Image Inpainting and Image Fusion Using Sparse Representation

SEPTEMBER, 2019

DOCTOR OF ENGINEERING

QIAOQIAO LI

AKITA PREFECTURAL UNIVERSITY

Abstract

Image inpainting refers to the process of restoring the missing or damaged regions in an image. The technology of image inpainting has received considerable research interests in the last several decades because of its numerous applications in various fields. And image inpainting method based on sparse representation is a main branch of the technology.

One aspect of this thesis considers the image inpainting based on sparse representation, which is one of the technologies of image inpainting. In the image inpainting method based on sparse representation, the dictionary is very important which directly affects the inpainting results. In order to make the filled target region is visually consistent with the source region, we directly use the patches in the source region to construct the dictionary. Furthermore, if all patches in the source region are used to construct the dictionary, there will be lead to some unrelated patches for the target patch which may affect the inpainting result. Therefore, in the proposed image inpainting method based on sparse representation, we adapt the histogram similarity comparison method to compare the similarity between the target patch and the patches from the source region. And then use the similar patches to generate the related dictionary and histogram dictionary. Experiment results have shown that the proposed methods based on related dictionary and histogram dictionary achieve a better image inpainting result than the method based on original dictionary.

Another aspect of this thesis is the image fusion for multifocus image and medical image. Specially, multifocus images are obtained from the same camera with different focus point, while the medical images are obtained from different modalities, e.g., computer tomography (CT) and magnetic resonance imaging (MRI) images. For multifoucs image fusion, a multifocus image fusion method is proposed using a structure-preserving filter in spatial domain. In particular, the latest recursive filter (RF) is introduced as the

structure-preserving filter in the proposed spatial domain method. Moreover, a focused region detection method is presented to determine initial weight maps based on an average low-pass filter. Then a fused image can be generated by the final weight maps, which are obtained by using the RF to refine the initial weight maps and can well preserve the structures of source images. Experimental results show that the proposed method is superior to the state-of-the-art multifocus fusion methods in terms of subjective and objective evaluations.

For medical image fusion, a transform domain method using segment graph filter and sparse representation is proposed. Medical image fusion is to generate a single image from different multimodal images. As the human visual system (HVS) is sensitive to the edge information of source images, edge information should be integrated into the fused image as much as possible. In this thesis, by using the so-called segment graph filter, i.e., an edge-preserving filter based on the segment graph, the source images can be decomposed into base images and detail images. Then the base images are fused by a fusion rule based on normalized Shannon entropy while the detail images are fused by using a sparse representation-based fusion method. Finally, the resultant fused image is computed from combining the fused base image and fused detail image. Compared with other edge-preserving filters, a segment graph filter can well preserve the edge. Experiment results indicate that the proposed method can achieve the fusion performance of the state-of-the-art methods in both subjective visual performance and objective quantification.

Acknowledgements

First of all, I would like to express my profound gratitude to my advisor, Professor Guoyue Chen, for his extremely specific guidance and fatherly support and encouragement during my stay at Akita Prefectural University as a doctoral candidate. Furthermore, throughout my three-year stay in Japan, Professor Chen always takes care of all aspects of my life. Without the strong and constant help and inspiration from Professor Chen, the thesis and even my study life in Japan would have never been completed.

Besides my advisor, I would like to thank the rest members of the examination committee for this thesis, Professor Kohji Dohsaka (Akita Prefectural University), associate professor Kazuki Saruta (Akita Prefectural University), Professor Yanwei Chen (Ritsumeikan University), for their enthusiasm and willingness to read the thesis, raise the instructive questions and give insightful comments, despite their tight schedules.

And I sincerely appreciate the research associate Yuki Terata (Akita Prefectural University) for the actual help and suggestions for my research.

I also would like to express my appreciation to associate professor Kun Zhan (Lanzhou University, China) for his kind recommendation and constant encouragement. And also, I am greatly indebted to Akita Prefectural University and China Scholarship Council for financial support.

I am deeply thankful to Haibo Wang, who is now working as an engineer at Huawei Research Institute, Xian, Shaanxi , for his sincere help and constant care to me and also

for his valuable suggestions and insistent trust for my study in Japan.

Many thanks also go to the teachers of Japanese in Akita Prefectural University for their kind help of improving my Japanese.

I also gratefully acknowledge many Chinese and Japanese friends whose kindness, friendship, hospitality and assistance made my stay in Akita so enjoyable and unforgettable.

Last, but not least, I would like to give my deep gratitude to my husband, my parents, my husband's parents and my elder brother and his family for their many years of selfless love, unlimited forgiveness, and overwhelming support.

Contents

List of Acronyms

PDEs	partial differential equations
DCT	discrete cosine transforms
CVT	curvelet transform
k-SVD	k means singular value decomposition
OMP	orthogonal matching pursuit
NNOMP	non-negative orthogonal matching pursuit
PSNR	signal-to-noise ratio
RF	recursive filter
MFIF	multifocus image fusion
BLF	bilateral filter
CBF	cross bilateral fiter
GFF	guide filter

HVS	human visual system
SSIM	structural similarity metric
RF	row frequency
CF	column frequency
MDF	main diagonal frequency
SDF	side diagonal frequency
MFRF	multifouces image fusion based on recursive filter
NSCT	nonsubsampled contourlet transform
NSCTSR	nonsubsampled contourlet transform and sparse representation
MFGD	multifocus image fusion in gradient domain
FFIF	fast structure filter
AVE	average filter
GAU	gaussian filter
CT	computer tomography
PET	positron emission tomography

SPECT	single photon emission tomography
MRI	magnetic resonance imaging
PCA	principal component analysis
ICA	independent component analysis
LP	laplacian pyramid
WT	wavelet transform
SGF	segment graph filter
WLS	weighted least square
DCT	discrete cosine transform
MOD	method of optimal directions
MSGF	medical image fusion using segment graph filter
NFCD	neuron-fuzzy-based method in contourlet domain
GFIS	medical image fusion based guided image filter and image statistics

List of Figures

Chapter 1

Introduction

1.1 Background of Image Inpainting

The modification of images in a way that is non-detectable for an observer who does not know the original image is a practice as old as artistic creation itself. Medieval artwork started to be restored as early as the Renaissance, the motives being often as much to bring medieval pictures "up to date" as to fill in any gaps [78, 25]. This practice is called retouching or inpainting. Image inpainting is a data completion problem that aims to recover or fill in missing information in a degraded image. And Bertalmio et al. firstly use the image inpainting in image processing in [7]. Image inpainting has been widely investigated in the application of digital effect (e.g., object removal), image restoration (e.g., scratch or text removal in photograph), image coding and transmission (e.g., recovery of the missing blocks) and so on [89]. All algorithms are guided by the assumption that pixels in the known and unknown parts of the image share the same statistical properties or geometrical structures. This assumption translates into different local or global priors, with the goal of having an inpainted image as physically plausible and as visually pleasing as possible. The image inpainting methods can be divided into three categories [89, 31]. The first category is diffusion-based inpainting approach [7, 13], in which the smoothness priors via parametric models or partial differential equations (PDEs) to diffuse local structures

from the exterior to the interior of the holes. These methods are suited for completing straight lines, curves, and inpainting small regions. However, they are not well suited for recovering the texture of large missing region. The second category is the examplar-based inpainting algorithm [22, 85], in which the best match patch or a set of candidate patches is chosen from the undamaged region by similarity comparing methods. Compared with the diffusion-based inpainting algorithm, the examplar-based inpainting algorithms have performed plausible results for inpainting the large missing region. However, this kind of methods always select the most suitable patch for the current place, a greedy method, which results in a risk of introducing unwanted object or artifact to the area to inpaint [70]. The third category is sparse representation-based inpainting methods [24, 70, 26]. The basic idea of this kind of approach is to represent an image by the sparse combination of an overcomplete set of transforms (e.g., wavelet, contourlet, discrete cosine transforms (DCT)), then the missing pixels are inferred by adaptively updating this sparse representation [89, 31]. In sparse representation, the number of patches and their coefficients are adaptively determined, instead of using only the best matching patch or a fixed number of the best matching patches. This overcomes the artifacts caused by the greedy search strategy used in exemplar-based algorithms. Therefore, the inpainting approach based on sparse representation will be particularly concerned in this thesis.

In order to compute the sparse representation of a signal, a dictionary is firstly constructed. There are three kinds of dictionaries are proposed for sparse representation. The first type is fixed dictionary which is often created by a prespecified set of function, such as discrete cosine transforms (DCT), short-time Fourier transforms, wavelet, Curvelet transform (CVT), and contourlet. However, these dictionaries are not well equipped for representation more complex natural and high-dimensional signal data [92, 69]. The second type is the learned dictionary, such as k-means singular value decomposition (K-SVD) [1]. Because a dictionary is learned with a set of training images or updated adaptive-

ly by using SVD operations, this method is termed as K-SVD which is the most widely used adaptive dictionary construction method. Specially, the training images can come from a set of natural images (globally trained dictionary) or input image itself (adaptively trained dictionary). And Elad and Aharon have indicated that an adaptive dictionary learned from a noisy input image sometimes outperforms and the fixed DCT or global trained dictionary. However, the process of K-SVD is iterative learning with a large number of training images which is more costly to be applied to practical application, thus the dimension of the dictionary is constrained due to its high computational complexity [39]. The third type is to use all the patches in the source region to construct the dictionary [70]. This type of dictionary can keep visual consistent between the target region and source region very well which makes the entire image looks plausible. However, this type dictionary contains all the known image patches in the image, there will be a large number of unrelated image patches to be restored [87]. In order to solve this problem, an aspect of this thesis is to find similar patches by using the similarity comparison method before generating the dictionary.

Therefore, the first objective of this research is to propose a new image inpainting based on related dictionary constructed by histogram. The related dictionary is constructed from the similarity comparison method using the histogram. In detail, the similarity comparison method using the histogram is divided into the four steps: Step 1, compute the histograms of the target patch and the candidate patches; Step 2, compare the histogram differences of three-channel histograms between the target patch and candidate patches; Step 3, sum the histogram differences and then sort them; At last, the similar patches are chosen from the order of arrangement (i.e., the more in front of the sequence, the higher the similarity between the target patch and candidate patch). Thus, the related dictionary which is constructed by these similar patches has a close relationship with the target patch. Because this similarity comparison method is sort by the sum difference value of R, G, B

three-channel histograms, there is a case the sum difference values of different patches are same which will make mistakes when we choose samples as the dictionary.

In order to solve this problem, a histogram dictionary is proposed for image inpainting. The histogram dictionary is also constructed by the similar comparison method using the histogram. The biggest difference from the related dictionary is in the third step of the similarity comparison method using the histogram. In step 3, we chose the max differences of the R, G, B three-channel histograms. And the other steps are the same as the similarity comparison method for the related dictionary. Therefore, the second objective of this thesis is to propose a new image inpainting based on sparse representation with histogram dictionary.

1.2 Background of Image Fusion

Images are taken by cameras usually suffer from the limited depth of field of a lens, so they can hardly get a picture in which all the objects are explicit, especially when the objects differ largely with each other in terms of their distances from the camera [45, 14, 108]. This effect may lead to a disadvantage for human observation or further computer processing. In order to solve this problem, the technology of multifocus image fusion has been proposed [44]. In this technology, it manages to detect the focused regions in multifocus images of the same scene and then integrate them to generate a composite image in which all the objects of interest are in focus. Until now, many image fusion algorithms have been invented [28, 32, 79, 98, 102]. According to the domains in which visual information is combined, these algorithms can be roughly grounded into two catalogs: transformed domain-based methods and spatial domain-based methods. Transformed domain-based methods process the transform coefficients while spatial domain-based method process the pixels directly. The process of the transformed domain-based methods is as follows: First, obtain the transform coefficients of input images by using the scale transformation;

Second, acquire a fused transformed coefficient according to a specific fusion rule to fuse the scale transformed coefficients of different images. Third, the fused image is got by taking an inverse scale transform on the fused coefficients. It should be noted that this kind of algorithm often requires transforming the source images to different frequency coefficients, i.e., high-frequency and low-frequency coefficients [53]. However, using the high-pass filters usually leads to the ringing effects and "halo" around the major structures [102, 108].

While the spatial domain-based image fusion methods directly combines the original information in the source images, which aim at synthesizing a fused image that is more information for visual perception and computer processing [43]. These algorithms use a fusion rule on the source images to obtain a fused image with all-in-focus. Most of these methods are based on the block or region. And the process of block-based multifocus image fusion methods are divided into the following steps: Firstly, divide the source image into blocks; Secondly, compute some focus measurements for each block. In this step, a binary map representing the blur and the focus regions can be obtained; Thirdly, based on this map, the focused regions are detected and integrated to obtain the fused image. Considering the spatial domain-based method process image pixels instead of the multiscale transform coefficients, and it can well preserve the original intensities of source images in the fusion result. Therefore, the third objective of this research is to propose a new multifocus image fusion in spatial domain. Furthermore, a structure-preserving filter is introduced into the proposed method, which aims to prevent smoothing across structures while still smoothing texture. And then original intensities of source images can well preserve in the fusion result [102].

However, single sensor image capture is insufficient to provide complete information about a targeted scene because of the sensor system limitations. Therefore, an image can be captured from different sensors which enhances visibility to human eyes or to mutually complement limitations of each image. And the entire information from several captures

should be integrated into a single image. Therefore, the multimodal image fusion technology has been developed to merge different multimodal images into a single image. Specially, multinodal medical image fusion has been emerging as a promising research field [112]. In medical imaging, different modalities such as positron emission tomography (PET), single-photon emission tomography (SPECT), computer tomography (CT), and magnetic resonance imaging (MRI) are used to capture complementary information [5]. CT imaging can exactly detect dense structure such as bones and implants, while MRI provides high-resolution soft substance information such as soft tissues. However, physicians often need to sequentially analyze different modalities medical images for better diagnosis and treatment.

Medical image fusion technique can be used to integrate the complementary information contained in multiple medical images to obtain a composite image [96]. The composite image not only can help the radiologist for better diagnosis and treatment, but also can help for the computer assisted surgery and radio surgery [5]. Recently, many medical image fusion methods have been proposed over the past decades [84, 2, 93, 18, 73, 50, 99, 100, 49, 102]. Because the human visual system (HVS) processes information in a multiresolution fashion which has been evidenced in [68], most medical image fusion methods are introduced under a multiscale transform-based framework to pursue perceptually good results. For example, Yang et.al has successfully used the contourlet transform into the medical image fusion because of its effectiveness in representing spatial structures [95]. However, contourlet has no shift-invariant property, as it contains downsampling process in the transform process. Recently, structure-preserving filter also has been applied to achieve multi-scale decomposition of images [27, 36]. Considering the ability of the to structure-preserving filter accurately separate an image into different scale structures which can help to reduce halo and aliasing artifacts in the fusion process, we will introduce it in the medical image fusion in this thesis to make the fusion results good

for human visual perception. Therefore, the forth objective of this research is to propose a new medical image fusion using segment graph filter and sparse representation.

1.3 Purpose of This Research

This study has aimed on developing the methods for image inpainting and image fusion and to obtain the good inpainted image and fused image for image processing.

In Chapter 1, the backgrounds of image inpainting and image fusion, and research purpose are particularly described.

In Chapter 2, the preliminaries is presented. The sparse representation, the structure-preserving filter and objective assessment metrics are given in this chapter.

In Chapter 3, an image inpainting method based on sparse representation is proposed. For sparse representation, the dictionary is very important. A related dictionary is obtained by sum histogram similarity comparison method. The sum histogram similarity comparison method is used to find out the patches which are similar to the target patch. And then the similar patches are applied to generate the related dictionary. Because the related dictionary is generated from the similar patches, a large number of unrelated patches will be avoided to effect the inpainting result. Although the inpainting method based on related dictionary can obtain good inpainting result, there also exist a special case. That is to say, when the sum differences of histogram for patches are same, it can not sort them and then can not choose a similar patch. Therefore, we also propose a histogram dictionary which is generated from the similar comparison by max difference of histogram. And then a new image inpainting method based on histogram dictionary is also proposed in this chapter.

