible-light image of a distant street scene; (b) Zoom-in view of a building at a distance of 670 m in (a); (c) Zoomed-in view of a building at a distance of 2.2 km in (a); (d) A raw LR infrared image of (a); (e) Reconstructed result of (a) by the proposed method; (f) Raw LR infrared image of (b); (g) Reconstructed result of (f); (h) Raw LR infrared image of (c); (i) Reconstructed result of (h).

ilar results to the above analyses. In the original LR image, details are almost indistinguishable. While among the reconstructed images, Fig. 11(d) presents the most satisfactory result without blurring in Fig. 11(b) and noise in Fig. 11(c). The other scene in Fig. 11(e) focused on a car at a short distance, which was captured by the camera XEVA-2.35 from *Xenics infrared solution company* with a resolution of 320*256. The enlarged wheel hub areas from the four images made a more intuitive comparison, of which the noise in Fig. 11(g) reconstructed by Zeyde's method is especially serious. These two groups of results demonstrated the positive ability of the proposed method in infrared image reconstruction.

The last experiment was conducted at a distant street scene which included complex information shown in Fig. 12(a). The raw infrared image of the distant street scene was shown in Fig. 12(d), and the reconstructed result in Fig. 12(e) was obtained by the proposed method. The original infrared image has a smooth appearance but indistinct details including edges, contours and so on. Two buildings at different distances were chosen to evaluate the reconstruction results, and their corresponding areas in the images were marked out and enlarged. The close one locates at 670 m away and the distant one 2.2 km away. The enhancement of the reconstructed image can be indicated by the Chinese characters on the right side of the close building shown by Fig. 12(g). They cannot even be seen in the original image, while after reconstruction they are distinguishable. The analogous result is also achievable in the distant building area.

4. Conclusion

For effective and efficient infrared image reconstruction, an SR reconstruction method based on classified dictionary learning was designed in this study. It originates from the idea of sparse representation and is realized by training several pairs of dictionaries according to feature clusters. During image reconstruction, the optimal dictionary pair is firstly determined in terms of the highest similarity between the LR image and the clusters. By this dictionary pair, the HR image can be reconstructed without increase in computational complexity and time cost. Satisfactory reconstruction results are obtained with reasonable resolution and distinguishable details. This method is insensitive to the contents or classifications of input images, and has the ability of enhancement advantages in resolution and vision feeling. Thus, this method has the potential to be applied to infrared images reconstruction captured in diverse situations. The further study will be focused on optimizing the dictionary training process to guarantee more reasonable clustering, and further to improve the reconstruction efficiency.

Funding

The Project was supported by the State Key Laboratory of Applied Optics (CS16017050001); Natural National Science Foundation of China (NSFC) (61475123, 61705175); the 61st General Program of China Postdoctoral Science Foundation (2017M613063); Fundamental Research Funds for the Central Universities (JB170503).

Conflict of interest statement

The authors declare that they have no conflict of interest to this work.

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work.

Acknowledgment

The authors acknowledge Rui Gong for the helpful guidance and meaningful discussion. We especially thank Juanjuan Zheng for the language revision.