In Chapter 4, an image fusion based on structure-preserving filter is proposed. The proposed fusion method is belong to the spatial domain method as the spatial domain method directly process pixels rather than processing the transform domain coefficients. And be-

cause the filtering is the most fundamental process of image fusion, and it is very important to exploit structures and details within images for fusion, the structure-preserving filter is used to prevent smoothing across structures while still smoothing texture in the proposed method. The propose fusion method is compared to the state-of-the-art methods, and the experimental results have shown that the performance of the proposed fusion scheme is better than the state-of-the-art methods in terms of subjective and objective evaluations. Also, a medical image fusion using segment graph filter and sparse representation is proposed. In medical imaging, different modalities such as computer tomography (CT) and magnetic resonance imaging (MRI) are used to capture complementary information. However, a doctor needs both CT and MRI information in a single image for better diagnosis and treatment. Therefore, a medical image fusion using segment graph filter and sparse representation to obtain a fusion medical image which contains the information with different modalities.

Finally, in Chapter 5, give the conclusions of this thesis.

Chapter 2

Preliminaries

Some basic preliminaries will be described in this chapter to increase this thesis more readable.

2.1 Sparse Representation

Sparse representation is based on the hypothesis that natural signal can be represented by a linear combination of a "few" atoms in dictionary matrix [53, 61]. That is to say given a dictionary $D \in \mathbb{R}^{j \times k}$ $(j < k)$ where each column of the dictionary $D = [d^1, d^2, ..., d^k]$ named an atom, and then a target signal $y = [y_1, y_2, ..., y_j]^T$ can be represented as a linear combination of atoms:

$$y \cong D\alpha, \tag{2.1}$$

where α are coefficients which represent the signal y in terms of the dictionary D.

In practice, the sparse representation problem of y over D is usually cast as

$$min_\alpha \|\alpha\|_0 \quad s.t. \quad y = D\alpha, \tag{2.2}$$

or

$$min_\alpha \|\alpha\|_0 \quad s.t. \|y - D\alpha\|_2 < \varepsilon, \tag{2.3}$$

where $\|.\|_0$ is l_0 norm, i.e., count the number of nonzero entries of a vector, and ε is the error tolerance. The optimization of it is an NP-hard problem and the greedy algorithms such as matching pursuit (MP) [58], orthogonal matching pursuit (OMP) [76] and other improved OMP [77] algorithms are always applied for solving this problem to estimate the coefficients α.

Because sparse representation and compressed sensing have been developed, the non-convex l_0-minimization problems in Eq.(2.2) and Eq.(2.3) can be relaxed to obtain the convex l_0-minimization problems in [86, 12]

$$min_\alpha \|\alpha\|_1 \quad s.t. \quad y = D\alpha, \tag{2.4}$$

and

$$min_\alpha \|\alpha\|_1 \quad s.t. \|y - D\alpha\|_2 < \varepsilon. \tag{2.5}$$

Solutions of this problem can be obtained by using linear programming methods [86, 21].

2.1.1 Design a Proper Dictionary

How to construct a proper dictionary is of particular importance problem in sparse representation. Recently, many dictionary generated methods have raised and these methods can be cast into three categories [105, 55]. One category is fixed basis, such as DCT dictionary [92]. The advantage of this dictionary is simple and easy to implement. However, a major problem with this category of dictionary is that it is often restricted to signals of a certain type and cannot be used for an arbitrary family of signals. The second category is based on some learning methods, such as PCA, MOD and K-SVD [1]. These learning-based methods mainly learn from two samples: one sample is a set of images and the other sample is source images. Comparing the two learning samples, learning a

dictionary directly from source images leads to better representation and provides superior performance in many image and vision application [64]. Since, the K-SVD is a popular learning dictionary, we introduce it as an example. The K-SVD algorithm is an iterative alternation between two steps: sparse coding (in order to find α) and dictionary updating (in order to find D). In the sparse sparse coding step, D is assumed to fixed, and the above optimization problem as a search for sparse representation with coefficients summarized in the matrix α. And this optimization problem can be solved by the MP [58] and OMP [76] algorithms mentioned in section 2.1. In the dictionary updating stage, the coefficient matrix α and the dictionary D are both presumed to be fixed. And only one column in the dictionary d_k and the coefficients which are correspond to it, the kth row in α, denoted as α_T^k. Then use the singular value decomposition (SVD) to find alternative d_k and α_T^k. However, this learned dictionary always has high computational complexity [39]. The third category is to use all the patches in the source region to construct the dictionary [70]. This category of dictionary can keep visual consistent between the target region and source region very well which makes the entire image looks plausible.

2.2 Structure-preserving Filter

2.2.1 Recursive Filter

The one-dimensional discrete recursive filtering process can be described by the following difference equation [71]:

$$J[n] = \sum_{i=0}^{M} a_i I[n-i] - \sum_{j=1}^{N} b_j J[n-j], \tag{2.6}$$

where $I[n]$ is the input signal, $J[n]$ is the filter output, a_i and b_j are the filter coefficients.

If $M = 0$ and $N = 1$, the filter of Eq. (2.6) becomes a first-order (1st-order) recursive

filter, i.e., the simplest one. Accordingly, the 1st-order recursive filter can be expressed by

$$J[n] = a_0 I[n] - b_1 J[n-1]. \tag{2.7}$$

As shown in the literature [29, 101], let $a_0 = 1 - a$ and $b_1 = -a$. Then, Eq. (2.7) is rewritten as

$$J[n] = (1-a)I[n] + aJ[n-1], \tag{2.8}$$

which is depicted in Fig. 2.1, where z^{-1} denotes a one-sample delay in time, $a \in [0, 1]$ and

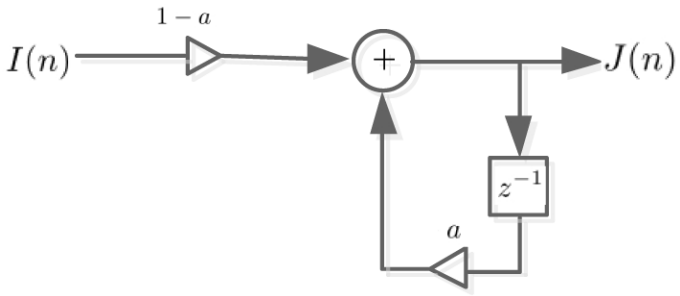


Figure 2.1: System diagram for the filter

$(1-a)$ are called the feedback and feedforward coefficients [71], respectively. Specifically, $J[n]$ is the output sample at time n which is based on the present input sample $I[n]$ and past output sample $J[n-1]$.

The transfer function of Eq. (2.8) is $h(n) = (1-a)a^n$ [71]. Given an impulse signal of $\delta(m-n)$, a response of $(1-a)a^{m-n}$ is generated. Here, $m-n$ is the distance between samples x_m and x_n [71, 29].

According to the above observation, Gastal *et al.* [29] propose a recursive edge-preserving filter which is defined by

$$J[n] = (1-a^d)I[n] + a^d J[n-1], \tag{2.9}$$

where d is the distance between neighborhood pixels. As d increases, a^d goes to zero,

stopping the propagation chain. Thus, it can preserve edges.

2.2.2 Segment Graph Fliter

The well known structure-preserving smoothing techniques can be divided into two types [103]: One is the optimization based filters, such as an edge-preserving filtering method based on weighted least square (WLS) optimization in [27] and L_0-smoothing filter [88]. The other is the weighted average based, such as bilateral filter (BF) [27] and guided filter (GF) [33]. Although these edge-preserving smoothing techniques are widely used in many methods, they may lead to the "halo" artifacts and large time consumption in the first type method, and cause the "halo" artifacts in the second type.

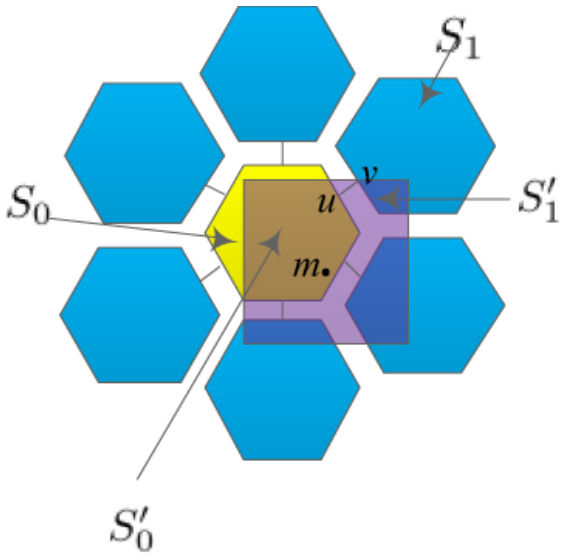


Figure 2.2: Filter kernel of structure preserving filter (SGF). The superpixels are presented as hexagons. The pixel m is in the superpixel S_0 (yellow hexagon). And its filtering window W_m is shown with violet mask.

Zhang *et al.* [103] has introduced tree distance into their filter to solve the "halo" problem appeared in most of these filters. And in order to solve "leak" problem, they design a segment graph technique which is a more reliable edge-aware structure to represent the image. According to these, a novel linear local filter based on segment graph, named as segment graph filter (SGF) is proposed [103]. Because superpixel decomposition of a

given image has been studied and the superpixel can run very fast in linear, they use the superpixel decomposition to construct the segment graph. A detailed introduction of the segment graph can be seen from Zhang *et al* .'s literature [103].

The SGF is based on the double weighted average, i.e., internal weight and external weight. Taking the tree distance into consideration, the internal weight function w_1 can be defined by

$$w_1(m,n) = exp(-\frac{D(m,n)}{\sigma}), \tag{2.10}$$

where $D(m,n)$ represents the tree distance between pixels m and n. As σ controls the attenuation speed of $D(m,n)$, the w_1 is inversely proportional to the tree distance $D(m,n)$.

For describe external weight a smoothing window W_m with radius r and the superpixels technology are introduced. The filter kernel of SGF is shown in Fig. 2.2. As shown in Fig. 2.2, several superpixel regions are denoted as $\{S_0,\ S_1,\ ...,\ S_k\}$ and the overlapped regions are represented by $\{S'_0,\ S'_1,\ ...,\ S'_k\}$, namely $S'_i = W_p \cap S_i$. Therefore, the external weight function w_2 can be defined by the area size ratio of S'_i and S_i

$$w_2(m,S_i) = \frac{|S'_i|}{|S_i|}, \tag{2.11}$$

where $|S'_i|$ and $|S_i|$ denote the area size of S'_i and S_i, respectively.

Once the internal weight and external weight are obtained, the filter output of an input image I at pixel n can be given by

$$J_m = \frac{1}{K_m} \sum_{0 \le i < k} w_2(m,S_i) \sum_{n \in S_i} w_1(m,n) I_n, \tag{2.12}$$

where K_m, S_i and J_m represent a normalizing term, superpixel region and filter output, respectively. w_1 and w_2 are the internal weight function and external weight function, respectively. The output J_m at pixel m is the double weighted average of the intensity value I_n in a specific neighbor region $\Omega = \cup_{0 \le i < k} S_i$ ($n \in S_0$).

In the linear implementation of SGF, a threshold τ is always set to cut off some of the edges between S_0 and its neighborhood S_i for taking full advantage of segment graph. In Fig. 2.2, the connecting edge E_{min} between S_0 and its neighborhood S_i during the segment graph construction can be defined by

$$E_{min}(S_0, S_i) = min\{W(u, v)|u \in S_0, v \in S_i\}, \tag{2.13}$$

$$W(u, v) = |I_u - I_v|, \tag{2.14}$$

where $u \in S_0$, $v \in S_i$ are the pixels/vertexes of the connecting edge.

Considering the above description, the filter output of the segment graph filter can be rewritten by

$$\begin{aligned} J_m &= \frac{1}{K_m} \sum_{0 \le i < k} \delta_i w_2(m, S_i) \sum_{n \in S_i} w_1(m, n) I_n \\ s.t. \quad \delta_i &= \begin{cases} 0, & \text{if} E_{min}(S_0, S_i) > \tau \\ 1, & \text{otherwise.} \end{cases} \end{aligned} \tag{2.15}$$

2.3 Objective Assessment Metrics

Objective assessment metrics are usually used to evaluate the process result for image processing. In this paper, we use the peak-signal-to-noise-ratio (PSNR) [81] to evaluate the inpainting result, and the fusion performances are evaluated by the six metrics, i.e., feature-based metric $Q_p^{ab|f}$ [90], structure-based metric $Q_w^{xy|f}$ [94], the normalized mutual information Q_{MI} [35], a nonlinear correlation information entropy Q_{NICE} [56], Chen-Blum metric Q_{CB} [56, 15], and spatial frequency error Q_{SF} based on spatial frequency [56, 111]. The PSNR is belong to the evaluation which requires a reference image and the large the PSNR value are, the better the inpainting result. Because it is often very difficult to obtain a ground truth, the aboved six metrics are the objective evaluation metrics without

requiring a reference image. And the larger the $Q_p^{ab|f}$, $Q_w^{xy|f}$, Q_{MI}, Q_{NICE} and Q_{CB} values are, the better the fusion results are, while the smaller the Q_{SF} value is, the better the fusion result is.

2.3.1 Peak-Signal-to-Noise-Ratio ($PSNR$)

Given a reference image r and a test image t, and the size of them are all $M \times N$, the PSNR between r and t

$$PSNR(r,t) = 10\log_{10}(255^2/MSE(r,t)), \tag{2.16}$$

where

$$MSE(r,t) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(r_{ij} - t_{ij})^2. \tag{2.17}$$

As the value of MSE approaches zero the PSNR approaches infinity, which show that a higher PSNR value provides a higher image quality [34].

2.3.2 Edge Evaluate Metric ($Q_p^{ab|f}$)

Because the human visual system (HVS) is sensitive to edge information, it is necessary to keep this information as much as possible. Xydeas and Petrovic [90] propose a metric $Q_p^{ab|f}$ to evaluate the amount of edge information which is preserved in the fused image.

The edge information preservation value Q^{AF} between source image A and the fused image F can be defined as

$$Q^{AF}(i,j) = Q_g^{AF}(i,j)Q_\alpha^{AF}(i,j), \tag{2.18}$$

where Q_g^{AF} and Q_α^{AF} are the edge strength and orientation preservation values, respectively.

Similarly, the edge information metric transferred from source image B into the fused

image F can also be computed. And then a normalized weighted performance metric $Q_p^{ab|f}$ between source images A, B and fused image F is computed by

$$Q_p^{ab|f}(i,j) = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M}[Q^{AF}(i,j)w^A(i,j) + Q^{BF}(i,j)w^B(i,j)]}{\sum_{n=1}^{N}\sum_{m=1}^{M}(w^A(i,j) + w^B(i,j))}, \tag{2.19}$$

where $w^A(i,j)$ and $w^B(i,j)$ are the weighted coefficients defined by

$$w^A(i,j) = [g_A(i,j)]^L, \tag{2.20}$$

$$w^B(i,j) = [g_B(i,j)]^L, \tag{2.21}$$

with a constant value L.

2.3.3 Image Structure Similarity Metric ($Q_w^{xy|f}$)

The structural similarity metric (SSIM) [82] for the source image X and fused image F with a sliding window w can be defined by

$$SSIM(X,F|w) = \frac{(2\bar{w_X}\bar{w_F} + C_1)(2\sigma_{w_XF} + C_2)}{(\bar{w_X^2} + \bar{w_F^2} + C_1)(\sigma_{w_X}^2 + \sigma_{w_F}^2 + C_2)}, \tag{2.22}$$

where C_1 and C_2 are small constants, w_X denotes the sliding window under consideration in X, $\bar{w_X}$ is the mean of w_X, $\sigma_{w_X}^2$ and $\sigma_{w_Xw_F}$ are the variance of w_X and covariance of w_X and w_F, respectively. The structural similarity metric for the source image Y and fused image F can also be computed.

Yang *et al.* [94] propose a new structural similarity with a threshold

$$Q_w^{xy|f} = \begin{cases} \lambda_w SSIM(X,F|w) + (1-\lambda_w)SSIM(Y,F|w), \\ \qquad \text{if} \quad \text{SSIM(X, Y|w)} \geq 0.75, \\ max\{SSIM(X,F|w), SSIM(Y,F|w)\}, \\ \qquad \text{if} \quad \text{SSIM(X, Y|w)} < 0.75, \end{cases} \tag{2.23}$$

where the weight λ_w is defined by

$$\lambda_w = \frac{s(X|w)}{s(X|w) + s(Y|w)}. \tag{2.24}$$

In an implementation, $s(X|w)$ and $s(Y|w)$ are the variances of images X and Y with the window w, respectively.

2.3.4 Normalized Mutual Information (Q_{MI})

Mutual information [35] is used to quantify the overall mutual information between the source images and fused image. For the source image A and fused image F, the mutual information is given by

$$MI(A,F) = H(A) + H(F) - H(A,F), \tag{2.25}$$

where

$$H(A) = -\sum_a p(a)\log_2 p(a), \tag{2.26}$$

$$H(F) = -\sum_f p(f)\log_2 p(f), \tag{2.27}$$

$$H(A,F) = -\sum_{a,f} p(a,f)\log_2 p(a,f), \tag{2.28}$$

where $p(a)$ and $p(f)$ are the marginal probability distribution functions of A and F, respectively and $p(a,f)$ is the joint probability distribution function of A and F.

Based on Eqs. (2.26) - (2.28), the mutual information can be rewritten as

$$MI(A,F) = \sum_A \sum_F p(a,f)\log_2 \frac{p(a,f)}{p(a)p(f)}. \tag{2.29}$$

Similarly, $MI(B,F)$ is the mutual information between the source image B and fused

image F which can be defined by

$$MI(B,F) = \sum_{B}\sum_{F} p(b,f)\log_2 \frac{p(b,f)}{p(b)p(f)}. \tag{2.30}$$

Finally, the normalized mutual information performance metric Q_{MI} is defined by

$$Q_{MI} = 2\left[\frac{MI(A,F)}{H(A)H(F)} + \frac{MI(B,F)}{H(B)H(F)}\right]. \tag{2.31}$$

2.3.5 Nonlinear Correlation Information Entropy (Q_{NICE})

Nonlinear correlation information entropy (Q_{NICE}) is a quality metric based on information theory [80]. A nonlinear correlation matrix R of the source images A, B, and fused image F is defined as

$$R = \begin{pmatrix} NCC_{AA} & NCC_{AB} & NCC_{AF} \\ NCC_{BA} & NCC_{BB} & NCC_{BF} \\ NCC_{FA} & NCC_{FB} & NCC_{FF} \end{pmatrix} = \begin{pmatrix} 1 & NCC_{AB} & NCC_{AF} \\ NCC_{BA} & 1 & NCC_{BF} \\ NCC_{FA} & NCC_{FB} & 1 \end{pmatrix}, \tag{2.32}$$

where $NCC_{X,Y}$ represents the nonlinear correlation coefficient between the source images A, B and the fused image F [80].

Let the eigenvalues of the nonlinear correlation matrix R be λ_i (i = 1, 2, 3). Then, the Q_{NICE} is computed as

$$Q_{NICE} = 1 + \sum_{i=1}^{3} \frac{\lambda_i}{3}\log_{256}\frac{\lambda_i}{3}. \tag{2.33}$$

2.3.6 Chen-Blum Metric (Q_{CB})

According to the human visual system (HVS), a quality assessment, named Q_{CB} is defined by Chen and Blum [15]. Let the masked contrast map for source image A be

$$C_A^{'} = \frac{t(C_A)^p}{h(C_A)^q + Z}, \tag{2.34}$$

where t, h, p, q, and Z are real scalar parameters that determine the shape of the nonlinearity of the masking function [15]. The information preservation value $Q_{AF}(x, y)$ from source image A to fused image F is defined as

$$Q_{AF}(x,y) = \begin{cases} \frac{C_A^{'}(x,y)}{C_F^{'}(x,y)}, \text{if } C_A^{'} < C_F^{'} \\ \frac{C_F^{'}(x,y)}{C_A^{'}(x,y)}, \text{otherwise.} \end{cases} \tag{2.35}$$

The global quality map between source images A, B and F is computed as

$$Q_C(x,y) = \lambda_A(x,y)Q_{AF}(x,y) + \lambda_B(x,y)Q_{BF}(x,y), \tag{2.36}$$

where $\lambda_A(x, y)$ and $\lambda_B(x, y)$ are the saliency maps of source images A and B, respectively. Finally, the metric value Q_{CB} is obtained by averaging the global quality

$$Q_{CB}(x,y) = \overline{Q_C}(x,y). \tag{2.37}$$

2.3.7 Image Fusion Metric Based on Spatial Frequency (Q_{SF})

Zhang *et al.* [111] propose an overall spatial frequency of an image $A(i, j)$ which is defined as

$$SF = \sqrt{(RF)^2 + (CF)^2 + (MDF)^2 + (SDF)^2}, \tag{2.38}$$

where RF, CF, MDF and SDF are the four first-order gradients along four directions

$$RF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}[A(i,j) - A(i,j-1)]^2}, \tag{2.39}$$

$$CF = \sqrt{\frac{1}{MN}\sum_{j=1}^{N}\sum_{i=2}^{M}[A(i,j) - A(i-1,j)]^2}, \tag{2.40}$$

$$MDF = \sqrt{w_d\frac{1}{MN}\sum_{i=2}^{M}\sum_{j=2}^{N}[A(i,j) - A(i-1,j-1)]^2}, \tag{2.41}$$

$$SDF = \sqrt{w_d\frac{1}{MN}\sum_{j=1}^{N-1}\sum_{i=2}^{M}[A(i,j) - A(i-1,j+1)]^2}, \tag{2.42}$$

where the distance weight w_d is set as $1/\sqrt{2}$. The four reference gradients can be obtained by taking the maximum of absolute gradient values between source images A and B along directions

$$Grad^D(I_R(i,j)) = \max\{abs[Grad^D(A(i,j))], abs[Grad^D(B(i,j))]\}, \tag{2.43}$$

where $D = \{H, V, MD, SD\}$ denote as horizontal, vertical, main diagonal and secondary diagonal, respectively. The four directional references RF_R, CF_R, MDF_R and SDF_R can be computed with the reference gradients substituting the differences in Eqs. (2.39) - (2.42). Thus, SF_R can be computed from Eq. (2.38).

Finally, the ratio of the SF error (metric Q_{SF}) is defined as

$$Q_{SF} = (SF_F - SF_R)/SF_R. \tag{2.44}$$

Chapter 3

Image Inpainting based on Sparse Representation

3.1 Image inpainting based on related dictionary

3.1.1 Introduction

Image inpainting which is to fill the missing or corrupted areas by using the known information of the image. These areas may be individual missing pixels in the damaged image or be continuous regions resulting from man-made degradation and other reasons. In recently, image inpainting has attracted growing interest from researchers because it has a wide variety of applications, such as image object removal [31], image restoration [37], transmission [26], noise removal [107] and disocclusion [59].

Formally, the problem of inpainting can be defined as following: given an input image I with a target region Ω which is filled in all pixels by using the information of the known areas $\Psi = I - \Omega$ [9].

Bertalmio *et al.* [7] firstly studied the image inpainting and then proposed an inpainting approach based on linear partial differential equations. Since then, many inpainting methods were proposed to address this problem. Generally, there are three types

of image inpainting approaches, i.e., diffusion-based inpainting approach, exemplar-based inpainting approach, and inpainting approach based on sparse representation [26]. The diffusion-based inpainting approach is such that the corrupted area is restored by diffusing the surrounding information to the target region [13, 7]. However, these diffusion-based inpainting methods are not well suitable for the textured, especially if the target region to be restored is larger than other regions. The exemplar-based inpainting approach, which is inspired by the idea of texture synthesis technique [7], is that a patch in the corrupted area is first selected, and then the unknown pixels in the selected patch is filled by copying the pixels in the best matching patch in the whole source region by comparing the similarity of selected patch [31, 22, 48, 17, 16, 89, 20]. Instead of using the best patch, inpainting approach based on sparse representation is to represent the image patches by using a sparse linear combination of atoms from a dictionary. Thus, the dictionary is very important for inpainting results. An effective sparse representation iterative inpainting algorithm is proposed [26]. And it is very suitable for recovering different structural components in the image. And the sparse representation can be adapted for restoring the color image [57]. This method is mainly extended from the denoising algorithm [23] whose dictionary is constructed by using the K-SVD [1]. After that, a similar algorithm is proposed in [70]. They directly used all the patches which are clipped from in the source region to construct a dictionary and obtained good inpainting results. They also analyzed their method which was better than the algorithm [57]. Although the dictionary generated from all the patches in source region is advantageous to the image inpainting, it will have some the unrelated atoms with image patch to be restored. Moreover, the unrelated atoms will introduce the interference into inpainting results. As a result, it will affect the inpainting result.

Therefore, in order to solve this problem, a new similarity comparison algorithm is proposed to find the similar patches before directly using the patches to generate the dictionary in this section. And then, use these similar patches to generate a related

dictionary for the target patch. At last, an image inpainting method is proposed based on related dictionary in this chapter. This proposed method makes every target patch has it corresponding related dictionary that can guarantee the inpainting result.

3.1.2 Algorithm

The overview of the proposed method will be presented in this section in detail. The proposed method can be divided into the following three parts: find a target patch, generation of a dictionary and sparse reconstruction algorithm. In the first part, the target patch is found by computing the filling order. And the whole algorithm is starting from this patch. In the second part, we use the similarity comparison method of histogram to compare the similarity between the target patch and the candidate patches and then to find the similar patches for the target patch. Then, a related dictionary is obtained by using the similar patches. In the last part, we use the known information from the target patch to estimate their unknown information according to the sparse representation. Details of the proposed method are shown in the following.

Filling Order

The filling order of image patches, which decides a patch on the missing region boundary with the highest priority for further inpainting, is crucial to inpainting results. In the inpainting methods, a different filling order can lead to different results [16]. In our method, we decide to use the filling order which is proposed in [17], since the structure information can be efficiently preserved. Recently, the filling order [17] has been used for their inpainting algorithms and obtained good results [70, 87]. Using the filling order to compute every patch which is centered at p located the boundary δU between damaged region (target region T) and undamaged region (source region S). Finding the target patch ψ_p has the maximum priority. As the patch ψ_p is on the boundary, the patch contains the known pixels A and the unknown pixels B. And we can see clearly the target patch in

Fig.3.1.

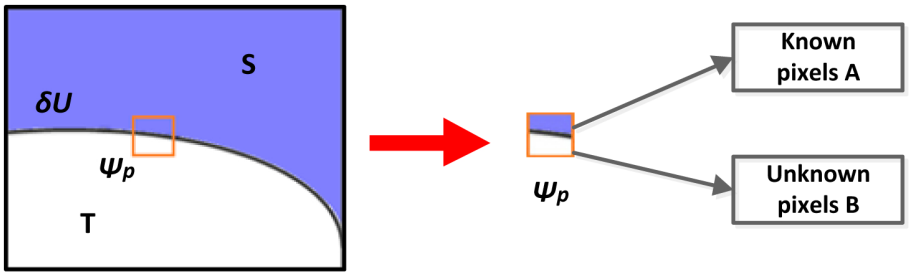


Figure 3.1: Using the filling order to select the target patch with higher priority

The filling order proposed by Criminisi *et al.* [17] is an iteration algorithm with the following three steps until all pixels have been filled:

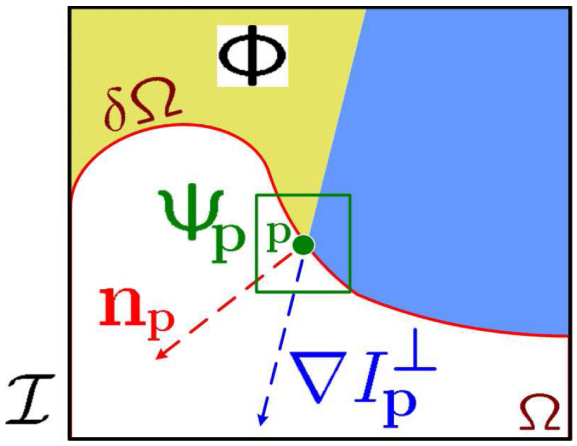


Figure 3.2: Notation diagram. Given the patch Ψ_p, n_p is the normal to the contour $\delta\Omega$ of the target region Ω and $\bigtriangledown I_p^{\perp}$ is the isophote (direction and intensity) at the point p. The entire image is denoted with I.

Step 1: Compute patch priorities. In the filling order algorithm, each pixel which is unfilled maintains a colour value and a confidence value, which reflects the confidence in the pixel value. Once a pixel is filled, the confidence will be frozen. That is to say, patches along the fill front are also given a temporary priority value during the process of the algorithm, which determines the order in which they are filled. Given a patch Ψ_p centred at the point p for some $P \in \delta\Omega$ which is shown in 3.2, the priority $P(p)$ of it is defined as

the product of two terms:

$$P(p) = C(p)D(p), \tag{3.1}$$

where $C(p)$ and $D(p)$ are the confidence term and the data term, and they are computed by

$$C(p) = \frac{\sum_{q \in \Psi_p \overline{\cap}} C(q)}{|\Psi_p|}, \tag{3.2}$$

$$D(p) = \frac{| \bigtriangledown I_p^{\perp} \cdot n_p|}{\lambda}, \tag{3.3}$$

where $|\Psi_p|$ is the area of Ψ_p, λ is a normalization factor (e.g., $\lambda = 255$ for a typical grey-level image), and n_p is an unit vector orthogonal to the front $\delta\Omega$ in the point p. The function $C(p)$ is set to $C(p) = 0 \ \forall \ p \in \Omega$, and $C(p) = 1 \ \forall \ p \in I - \Omega$.

Step 2: Propagate texture and structure information. Once all priorities on the fill front have been computed, the patch $\Psi_{\hat{p}}$ with the highest priority is found. And then fill it with the inpainting algorithm from the source region Φ.

Step 3: Update confidence values. After the patch Ψ_p has been filled with new pixel values, the confidence $C(p)$ is updated in the area delimited by $\Psi_{\hat{p}}$ as follows:

$$C(q) = C(\hat{p}), \qquad \forall q \in \Psi_{\hat{p}} \cap \Omega. \tag{3.4}$$

This is a simple update rule which allows us to measure the relative confidence of patches on the fill front, without image-specific parameters.

Generation of Dictionary

In order to compute the sparse representation of the image patch, a dictionary must be determined firstly. In the traditional method of generating a dictionary, there are three kinds of dictionaries. The first kind is the fixed dictionary. For example, an over-complete separable version of the DCT dictionary is constructed by sampling cosine waves with

different frequency in [92]. However, this dictionary is not customized by using appropriate input image data, the adaptability for certain types of data is not good. The second one is the learning dictionary. For instance, the K-SVD dictionary learning algorithm. Although this dictionary is adaptive to the input image date, it has high computational efficiency [1, 105]. The third is the dictionary which is directly generated from the whole patches clipped from the original region [70]. A problem of it is that some unrelated patches can lead to some interference. In order to solve this proplem, we preprocess the candidate patches by using the similarity comparison method of histogram in the proposed method, as shown in Fig. 3.3.