References

[1] W. Wu, X.M. Yang, K. Liu, Y.G. Liu, B.Y. Yan, H. Hua, A new framework for remote sensing image super-resolution: sparse representation-based method by processing dictionaries with multi-type features, J. Sys. Architect. 64 (2016) 63–75.
[2] R. Heintzmann, M.G.L. Gustafsson, Subdiffraction resolution in continuous samples, Nat. Photon. 3 (2009) 362–364.
[3] T. Cheng, D. Chen, B. Yu, H.B. Niu, Reconstruction of super-resolution STORM images using compressed sensing based on low-resolution raw images and interpolation, Biomed. Opt. Express 8 (5) (2017) 2445–2457.
[4] S. Kishk, B. Javidi, Improved resolution 3D object sensing and recognition using time multiplexed computational integral imaging, Opt. Exp. 11 (26) (2003) 3528–3541.
[5] D.T. Huang, W.Q. Huang, P.T. Gu, P.Z. Liu, Y.M. Luo, Image super-resolution reconstruction based on regularization technique and guided filter, Infrared Phys. Technol. 83 (2017) 103–113.
[6] S.W. Hell, Far-field optical nanoscopy, Science 316 (2007) 1153–1158.
[7] P. Gao, B. Prunsche, L. Zhou, K. Nienhaus, G.U. Nienhaus, Background suppression in fluorescence nanoscopy with stimulated emission double depletion, Nat. Photon. 11 (3) (2017) 163–169.
[8] E. Betzig, G.H. Patterson, R. Sougrat, O. Lindwasser, S. Olenych, J. Bonifacino, M. Davidson, J. Lippincott-Schwartz, H. Hess, Imaging intracellular fluorescent proteins at nanometer resolution, Science 313 (2006) 1642–1645.
[9] Y. Zhao, Q. Chen, X.B. Sui, G.H. Gu, A novel infrared image super-resolution method based on sparse representation, Infrared Phys. Technol. 71 (2015) 506–513.
[10] P. Milanfar, Super-resolution Imaging, CRC Press, 2010.
[11] A.V. Kanaev, C.W. Miller, Multi-frame super-resolution algorithm for complex motion patterns, Opt. Express 21 (17) (2013) 19850–19866.
[12] T. Peleg, M. Elad, A statistical prediction model based on sparse representations for single image super-resolution, IEEE Trans. Image Proc. 23 (6) (2014) 2569–2582.
[13] N.B. Hao, H.B. Liao, Y.M. Qiu, J. Yang, Face super-resolution reconstruction and recognition using non-local similarity dictionary learning based algorithm, IEEE/CAA J. Automatica Sin. 3 (2) (2016) 213–224.
[14] X.Y. jing, X.K. Zhu, F. Wu, R.M. Hu, X.G. You, Y.H. Wang, H. Feng, J.Y. Yang, Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning, IEEE Trans. Image Proc. 26 (3) (2017) 1363–1378.
[15] R. Horisaki, R. Takagi, J. Tanida, Learning-based imaging through scattering media, Opt. Exp. 24 (13) (2016) 13738–13743.
[16] V.M. Patel, Y.C. Chen, R. Chellappa, P.J. Phillips, Dictionaries for image and video-based face recognition, J. Opt. Soc. Am. A 31 (5) (2014) 1090–1103.
[17] M. Shi, Y. Bai, Q.M. Yi, SR Reconstruction algorithm based on adaptive multiple dictionary pairs, Front. Comput. Sci-Chi. 9 (5) (2015) 604–610.
[18] Y.X. Mao, Y. Wang, J.T. Zhou, H.W. Jia, An infrared image super-resolution reconstruction method based on compressive sensing, Infrared Phys. Technol. 76 (2016) 735–739.
[19] J.C. Yang, J. Wright, T. Huang, Y. Ma, Image SR via sparse representation, IEEE Trans. Image Proc. 19 (11) (2010) 2861–2873.
[20] M. Cheng, C. Wang, J. Li, Single-image super-resolution in RGB space via group sparse representation, IET Image Process. 9 (6) (2015) 461–467.
[21] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, Proceedings of the 7th International Conference on Curves and Surfaces(Curves and Surfaces 2012), pp. 711–730.
[22] K. Zhang, X. Gao, D. Tao, X. Li, Multi-scale dictionary for single image super-resolution, IEEE Computer Vision Pattern Recognition (IEEE CVPR 2012), pp. 1114–1121.
[23] W.S. Dong, L. Zhang, G.M. Shi, X.L. Wu, Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, IEEE Trans. Image Proc. 20 (7) (2011) 1838–1857.
[24] U. Maulik, S. Bandyopadhyay, Performance evaluation of some clustering algorithms and validity indices, IEEE Trans. Pattern Anal. 24 (12) (2002) 1650–1654.
[25] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Proc. 54 (11) (2006) 4311–4322.
[26] M. Okui, M. Kobayashi, J. Arai, F. Okano, Moire fringe reduction by optical filters in integral three-dimensional imaging on a color flat-panel display, Appl. Opt. 44 (21) (2005) 4475–4483.
[27] S. Baker, T. Kanade, Limits on super-resolution and how to break them, IEEE Trans. on Pattern Anal. 24 (9) (2002) 1167–1183.
[28] G. Kreymerman, Adjustable active optical low-pass filter, Appl. Opt. 51 (2) (2012) 268–272.

[29] L.M. Xu, W.H. Fan, J. Liu, High-resolution reconstruction for terahertz imaging, Appl. Opt. 53 (33) (2014) 7891–7897.
[30] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, IEEE Computer Vision and Pattern Recognition (IEEE CVPR 2017), pp: 105–114.
[31] I. Tosic, P. Frossard, Dictionary learning, IEEE Signal Process. Magazine 28 (2) (2011) 27–38.
[32] M. Mahajan, P. Nimbhorkar, K. Varadarajan, The planar k-means problem is NP-Hard, Lecture Notes Comput. Sci. 5431 (2009) 274–285.
[33] N.A.M. Isa, S.A. Salamah, U.K. Ngah, Adaptive fuzzy moving K-means clustering algorithm for image segmentation, IEEE Trans. on Consum. Electr. 55 (4) (2009) 2145–2153.
[34] T. Rosyadi, A. Arif, Nopriadi, B. Achmad, Faridah, Classification of leukocyte images using K-means clustering based on geometry features, on 6th International Annual Engineering Seminar (2016 InAES), pp. 245–249.
[35] X.F. Mai, J. Liu, X. Wu, Q. Zhang, C.J. Guo, Y.F. Yang, Z.H. Li, Stokes space modulation format classification based on non-iterative clustering algorithm for coherent optical receivers, Opt. Exp. 25 (3) (2017) 2038–2050.
[36] R. Gong, Y. Wang, Y.L. Cai, X.P. Shao, How to deal with color in super resolution reconstruction of images, Opt. Express 25 (10) (2017) 11144–11156.
[37] B. Yang, S.T. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuit, Information Fusion 13 (1) (2012) 10–19.
[38] L.H. Zhang, L. Dong, D.W. Zhang, X.M. Gao, X.H. Ma, Study of spectral reflectance reconstruction based on an algorithm for improved orthogonal matching pursuit, J. of the Opt. Soc. of Kor 20 (4) (2016) 515–523.