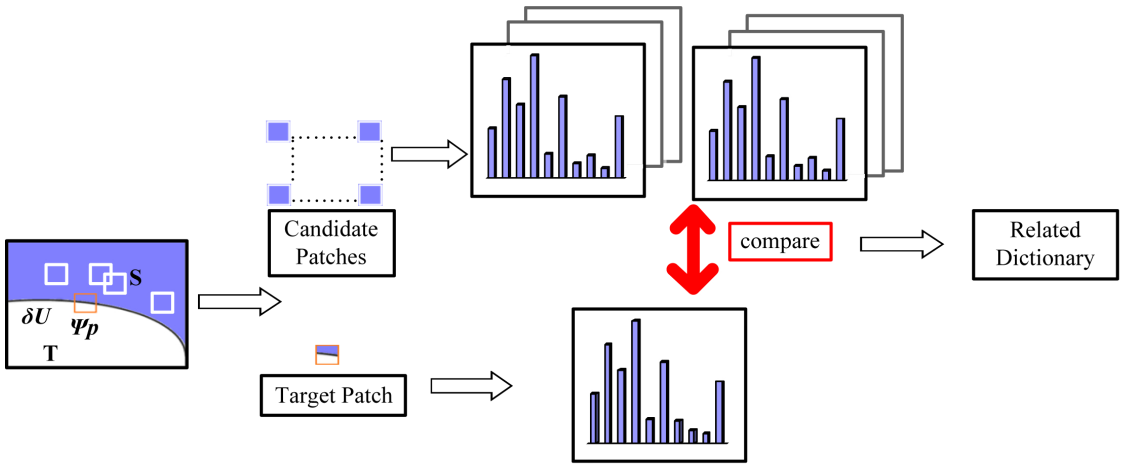


Figure 3.3: Histogram comparison between target patch and candidate patches

In detail, the similarity comparison method by using the sum histogram difference is shown in Fig. 3.4. First, select a target patch Ψ_p using the filling order which is described by Criminisi *et al.* [16, 17] and compute the histogram of it. Let the histogram have N bins. Therefore, the histogram of color image have N values. For three channels R, G, B

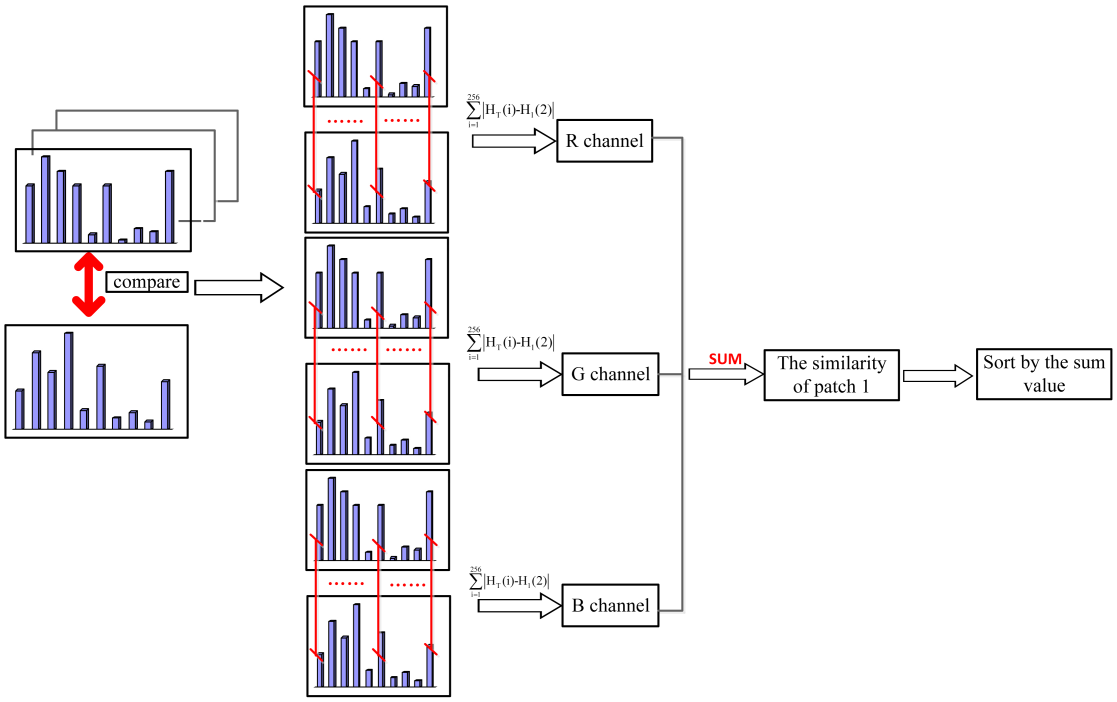


Figure 3.4: The schematic diagram of the proposed sum histogram difference

of target patch Ψ_p, the histogram of each channel can be represented by

$$\mathbf{h}_{\mathbf{\Psi_p}\mathrm{R}} = [h_{\mathbf{\Psi_p}\mathrm{R}1}, \cdots, h_{\mathbf{\Psi_p}\mathrm{R}N}]^T, \tag{3.5}$$

$$\mathbf{h}_{\mathbf{\Psi_p}\mathrm{G}} = [h_{\mathbf{\Psi_p}\mathrm{G}1}, \cdots, h_{\mathbf{\Psi_p}\mathrm{G}N}]^T, \tag{3.6}$$

$$\mathbf{h}_{\mathbf{\Psi_p}\mathrm{B}} = [h_{\mathbf{\Psi_p}\mathrm{B}1}, \cdots, h_{\mathbf{\Psi_p}\mathrm{B}N}]^T. \tag{3.7}$$

Second, cut the whole known patches f_i from the image I and compute the histogram of them. f_i, $(i = 1, \ldots, L)$, with L denotes the number of known patches, $f_{\mathrm{R}i}$, $f_{\mathrm{G}i}$ and $f_{\mathrm{B}i}$ denote three channels (RGB) of the patch, respectively. The histogram of each channel

can be represented by

$$\mathbf{h_{fR}}_i = [h_{f\mathrm{R}i1}, \cdots, h_{f\mathrm{R}iN}]^T, \tag{3.8}$$

$$\mathbf{h_{fG}}_i = [h_{f\mathrm{G}i1}, \cdots, h_{f\mathrm{G}iN}]^T, \tag{3.9}$$

$$\mathbf{h_{fB}}_i = [h_{f\mathrm{B}i1}, \cdots, h_{f\mathrm{B}iN}]^T. \tag{3.10}$$

Thus, two patches Ψ_p and f_i can be use the simplest method to measure similarity by comparing corresponding bins of the histogram. In R, G, B three channels, the difference can be defined by comparing their histogram as follows:

$$V_{Ri} = ||\mathbf{h_{\Psi_p}}_{\mathrm{R}} - \mathbf{h_{fR}}_i||_1 = \sum_{j=1}^{N} (|h_{\Psi_p \mathrm{R}j} - h_{f\mathrm{R}ij}|), \tag{3.11}$$

$$V_{Gi} = ||\mathbf{h_{\Psi_p}}_{\mathrm{G}} - \mathbf{h_{fG}}_i||_1 = \sum_{j=1}^{N} (|h_{\Psi_p \mathrm{G}j} - h_{f\mathrm{G}ij}|), \tag{3.12}$$

$$V_{Bi} = ||\mathbf{h_{\Psi_p}}_{\mathrm{B}} - \mathbf{h_{fB}}_i||_1 = \sum_{j=1}^{N} (|h_{\Psi_p \mathrm{B}j} - h_{f\mathrm{B}ij}|). \tag{3.13}$$

The similarity based on sum V_{Si} is defined as

$$V_{Si} = V_{Ri} + V_{Gi} + V_{Bi}. \tag{3.14}$$

Considering the number of the known patches is L, the similarity based on sum is written by $V_S = [V_{S1,\cdots,V_{SL}}]^T$.

Last, sort the V_S and find the top TN ($TN < L$) known patches to generate the related dictionary.

In the example shown in Fig. 3.5, the blue-green rectangle is denoted as target patch which is the patch we want to inpaint. The mulberry rectangles are denoted as the patches which we choose to generate the related dictionary. For example, we choose the

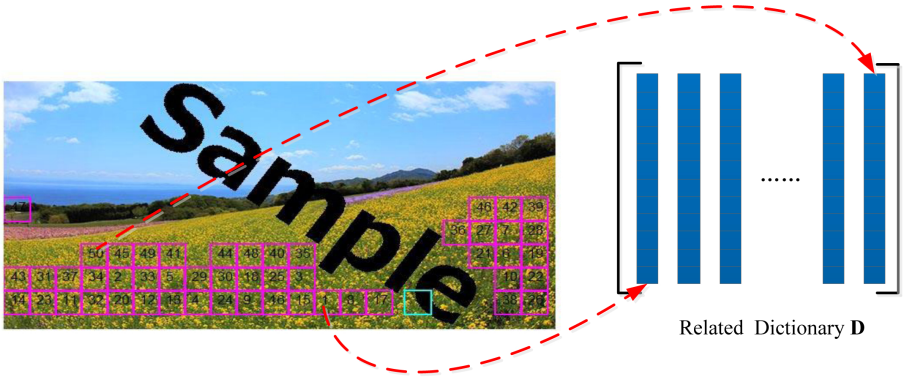


Figure 3.5: Generate the related dictionary by using similar patches

top $TN = 50$ similar patches $\{\Psi_{qj}\}_{j=1}^{50}$, among of them Ψ_{q1} means the most similar one and it represents as the first column of the related dictionary. The rest patches can be done in the same manner.

Signal Recovery

After finding the patch Ψ_p, the aim of the sparse reconstruction is using the known information A to estimate the unknown information B. Matrix M has a special structure determined by the layout of the known pixels. Thus, the known information can be computed by

$$A = M\Psi_p. \tag{3.15}$$

In the problem of image inpainting, it aims to use the known information A to estimate the unknown information B. And A can be seen as the signal y in the theory of the sparse representation, and then the sparse representation can be rewritten by

$$A = MD\alpha, \tag{3.16}$$

where D is the new related dictionary which is generated by the similarity comparison method of the histogram. In the proposed method, we will use the non-negative orthogonal matching pursuit (NNOMP) [91] algorithm which is an improved method of OMP to obtain the estimation of sparse coefficients $\hat{\alpha}$.

The NNOMP algorithm can be described as follows:

- Step 1: Initialize the residual $r_0 = A$ and initialize the set of selected variable $D(c_0) = \emptyset$. And let iteration counter $i = 1$.

- Step 2: Loop over all prototype signal atoms and find the index of the best atom function in D as

  $t_i = argmax < r_{i-1}, d_t >$,

  and add the variable D_{t_i} to the set of selected variables. Update $c_i = c_{i-1} \cup t_i$.

- Step 3: Estimate the sparse representation $\hat{\alpha_i}$ by using Non-Negative Least Square (NNLS)

  $\hat{\alpha_i} = argmin_{\alpha_i \geq 0} \parallel A - D_{c_i}\alpha_i \parallel_2 = (D_{c_i}^T D_{c_i})^{-1} D_{c_i}^T A$.

- Step 4: Update the residual $r_i = A - D_{c_i}\hat{\alpha_i} = A - D_{c_i}(D_{c_i}^T D_{c_i})^{-1} D_{c_i}^T A$.

  Let $P_i = D_{c_i}(D_{c_i}^T D_{c_i})^{-1} D_{c_i}^T$ denote the projection onto the linear space spanned by the elements of D_{c_i}. Then the residual can be rewritten by $r_i = (I - P_i)A$.

- Step 5: Stop the algorithm, if the stopping condition is achieved. Otherwise, set the i=i+1 and return to Step 2.

If we obtain the estimation of sparse coefficients $\hat{\alpha}$ by using the NNOMP, the unknown information B can be restored approximately using

$$B = \bar{M} D \hat{\alpha}, \tag{3.17}$$

where $\bar{M} = E - M$ is a matrix which is decided by the layout of missing pixels and E is also a matrix with each entry being one. In detail, missing pixels in the target patch can be inpainted as follows:

$$\hat{\Psi}_p^i = \begin{cases} \Psi_p^i, & if \quad i \in A \\ (\bar{M}D\hat{\alpha})_i, if & otherwise. \end{cases} \tag{3.18}$$

As discussed above, the schematic diagram of we proposed inpainting algorithm is shown in Fig. 3.6.

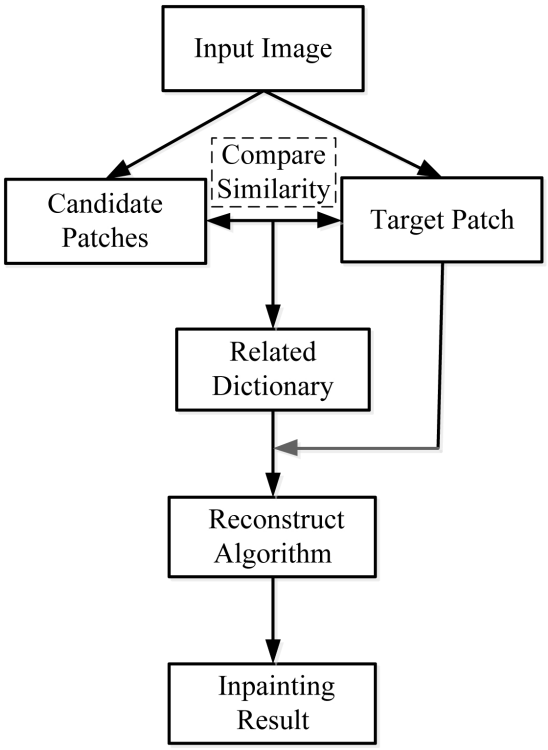


Figure 3.6: Schematic diagram for the proposed inpainting algorithm

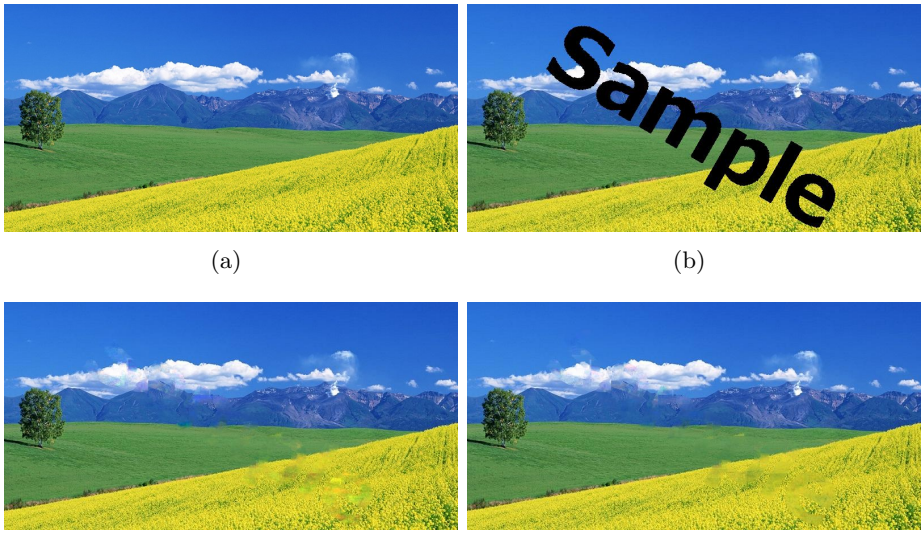
3.1.3 Experiment Results

The experiment is conducted to demonstrate that generating related dictionary by using histogram is more effective than the original dictionary for image inpainting. In this experiment, we adopt different dictionaries to compare with the inpainting results of

them.

Image Inpainting with Different Dictionaries

In order to assess the performance of different dictionaries objectively, two dictionaries are used for comparison: original dictionary and related dictionary. Two pairs of testing images Fig. 3.7 (a) and Fig. 3.7 (b) show the original image and input image of one natural landscape, respectively. Fig. 3.8 (a) and Fig. 3.8 (b) show the original image and input image of another natural landscape, respectively.



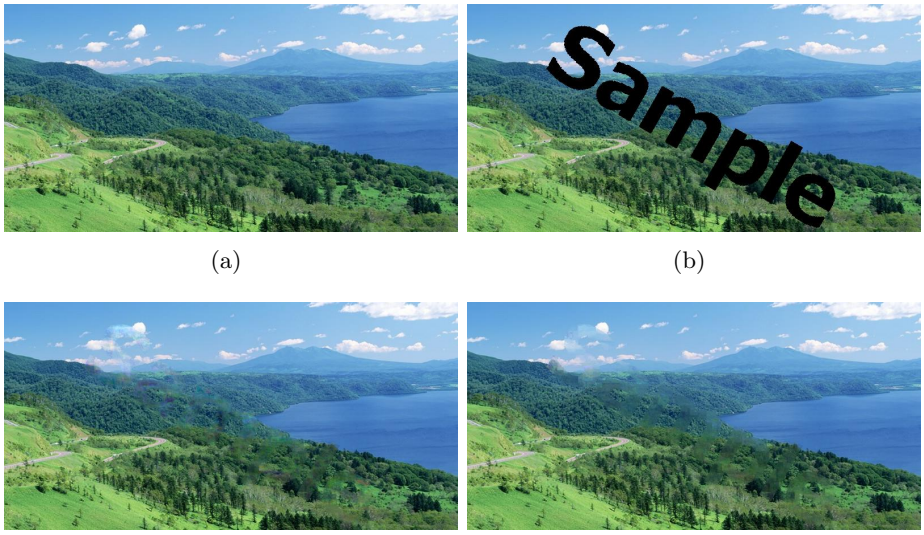
(a) (b)

(c) (d)

Figure 3.7: Inpainting result with different dictionaries for image 1: (a) original image (b) corrupted image including test region (c) inpainting result obtained by using original dictionary (d) inpainting result obtained by using related dictionary.

Evaluation Index

To measure the experimental results of quantitative evaluation by using Peak Signal-to-Noise Ratio (PSNR) obtained from MSE [81]. In addition, the values of PSNR in each color channel (R, G, B) are also computed in experiments. Different from computing mean square error (MSE) of the whole image, we just compute the mean square error of the defective region. Note that the mean square error is computed by using



(a) (b)

(c) (d)

Figure 3.8: Inpainting result with different dictionaries for image 2: (a) Original image (b) corrupted image including test region (c) inpainting result obtained by using original dictionary (d) inpainting result obtained by using related dictionary

$$MSE = \frac{\sum_{(i,j)} \in T(I_{ori}(i,j) - I_{inp}(i,j))^2}{N(T)}, \tag{3.19}$$

where $I_{ori}(i,j)$ is the brightness values of the original image and $I_{inp}(i,j)$ is the brightness values of the inpainted image. T represents the target region, that is to say the defective region, it is noted in Fig. 3.1. And $N(T)$ represents the number of pixels in the defective region.

The PSNR is defined by

$$PSNR = 10log_{10}(\frac{255^2}{MSE}). \tag{3.20}$$

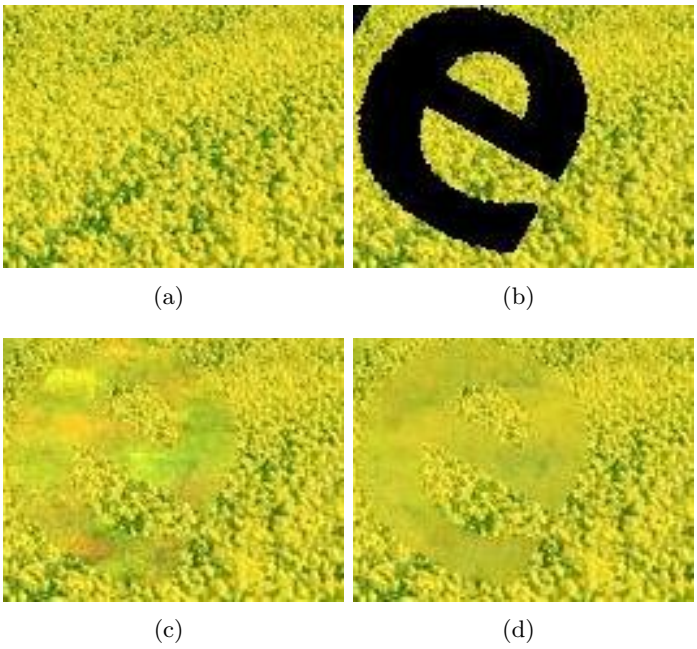
The experiment is conducted based on Fig. 3.6. And then we use the peak Peak Signal-to-Noise Ratio (PSNR) between original image and the inpainted image as the evaluation index to quantify the inpainting results. The results of quantitative metrics are depicted

in Table 3.1. From the Table 3.1, we can learn that the proposed method by using the related dictionary has a better inpainting effect than the method by using the original dictionary.

Table 3.1: Objective performance

	Metric	original dictionary	sub-dictionary
Image 1	R	18.45	18.78
	G	20.58	21.01
	B	23.66	24.19
	RGB	20.40	**20.80**
Image 2	R	18.28	18.79
	G	19.26	19.71
	B	20.25	20.91
	RGB	19.19	**19.72**

In order to view the experiment results easily, Fig. 3.9 and Fig. 3.10 are cropped to show the inpainting result and the same parts of original image and corrupted image.

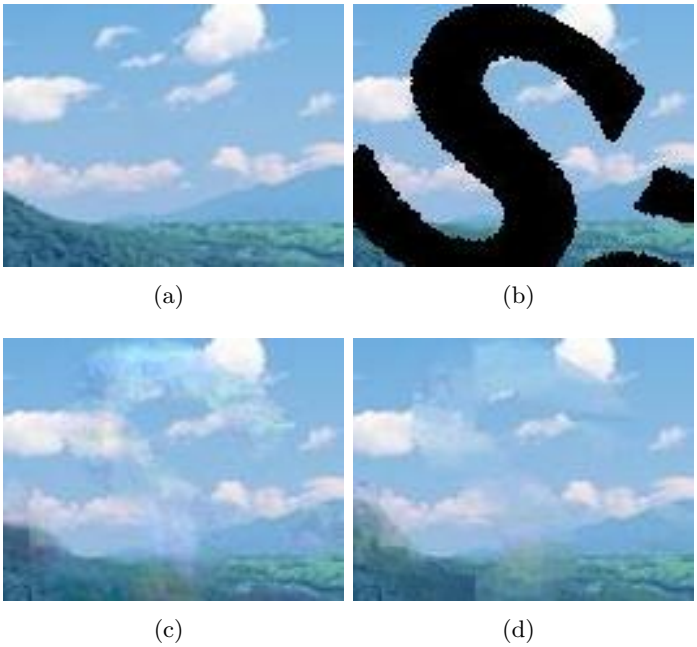


(a) (b)

(c) (d)

Figure 3.9: The cropped part from the image 1 and inpainting result: (a) original image (b) corrupted image (c) inpainting result obtained by using original dictionary (d) inpainting result obtained by using related dictionary.

As shown in Fig. 3.7 (c) and Fig. 3.8 (c), there are some artifacts in the inpainting



(a) (b)

(c) (d)

Figure 3.10: The cropped part from the image 2 and inpainting result: (a) original image (b) corrupted image (c) inpainting result obtained by using original dictionary (d) inpainting result obtained by using related dictionary.

results by using the original dictionary, and they can be shown in Fig. 3.9 (c) and Fig. 3.10 (c) easily. These artifacts are produced because the original dictionary contains the unrelated atoms with the image patch. And the inpainting results by using related dictionary in Fig. 3.7 (d) and Fig. 3.8 (d) obtain good visual effect without a feeling of oddness, and they can also be shown in Fig. 3.9 (d) and Fig. 3.10 (d) easily. From the subjective evaluation and objective evaluation, we can see clearly that the performance of inpainting algorithm using related dictionary are better than inpainting algorithm using original dictionary. This implies that the patch inpainting using related dictionary is more effective than using the original dictionary.

3.1.4 Conclusion

In this paper, we have proposed a new inpainting method based on sparse representation with a related dictionary. The related dictionary is well defined for each target patch,

which consists of the patches in the uncorrupted area having a similar histogram with that of the patch. This is an important step, because it guarantees the inpainting result is more accurate. Thus, the target patch can be inpainted by a sparse representation of the patches in the related dictionary. As shown in the experimental results, the proposed method using related dictionary has obtained better performance than using the original dictionary. The objective quantitative evaluation is consistent with the subjective visual effect of the inpainting result.

In the future, we will concern how to choose out the patches are more relational to the target patch by using the histogram. The chosen patches are more similar to the target patch, the inpainting results will be better. In addition, we also want to study other methods for image inpainting, especially the methods for the inpainting region is large.

3.2 Image Inpainting Based on Histogram Dictionary

3.2.1 An Improved Comparison Method of Histogram

In Section 3.1.1, the related dictionary is proposed to inpaint image. However, there is a problem of it and the problem is shown as following.

Table 3.2: Objective performance

Metric	R Channel	G Channel	B Channel	Sum(R,G,B)
$Sample1$	4.2	3.5	2.3	10
$Sample2$	0.8	8.1	1.1	10
$Sample3$	6.1	1.5	2.4	10

As shown in Table 3.2, the sum values of three samples are all 10, so it will make mistakes when we choose samples as the dictionary. In order to solve this problem, an improved comparison method of histogram has been proposed. Considering the three channels of the color image, we see the similarity of patches as a 3-D vector and find the maximum value of it, then sort them. Using this max method, the order of the case in the Table 3.2 can be shown in Table 3.3. And then we can find the similar patches to generate

the histogram dictionary.

Table 3.3: Objective performance

Metric	R Channel	G Channel	B Channel	Sum(R,G,B)	Max(R,G,B)	Order
*Sample*1	4.2	3.5	2.3	10	4.2	1
*Sample*2	0.8	8.1	1.1	10	8.1	3
*Sample*3	6.1	1.5	2.4	10	6.1	2

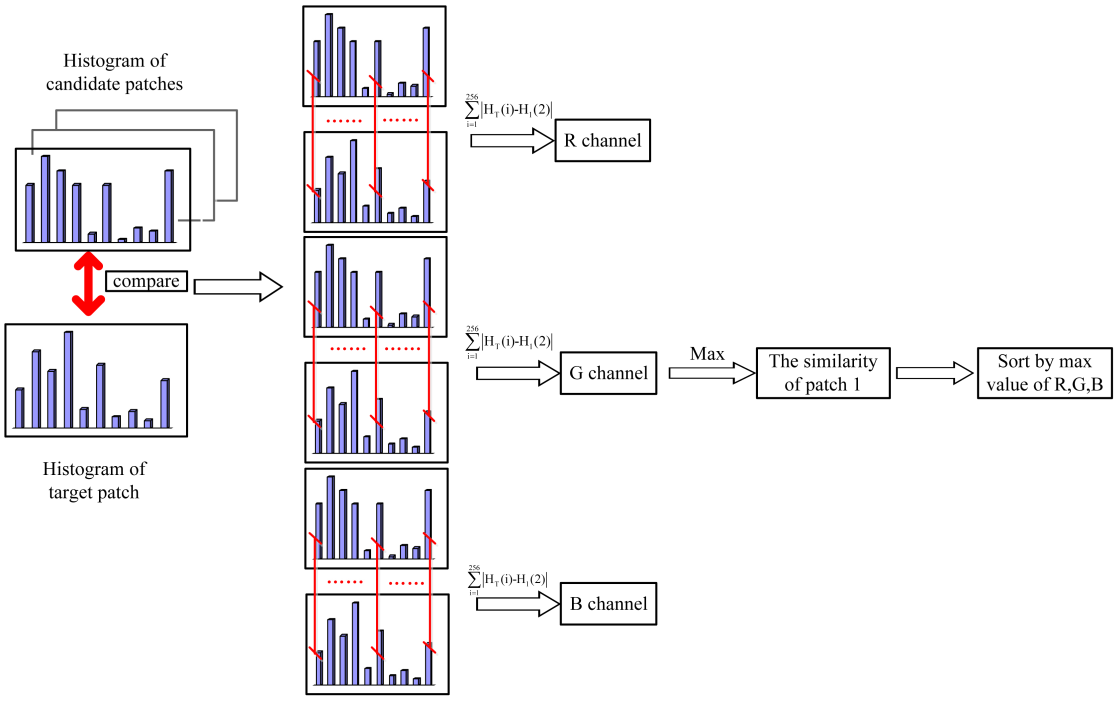


Figure 3.11: The schematic diagram of the proposed maximum histogram difference

The schematic diagram of we proposed maximum histogram difference is summarized in Fig. 3.11. As shown in Fig. 3.11, the big difference between the sum histogram and the max histogram is the processing the differences of R, G, B three channels. Similar to the Eq. (3.14), we denote the max difference of three channels as V_{Mi}. The similarity based on max V_{Mi} is defined as

$$V_{Mi} = max(V_{Ri}, V_{Gi}, V_{Bi}). \tag{3.21}$$

Considering the number of the known patches is L, the similarity based on max is

written by $V_M = [V_{M1}, \cdots, V_{ML}]^T$. And then, sort the V_M and find the top TN_1 ($TN_1 < L$) known patches to generate the histogram dictionary.

Histogram Dictionay

For each target patch, we choose the top $TN_1 = 50$ sample patches, and then use them to construct a histogram dictionary as an example. Fig. 3.12 (a) and (b) are shown the two target patches and their similar patches by using the improved comparison method of histogram. Fig. 3.12 (c) and (d) are shown the chosen patches of them in scatter plots.

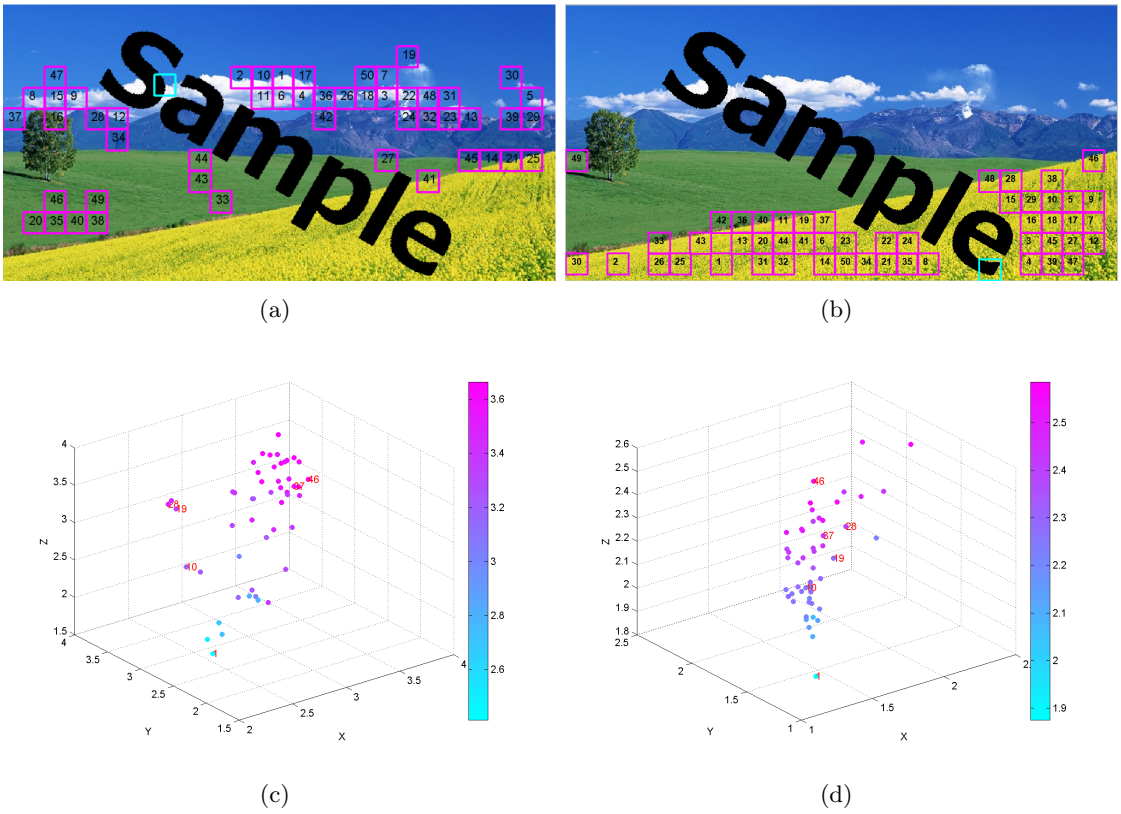
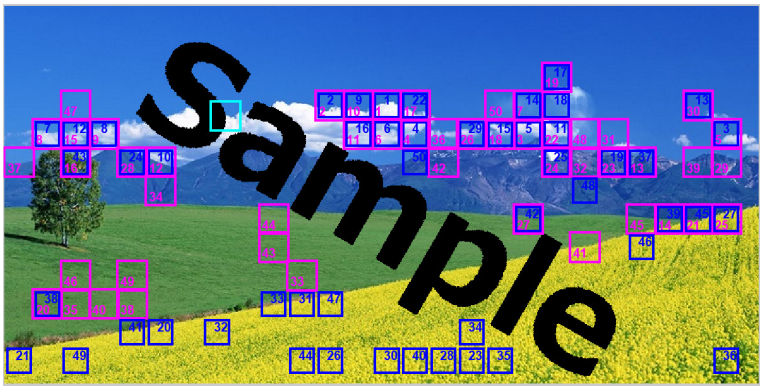


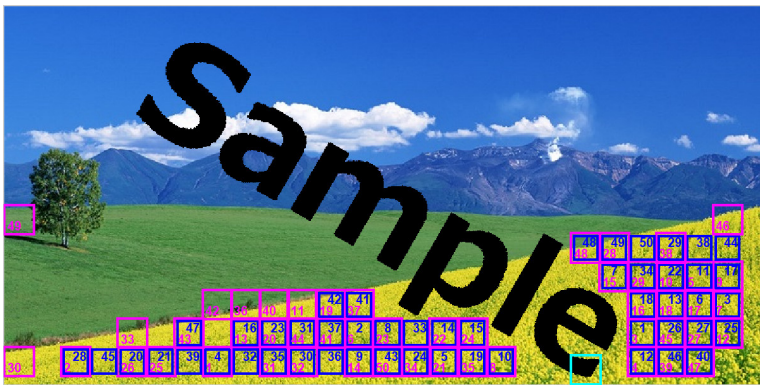
Figure 3.12: The chosen patches by using max histogram (3-D): (a) target patch 1 and its chosen patches, (b) target patch 2 and its chosen patches. (c) the histogram of the target patch 1. (d) the histogram of the target patch 2.

Furthermore, in order to clearly show the difference by using the two different method of histogram for comparing the similarity. Fig. 3.13 shows the chosen patches with the same target patch by using the different comparing methods. It is shown the chosen top

50 patches with the two methods in the same picture. The mulberry block is shown the results of based on max method and the blue block is shown the based on sum method, respectively. From Fig. 3.13 (a), the chosen top 50 sample patches by using similarity based on max contain more relevant patches than that by using similarity based on sum. From Fig. 3.13 (b), the chosen top 50 sample patches by using similarity based on max are generally as relevant as the chosen patches by using similarity based on sum.



(a)



(b)

Figure 3.13: the chosen patches with sum and max histogram are shown in the same picture (a) target patch 1 and its chosen patches by using sum and max histogram, (b) target patch 2 and its chosen patches by using sum and max histogram.

3.2.2 Image Inpainting based on Histogram Dictionary

The process of the image inpainting based on histogram dictionary is similar to the image inpainting based on related dictionary. And the proposed inpainting algorithm is

described in detail in Algorithm 1.

Algorithm 1: The proposed method based on histogram dictionary

Input: the observed image I
Output: inpainting image $\hat{I}$
Initialization : $block = n \times n$, KP, $l = 0$
Repeat :
Clipped the patch f_{l+1} *from the image* I *in a window* $(n \times n)$
if f_{L+1} *is in the source region* S *of image* I **then**
  f_{L+1} *is as a candidate patch compute the histogram of the patch by Eq.*(3.8)
  $L \leftarrow L + 1$
  $KP_{L+1} \leftarrow [KP_L \quad f_{L+1}]$
end
while I *has defective pixels* **do**
  initialization:$V_M = [\ \ ]$
  use the filling order to find the target patch ψ_p
  compute the histogram of it by Eq.(3.5)
  for $i = 1 : L$ **do**
    compute the difference between target patch and the candidate patches by Eq.(3.11)
    $V_{Mi} \leftarrow max(V_{Ri}, V_{Gi}, V_{Bi})$;
    $V_M \leftarrow [V_M \quad V_{Mi}]$;
  end
  $index \leftarrow sort(V_M)$;
  $D_h \leftarrow KP(:, index(1 : L))$;
  $\hat{\alpha} \leftarrow FNNOMP(\psi_p, D_h)$;
  reconstruct the ψ_p *by using Eq.*(3.18);
end
return $\hat{I}$

3.2.3 Experiment

In this part, we test the performance of the proposed method on a variety of natural images. We compare the proposed method with the inpainting algorithm of Criminisi et al. [16, 17]. The proposed method is also employs the inpainting filling order which is described in [16, 17]. And the proposed method is implemented by using Algorithm 1. The number of L is set as 400, that is to say we choose the top 400 patches as the similar dictionary. For fairness, the window size of all methods are set 9×9.

The performance of the proposed method is compared with different image inpainting

methods in quantitative evaluations. We use the peak signal-to-noise ratio (PSNR) as the metrics to evaluate the imapinting results. Furthermore, in order to see clearly, PSNR values in three channels (R, G, B) are also presented.

Figure 3.14: Obtained results of three natural images. The first row shows three original images. The second row shows the degraded images. The third to fifth row show the result of Criminisi *et al.* , the proposed method based on related dictionary and the proposed method based on histogram dictionary.

Considering the two similarity methods, the proposed method inpaint the missing region by using the related dictionary and histogram dictionary. Fig. 3.14 presents three

testing images for this experiment. The first row are the original noncorrupted images. In the remaining rows, from the first and the fourth rows are the corrupted images, results of Criminis *et al.* inpainting algorithm [16, 17], the proposed method by using the related dictionary and the proposed method by using the histogram dictionary. The peak signal-to-noise ratio (PSNR) between the result and the original image is summarized in Table 3.4. It can be seen in Fig. 3.14, the results of Criminisi *et al.* algorithm cause obvious

Table 3.4: Objective performance

image	RGB	Criminisi	Similarity Method	
			sum	**max**
N01	R	31.2285	36.2735	38.8340
	G	32.9927	37.0261	40.4300
	B	33.8244	37.0565	41.3951
	RGB	32.6819	36.7854	**40.2197**
N06	R	31.2285	36.2735	38.8340
	G	32.9927	37.0261	40.4300
	B	33.8244	37.0565	41.3951
	RGB	32.6819	36.7854	**40.2197**
N05	R	31.2285	36.2735	38.8340
	G	32.9927	37.0261	40.4300
	B	33.8244	37.0565	41.3951
	RGB	32.6819	36.7854	**40.2197**
N02	R	31.2285	36.2735	38.8340
	G	32.9927	37.0261	40.4300
	B	33.8244	37.0565	41.3951
	RGB	32.6819	36.7854	**40.2197**

miscopies in the third row. For instance, the snow mountain in (c) of the third row appears unwanted structure of the result of Criminis *et al.*. This is because the method of Criminisi *et al.*, only choose a best match patch to inpaint missing region, some unwanted artifacts appear the results. In the fourth row, the result of the proposed method by using the related dictionary are shown. The edge of the mountain is not inpainted very well in (a) of fourth row. And in (c) of fourth row, extra color produces in the result. For the proposed method by using the reltaed dictionary, the similar patches are chosen by similarity based on max in the framework of sparse representation, so it can overcome the

influences which caused by the inpainting method of Criminisi *et al.*. Furthermore, the histogram dictionary is generated by comparing the difference in 3-D vector, it is more suitable for the color image than the related dictionary for the proposed method. And the fact confirms the quantitative metrics as shown in Table 3.4. The objective quantitative evaluation is consistent with the subjective visual effect of the inpainting result images.

3.2.4 Conclusion

A new image inpainting method based sparse representation has been proposed in this chapter. In order to solve the poor adaptability which caused by the fix dictionary, the proposed method based on sparse representation which use the dictionary constructed directly from all the patches in the known region. And the dictionary constructed all the patches which will be a large number of unrelated atoms came from the image patch to be restored, so they may affect the inpainting result. In order to solve this problem, two measure similarity methods, i.e., based on sum histogram and based on max histogram, are proposed for comparing the similarity between the target patch and all candidate patches. And then the similar patches are chosen to form the related dictionary and histogram dictionary. In this way, the interference of the non-related patches to the sparse construction can be avoided.

The experiment results show that the inpainting results using the histogram are better than that of the related dictionary. And we also compare the proposed inpainting method using the histogram dictionary with the method of Criminisi *et al.*, the experiment results show better performance both in the sense of PSNR quality and visual quality.

Chapter 4

Image Fusion Using Structure-preserving Filter

4.1 Multifoucs Image Fusion Using Structure-preserving Filter

4.1.1 Introduction

Images taken by digital cameras generally suffer from a certain level of degradation, e.g., due to the limited depth of field of lenses, which leads to that parts of imaging objects are focused while others are blurred [28, 102]. It is of great interest to develop multi-focus image fusion (MFIF) [28, 32, 79, 98, 102] techniques, which manage to detect the focused regions in multi-focus images of the same scene and then integrate them to generate a composite image in which all the objects of interest are in focus.

With in this context, numerous MFIF techniques have been developed [28, 32, 79, 98, 102, 53, 44, 66, 41, 40, 36, 92]. According to domains in which the image information is combined, these techniques are roughly divided into transform domain methods, i.e., multi-scale fusion methods, and spatial domain methods, i.e., single scale fusion methods [43, 92]. The former ones mainly cast into three steps. First, compute the transform coeffi-

cients of source images. Second, fuse these coefficients to composite coefficients with some fusion rules. Finally, obtain a fused image by the inverse transform from the composite coefficients. With this basic frame, various transforms have been fully explored for image fusion [40, 54, 42, 62, 104, 52].

It should be noted that this kind of algorithm often requires transforming the source images to different frequency coefficients, i.e., high-frequency and low-frequency coefficients [53]. However, using the high-pass filters usually result to the ringing effects and "halo" around the major structures [102, 108].

In order to well preserve the structure of the original images, structure-preserving filters have been introduced into the multi-scale decomposition, which aim to prevent smoothing across structures while still smoothing texture [102, 88, 75, 27, 33]. Hu and Li [36] integrate edge-preserving characteristic of the bilateral filter and the image directional feature extracted by the filter bank. Bilateral filter-based methods usually involve artifacts around the edge, such as a hole artifact. Further, Kumar [41] uses the cross bilateral filter (a variant of the bilateral filter) to extract the detail images, which is applied to compute weights from the source images and then fuse the multi-sensor and multi-focus images by weighted average. The cross bilateral filter-based method may introduce gradient reversal artifacts in the fused image. Farbman *et al.* [27] use the weighted least squares filter to decompose images for multi-scale tone and detail manipulation, and they indicate that the filter is particularly suited for progressive coarsening of images for multi-scale detail extraction. Because of the property of the method based on weighted least squares filter, it requires the solution of a sparse linear system which limits the performance of the technique. Zhao *et al.* [110] design a detail preserving multi-scale decomposition based on L_0-smoothing filter, which may lose some details since it depends on image gradients. Li *et al.* [44] use the guided filter to refine the weight map and fuse images by a weight averaging rule. Because the input images of the guide filter are the guided and target images, the challenge

of the guided filter-based method is the structure inconsistency between the two input images [51]. Moreover, these structure-preserving filter-based methods in the transform domain have a relatively high time complexity and the original intensities of source images are not well transformed into fusion results [102].

Different from transform domain methods, spatial domain methods process specific pixels directly. Also, the structure-preserving filter used in spatial domain methods is different from it used in transform domain [102, 47]. In particular, the former processes most scales of source images simultaneously, i.e., the intrinsic structure with large scales and details with small scales, but the latter only processes the limited scales which are determined by the decomposition levels. However, compared with the great concentration on applying the structure-preserving filter in transform domain MFIF methods [36, 41, 27, 110, 44, 51], less attention has been paid to the application in spatial domain MFIF methods [102].

The purpose of this paper is to develop a new spatial domain method based on structure-preserving filter for image fusion. The latest recursive filter [29] is introduced which processes image pixels instead of the multi-scale decomposition coefficients, so it can well preserve the original intensities of source images in fusion result. Moreover, a new focused region detection method is proposed based on an average low-pass filter. The method detects the focused regions by the following three steps. First, the rough saliency regions are obtained by comparing the source images and their corresponding smoothed images with an average low-pass filter. This is based on a fact that the intensity error between the source image and its corresponding smoothed image is large for a focused pixel while relatively small for a defocused pixel. Second, in order to make the focus regions more correctly, the average low-pass filter is reapplied to the rough saliency regions. Finally, the initial weight maps are determined by saliency comparison.

4.1.2 Multi-focus Image Fusion

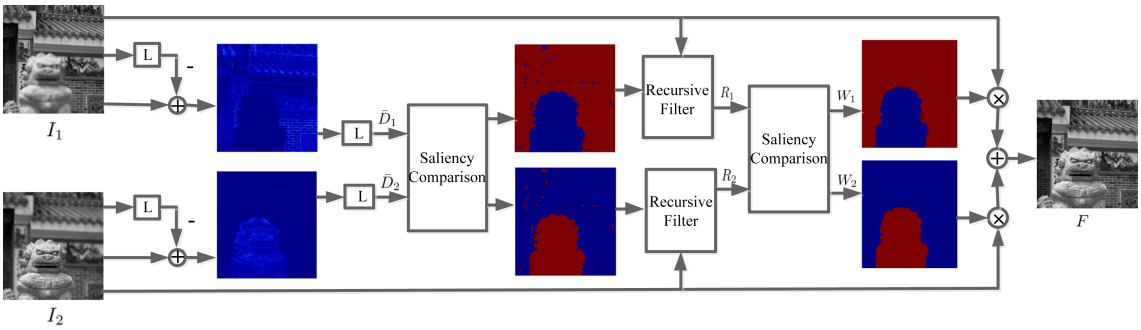


Figure 4.1: Schematic of the proposed image fusion method. The letter 'L' denotes a fast low-pass smoothing filter.

Schematic of the proposed image fusion method is depicted in Fig. 4.1. First, obtain initial weight maps by a new focused region detection method. Second, refine the initial weight maps using the recursive filter to obtain final weight maps. Finally, fuse the source images and final weight maps to achieve the fusion result.

Obtain The Initial Weight Map by a New Focused Region Detection Method

The proposed focused region detection method consists of three steps. First, obtain the rough saliency regions by comparing the source images and their corresponding smoothed images obtained by an average low-pass filter. Second, to make the focus regions more clearly we reapply the average low-pass filter to the rough saliency regions. Finally, the initial weight maps are obtained by saliency comparison.

In the common case of MFIF, there are always two input images, where each one has focused and defocused regions. Khan *et al.* [38] suggest that focused regions are more salient than defocused ones. This fact gives us an idea that the salient regions can be separated from source images by a simple average low-pass filter. In particular, the saliency maps are obtained pixel-by-pixel by comparing original images with their corresponding smoothed images. Since the proposed method does not require breaking source images into blocks, then it can also effectively reduce block artifacts.

As shown in Fig. 4.1, the two source images I^1 and I^2 are first smoothed by an average low-pass smoothing filter

$$\bar{I}_p^m = I_p^m * L_1, \quad m = 1, 2, \tag{4.1}$$

where p is pixel index and L_1 is an average low-pass filter with a default size of 5×5.

Once the smoothed images, i.e., $\bar{I}_p^m$, are obtained, the prediction residual images can be computed easily by comparing the smoothed images with their corresponding source images as

$$D_p^m = |\bar{I}_p^m - I_p^m|, \quad m = 1, 2. \tag{4.2}$$

Because the intensity absolute differences of a focused pixel of the D_p^m have a relatively large value than those in a defocused pixel, we can use them to detect image sharpness. Then, reapply the average low-pass smoothing filter to the residual images:

$$\bar{D}_p^m = D_p^m * L_2, \quad m = 1, 2, \tag{4.3}$$

where $\bar{D}_p^m$ [see Fig. 4.1] are the smoothed images of D_p^m and L_2 is also an average low-pass filter with a default size of 7×7.

Then, the initial weight maps are generated by a saliency comparison

$$P_p^m = \begin{cases} 1, \text{if } \bar{D}_p^m = \max[\bar{D}_p^1, \bar{D}_p^2], m = 1, 2 \\ 0, \text{otherwise.} \end{cases} \tag{4.4}$$

The saliency comparison can be implemented by Algorithm 2.

Algorithm 2: Saliency comparison

```
if D^1 =max(D^1, D^2) then
|  P^1 = 1
else
|  P^1 = 0
end
```

Achieve Final Weight Map Produced by Recursive Filter

It should be noted that the initial weight maps may cause holes and gaps in a homogeneous region [see Fig. 4.1] which may influence the fusion performance. To solve this problem, we refine P_p^m $(m = 1, 2)$ by a real-time structure-preserving smoothing filter, i.e., the recursive filter (RF) [29]. The RF is performed on the P_p^m by using the corresponding I_p^m as the reference images

$$R_p^m = RF(P_p^m, I_p^m), \quad m = 1, 2. \tag{4.5}$$

As shown in Fig. 4.1, R_p^m $(m = 1, 2)$ represent refined weight maps which are the results of RF. Since the RF refines the saliency, it can improve the spatial consistency of weight maps.

The final weight maps are decided by comparing the R_p^m as follows:

$$W_p^m = \begin{cases} 1, \text{if } R_p^m = \max[R_p^1, R_p^2], m = 1, 2 \\ 0, \text{otherwise.} \end{cases} \tag{4.6}$$

Note that this process is a saliency comparison, so it can also be implemented by Algorithm 2.

Generate Fusion Result

Once the final weight maps are obtained, the fused image can be generated directly by

$$F = W_p^1 I_p^1 + W_p^2 I_p^2. \tag{4.7}$$

According to the above description, the proposed fusion method is summarized by Algorithm 3.

Algorithm 3: The proposed method based on recursive filter

Input: the observed image I^1 and I^2
Output: Fused image F
Function Fusion Scheme (I^1, I^2);
$\bar{I}^1 = I^1 * L_1$;
$\bar{I}^2 = I^2 * L_1$;
$D^1 = |\bar{I}^1 - I^1|$;
$D^2 = |\bar{I}^2 - I^2|$;
$\bar{D}^1 = D^1 * L_2$;
$\bar{D}^2 = D^2 * L_2$;
$D_{\max} = \max(\bar{D}^1, \bar{D}^2)$;
for $m \in \{1, 2\}$ **do**
  $P^m = Saliency\ Comparison(\bar{D}^m, D_{\max})$;
  $R^m = Recursive\ Filter(P^m, I^m)$
end
$R_{\max} = \max(R^1, R^2)$;
$W^1 = Saliency\ Comparison(R^1, R_{\max})$;
$W^2 = Saliency\ Comparison(R^2, R_{\max})$;
$F = W^1 I^1 + W^2 I^2$;
return F;
End Function;
Function SaliencyComparsion$S^1, S_{\max}$ **if** $S^1 = S_{\max}$ **then**
  $T^1 = 1$
else
  $T^1 = 0$
end
ReturnT^1;
End Function;
Function RecursiveFilter(P, I) *Obtainging the* R from P and I;
return R;
EndFunction;

4.1.3 Experiments

Experiment Setup

In the experiments, the six pairs of multifocus images: "disk", "lab", "leaf", "newspaper", "clock" and "temple" are used for evaluating the performances of the proposed method. We set $\sigma_s = 40$ and $\sigma_r = 0.2$ as the default parameters in the experiments. And the detailed parameter influence is analyzed in Section 4.1.3.

Objective Assessment Metrics

In this paper, the fusion performances are evaluated by the six metrics, i.e., feature-based metric $Q_p^{ab|f}$ [90], structure-based metric $Q_w^{xy|f}$ [94], the normalized mutual information Q_{MI} [35], a nonlinear correlation information entropy Q_{NICE} [56], Chen-Blum metric Q_{CB} [56, 15], and spatial frequency error Q_{SF} based on spatial frequency [56, 111]. The larger the $Q_p^{ab|f}$, $Q_w^{xy|f}$, Q_{MI}, Q_{NICE} and Q_{CB} values are, the better the fusion results are, while the smaller the Q_{SF} value is, the better the fusion result is.

Compared Methods

The proposed multifocus image fusion method based on recursive filter (MFRF) is compared with the representative image fusion methods including NSCTSR [53], GFF [44], MFGD [66], FFIF [102], and CBF [41]. Liu *et.al* [53] present a general image fusion framework with multi-scale transform and sparse representation. They perform the multi-scale transform on the source images to obtain the low-pass and high-pass coefficients, and then use the sparse representation and the "max-absolute" rule to fuse them, respectively. In the last, they also indicate that the framework based on a nonsubsampled contourlet transform (NSCT) and sparse representation (NSCTSR) [53] is best for multi-focus image fusion. The fusion methods based on guide filter (GFF) [44], a fast structure filter (FFIF) [102] and cross bilateral filter (CBF) [41] are all fusion methods

based on structure-preserving filters as introduce in Section 1. Paul *et.al* [66] propose a novel method for multi-exposure and multifocus image fusion in gradient domain (MFGD). The fusion processes of the luminance and chrominance channels are treated in different ways. The MFGD is based on blending the gradients of the luminance components of the source multifocus images and use the image reconstruction technique of Harr wavelet-based to fuse luminance.

To show the effectiveness of MFRF comparing with the fusion methods in transform domain, so we choose one of them, i.e., the fusion method based NSCT (NSCTSR) [53]. We also select a gradient domain method, i.e., MFGD [66], to demonstrate the proposed spatial method can get better fusion performances than the existing methods. And as the MFRF is based on structure-preserving filter, we also choose three fused methods based on structure-preserving filters, i.e., GFF [44], FFIF [102] and CBF [41], to show the effectiveness of MFRF. For all these methods, we use the default parameters given in authors' papers and source codes provided by them.

Experiment Results

The first experiment is conducted to demonstrate that the average filter chosen as the low-pass filter in the proposed method is more effective than other low-pass filters. The second experiment is conducted to demonstrate that the proposed method achieves the state-of-the-art fusion performance.

Compare with Different Low-pass Filters

There are some common low-pass filters such as average filter (AVE), Gaussian filter (GAU), median filter (MED) and bilateral filter (BLF). In order to find an effective low-pass filter for the proposed method, we conduct an experiment for six pairs of testing images on this subsection. For a fair comparison, we use the same framework as shown in Fig. 4.1 and the same window size with the four low-pass filters. The fusion performances

are evaluated by the values of $Q_p^{ab|f}$, $Q_w^{ab|f}$, Q_{MI}, Q_{NICE}, Q_{CB} and Q_{SF}. The experimental results are shown in Fig. 4.2. As shown in the Fig. 4.2, the fusion results of the AVE and GAU get the higher metrics value than that of the other two low-pass filters for the top five metrics while getting the smaller value of Q_{SF}. And considering the AVE is a simple one without setting the standard deviation parameter σ which is important in the GAU, thus we choose it for the proposed method.

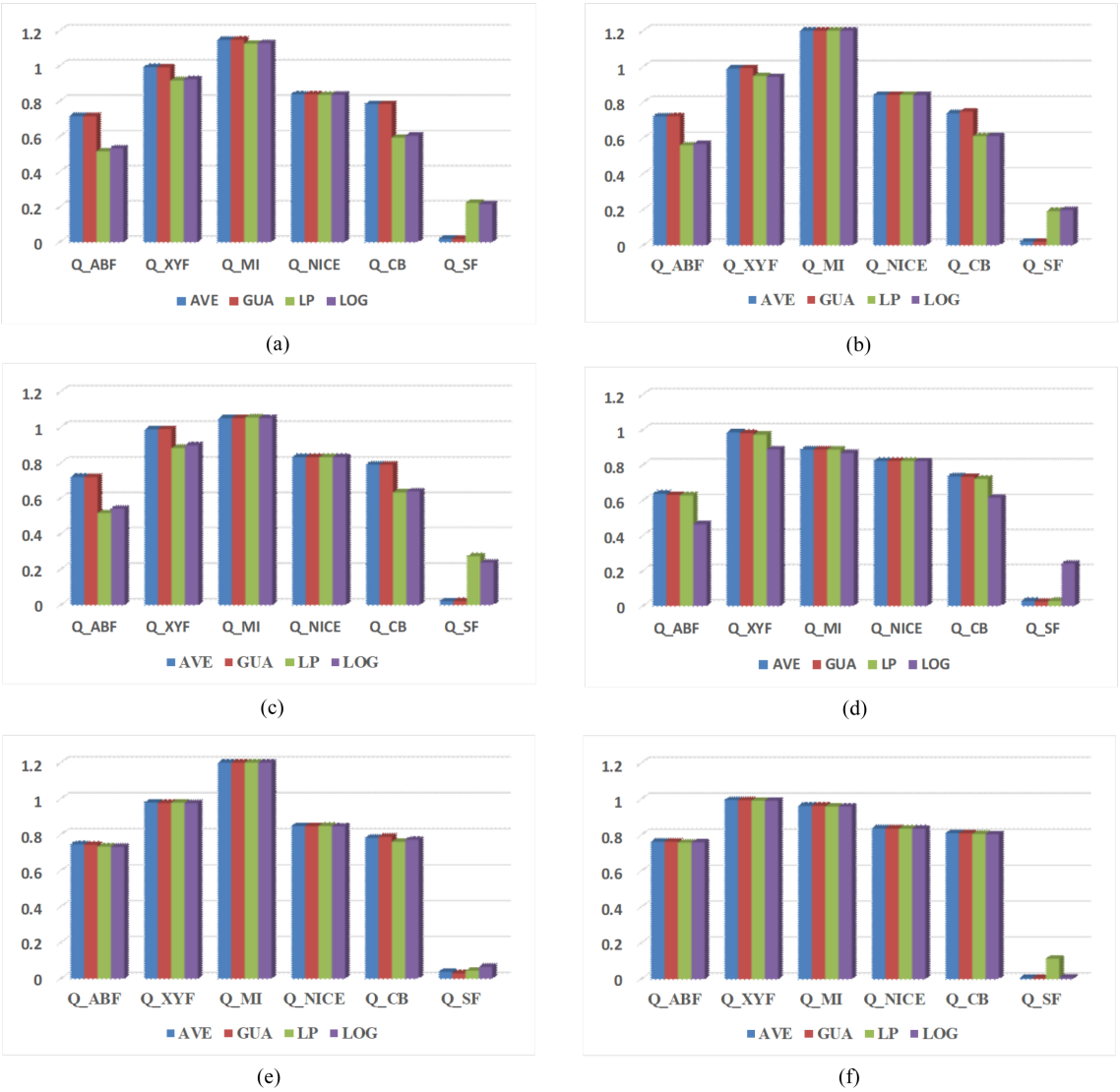
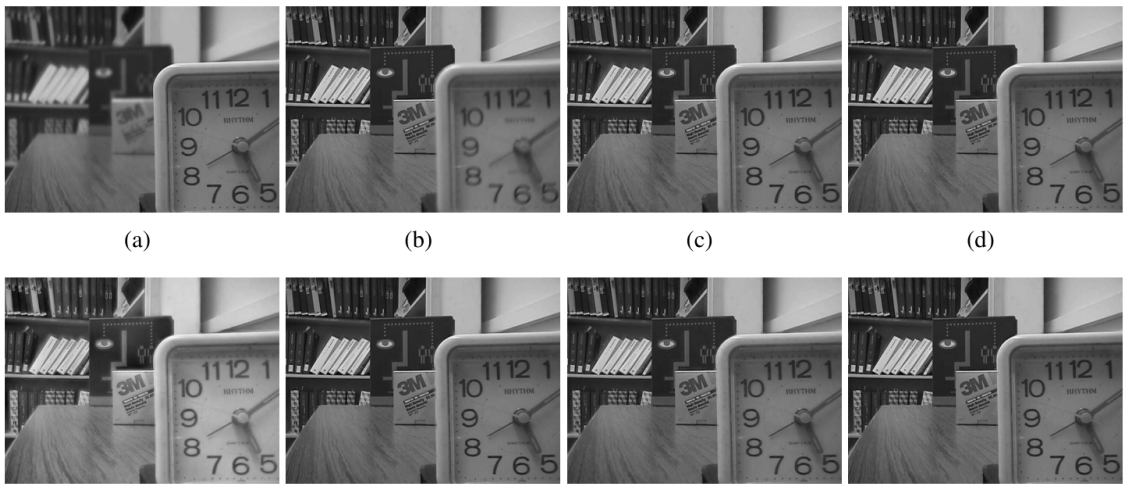


Figure 4.2: Fusion performance of different low-pass filters (a) disk, (b) lab, (c) leaf, (d) newspaper, (e) clock, (f) temple.



(a) (b) (c) (d)

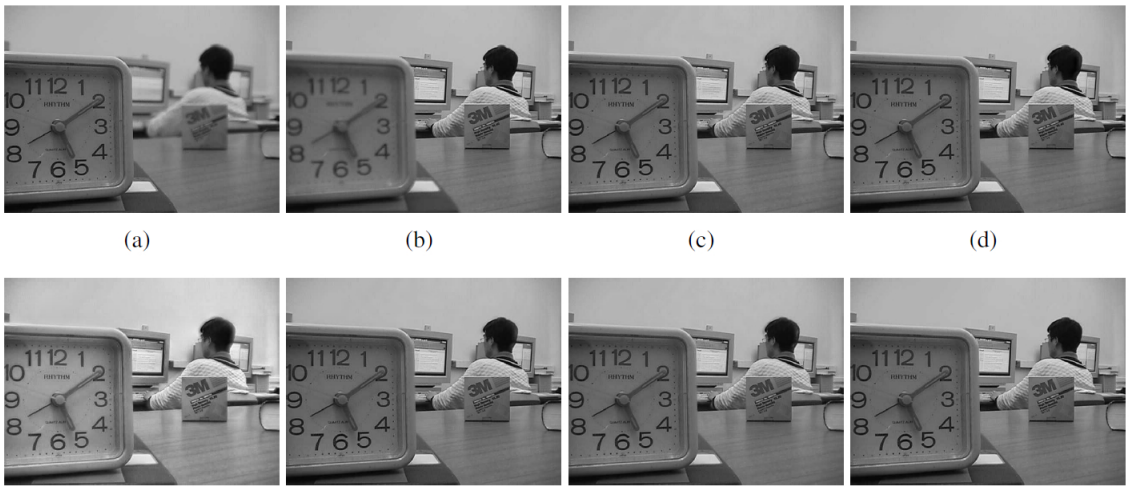
(e) (f) (g) (h)

Figure 4.3: Fusion results obtained by different methods for multi-focus image "disk" (a) disk 1 (b) disk 2 (c) NSCTSR (d) GFF (e) MFGD (f) FFIF (g) CBF (h) MFRF.

Compare with Different Fusion Methods

In this section, the proposed method is compared with five state-of-the-art algorithms : NSCTSR [53], GFF [44], MFGD [66], FFIF [102], and CBF [41] which are described in detail in Section 4.1.3. Experiments are conducted on the six pairs of multi-focus images as shown in Figs. 4.3 - 4.8 (a) and (b). Fig. 4.3 represents the fused images of different fusion algorithms for the "disk" dataset. As shown in Figs. 4.3 (c) and (g), there are some artifacts in the NSCTSR and CBF fusion results. Fusion results of NSCTSR, GFF and CBF methods have "halo" artifacts in the boundary of the clock [see Figs. 4.3 (c), (d) and (g)]. It can be seen in Fig. 4.3 (e), the fusion result of MFGD method increases the brightness and causes many blurring artifacts. The fusion result of FFIF has artifacts in vertical edges of the clock [see Fig. 4.3 (f)]. That is to say, the edges are not well preserved. The proposed MFRF method obtains better performance in visual quality [see Fig. 4.3 (h)].

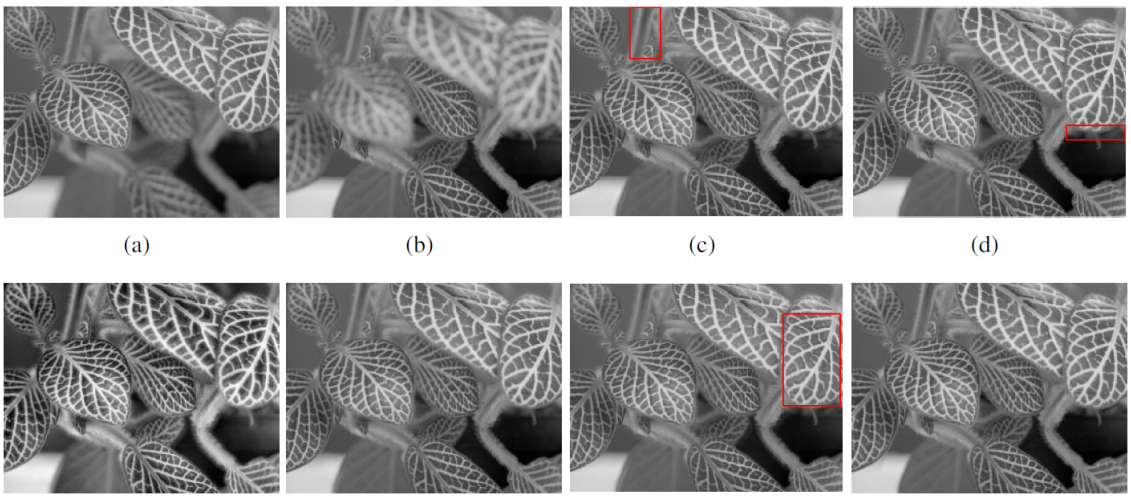
Fig. 4.4 displays the fusion results on "lab" dataset. From Fig. 4.4 (c), Fig. 4.4 (d), Fig. 4.4 (f) and Fig. 4.4 (g), it is clear that the fusion results produced by NSCTSR, GFF,



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4.4: Fusion results obtained by different methods for multi-focus image "lab" (a) lab 1 (b) lab 2 (c) NSCTSR (d) GFF (e) MFGD (f) FFIF (g) CBF (h) MFRF.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4.5: Fusion results obtained by different methods for multi-focus image "leaf" (a) leaf 1 (b) leaf 2 (c) NSCTSR (d) GFF (e) MFGD (f) FFIF (g) CBF (h) MFRF.

FFIF and CBF methods have "halo" artifacts around the head of the human. It can be seen from Fig. 4.4 (e) that the fusion result of MFGD also increases the brightness of the source images and has blurring artifacts and ringing artifacts in the edge of the head and body of the human. Compared with other methods, the fusion result of the proposed MFRF has fewer artifacts [see Fig. 4.4 (h)].

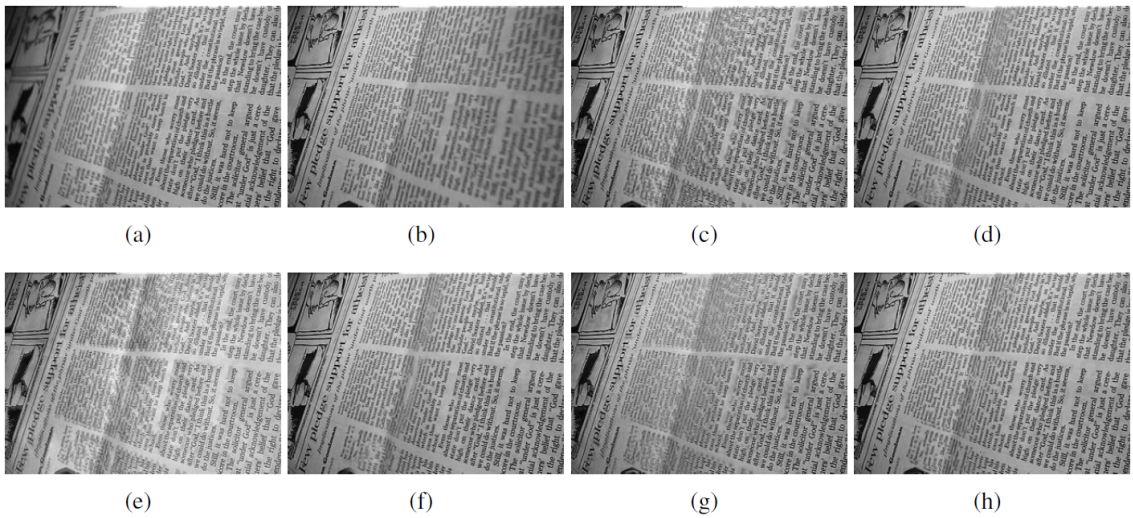


Figure 4.6: Fusion results obtained by different methods for multi-focus image "newspaper" (a) newspaper 1 (b) newspaper 2 (c) NSCTSR (d) GFF (e) MFGD (f) FFIF (g) CBF (h) MFRF.

The fusion results with different methods on "leaf" dataset are shown in Fig. 4.5. The fusion result of NSCTSR method has some blurring artifacts in the strong edge indicated in the red rectangle [see Fig. 4.5 (c)]. As shown in Fig. 4.5 (d), some artifacts are observed in the edge of a leaf which is also indicated by the red rectangle in the fusion result of GFF method. The brightness of the fusion result of MFGD method is higher than those of other fusion methods [see Fig. 4.5 (e)]. As seen from Fig. 4.5 (f), the fusion result of FFIF method causes blurring artifacts in the small leaf which are seen clearly in the red rectangle. From Fig. 4.5 (g), it can be seen that the fusion result of CBF method decreases the sharpness of some region shown in the red rectangle.

Fig. 4.6 shows the fusion results obtained by different fusion algorithms on the "newspa-

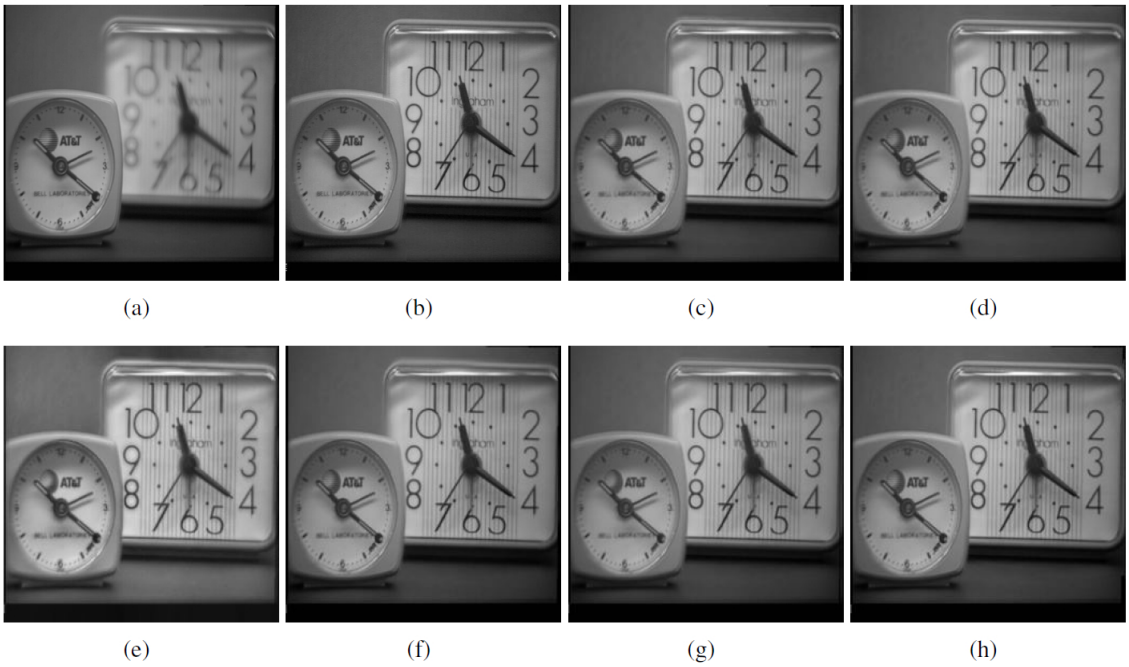


Figure 4.7: Fusion results obtained by different methods for multifocus image "clock" (a) clock 1 (b) clock 2 (c) NSCTSR (d) GFF (e) MFGD (f) FFIF (g) CBF (h) MFRF.

per" dataset. As can be seen from Fig. 4.6 (c), the fusion result of method CBF generates the ringing and blurring artifacts. Fig. 4.6 (d), Fig. 4.6 (f) and (g) display the fusion results obtained by NSCTSR, GFF and CBF algorithms, there are some blurring artifacts in the boundary between the focused and defocused regions. Some shadows like artifacts edge occur around the letter in the fusion result of method MFGD [see Fig. 4.6 (e)]. In Fig. 4.6 (e) most of the words are blurred, so it is hard to get some information from them. It can also be observed that there are still some blurring artifacts in the fusion result of the MFRF method [see Fig. 4.6 (h)]. However, it has fewest artifacts in the fusion results.

Fig. 4.7 shows the fusion results of "clock" dataset. As shown in Fig. 4.7 (c) and (g), the fusion results obtained by NSCTSR and CBF have artifacts in the top left corner of the big clock. For the image brightness, the fusion result in Fig. 4.7 (e) increases image brightness and also has blurring artifacts. The fusion result of FFIF method is blurred as shown in the top-right edge of the bigger clock and decreases the sharpness [see Fig. 4.7

(f)]. It is observed from Fig. 4.7 (d) and (h) that GFF and MFRF algorithms work better than the other algorithms.

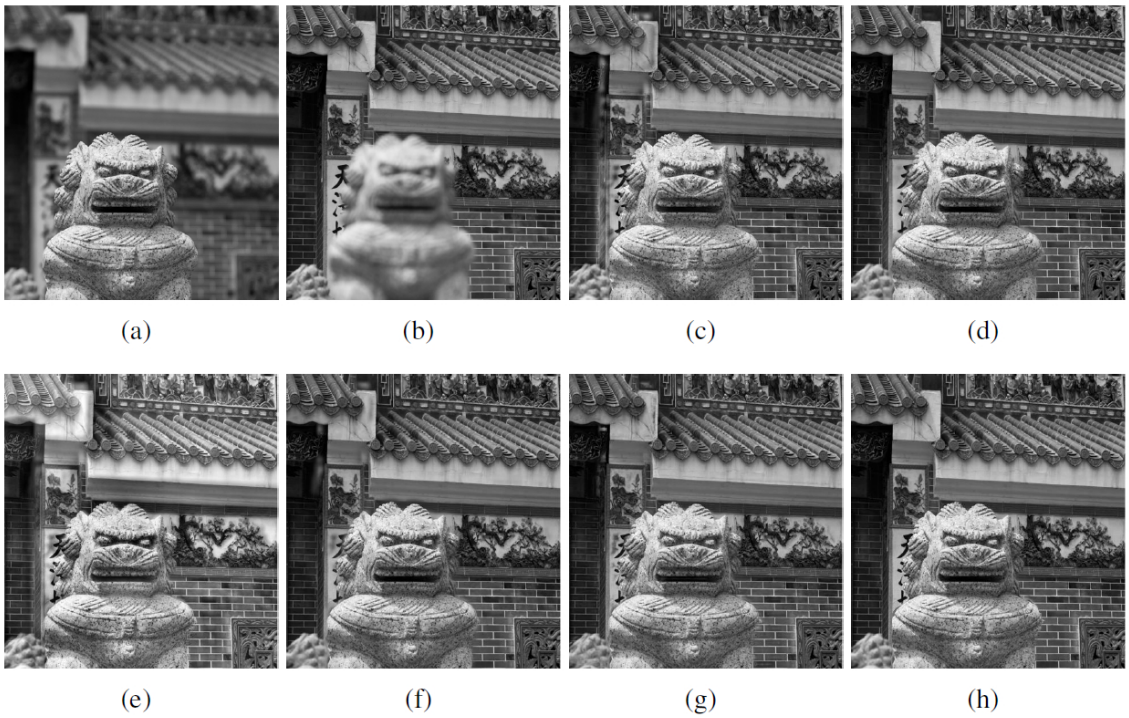


Figure 4.8: Fusion results obtained by different methods for multifocus image "temple" (a) temple1 (b) temple2 (c) NSCTSR (d) GFF (e) MFGD (f) FFIF (g) CBF (h) MFRF.

The fusion results of different algorithms for "temple" dataset are shown in Fig. 4.8. It can be seen from Fig. 4.8 (c) that, there are some blurring artifacts around the stone lion and additional shadows in the background. For the image brightness, the fusion result obtained by the GFF method decreases image brightness [see in Fig. 4.8 (d)]. As shown in Fig. 4.8 (e), the fusion result of MFGD method not only produces the blurring artifacts but also increases image brightness. From Fig. 4.8 (g) and (f) shown the fusion results of NSCTSR and FFIF algorithms, there are additional shadows in the background. In general, Figs. 4.4 - 4.8 (e) have blurring artifacts and increase image brightness among the results. From Figs. 4.4 - 4.8 (h), it can be seen that the proposed MFRF method obtain good fusion results in the test six pairs of images for subjective evaluation.

The objective evaluation of the fused results for the test six pairs of images are shown

in Table. 4.4. In order to see more clearly, the highest values of each objective metrics have been marked in bold. From Table. 4.4, it can be seen that the fusion results of MFRF usually achieve the highest values of the evaluation metrics. Based on above analysis of subjective and objective evaluation, it can be seen that the proposed algorithm obtains the better fusion results than the other methods [53, 44, 66, 102, 41].

Computational Time Analysis

We evaluate the computation time of different methods to explore the efficiency of the MFRF. All experiments are performed on a PC with an Intel (R) Core (TM) i7-2600K 3.40 Hz CPU and 16.0 GB RAM in MATLAB. Because the different image pairs have different spatial resolutions, e.g, the images disk and lab have the same resolutions 640 $\times$ 480, the images leaf have 268 $\times$ 204 resolutions, the images newspaper have 322 $\times$ 234 resolutions, the images clock have 256 $\times$ 256 resolutions and the images temple have 481 $\times$ 516 resolutions, the experiments are conducted on five resolutions.

The computation time of different fusion algorithms is executed 8 times and the average time for each test image pair is shown in Table. 4.2. The execution time for the FFIF algorithm is comparatively short when compared to the other methods, which is quite obvious since it uses the fast algorithm. In contrast, the execution time of MFRF is less than the NSCTSR and CBF methods and is comparable with that of the other two methods. Therefore, the MFRF is promising for a real-time implementation.

Parameter Influence

In order to demonstrate the MFRF robustness, we study the influences of the parameters σ_s and σ_r. The σ_s and σ_r control the space and range supports of the RF, respectively. Because the six metrics are the same order of magnitude and in order to save space, we just reveal the experiment result of metric $Q_p^{ab|f}$ which is shown in Fig. 4.9. As shown in Fig. 4.9, the influence of σ_r and σ_s are varying from 0.2 to 2 with intervals 0.2 and

Table 4.1: Objective performance

Images	Metric	NSCTSR	GFF	MFGD	FFIF	CBF	MFRF
disk	$Q_p^{ab\|f}$	0.6918	0.7054	0.6744	0.7124	0.6813	**0.7148**
	$Q_w^{xy\|f}$	0.9179	0.9405	0.8312	0.9583	0.8966	**0.9926**
	Q_{MI}	0.8628	0.9732	0.5122	1.0573	0.9203	**1.1482**
	Q_{NICE}	0.8236	0.8291	0.8122	0.8338	0.8263	**0.8390**
	Q_{CB}	0.6987	0.7247	0.5863	0.7560	0.6747	**0.7841**
	Q_{SF}	0.0281	0.0331	0.1407	0.0295	0.0649	**0.0187**
lab	$Q_p^{ab\|f}$	0.6960	0.7134	0.6615	0.7188	0.6942	**0.7208**
	$Q_w^{xy\|f}$	0.8919	0.9417	0.7894	0.9230	0.8661	**0.9884**
	Q_{MI}	1.0249	1.1331	0.5987	1.1846	1.0690	**1.2347**
	Q_{NICE}	0.8310	0.8360	0.8165	0.8389	0.8330	**0.8414**
	Q_{CB}	0.6648	0.6919	0.5877	0.6983	0.6354	**0.7379**
	Q_{SF}	0.0283	0.0328	**0.0094**	0.0321	0.0800	0.0203
leaf	$Q_p^{ab\|f}$	0.7078	0.7177	0.6353	0.7199	0.7074	**0.7212**
	$Q_w^{xy\|f}$	0.9566	0.9697	0.8276	0.9794	0.9553	**0.9880**
	Q_{MI}	0.9566	0.7738	0.3728	0.9181	0.7553	**1.0501**
	Q_{NICE}	0.8146	0.8184	0.8073	0.8249	0.8176	**0.8318**
	Q_{CB}	0.7465	0.7684	0.5758	0.7799	0.7326	**0.7911**
	Q_{SF}	0.0264	0.0371	0.3883	0.0365	0.0681	**0.0207**
newsp	$Q_p^{ab\|f}$	0.5695	0.6226	0.5481	0.6290	0.5471	**0.6367**
	$Q_w^{xy\|f}$	0.9395	0.9824	0.8871	**0.9898**	0.9030	0.9847
	Q_{MI}	0.3048	0.6085	0.2635	0.8153	0.3655	**0.8871**
	Q_{NICE}	0.8046	0.8119	0.8040	0.8198	0.8057	**0.8229**
	Q_{CB}	0.6660	0.7297	0.5803	0.7423	0.6239	**0.7352**
	Q_{SF}	0.0395	0.0389	0.1875	0.0350	0.1130	**0.0268**
clock	$Q_p^{ab\|f}$	0.7375	0.7403	0.7142	0.7435	0.7380	**0.7464**
	$Q_w^{xy\|f}$	0.9319	0.9418	0.7992	0.9701	0.9390	**0.9786**
	Q_{MI}	1.0507	1.1031	0.5989	1.1953	1.0791	**1.2513**
	Q_{NICE}	0.8350	0.8384	0.8174	0.8428	0.8364	**0.8473**
	Q_{CB}	0.7622	0.7666	0.6514	0.7763	0.7266	**0.7882**
	Q_{SF}	0.0424	0.0546	**0.0049**	0.0621	0.1032	0.0376
temple	$Q_p^{ab\|f}$	0.7152	0.7556	0.7367	0.7543	0.7449	**0.7631**
	$Q_w^{xy\|f}$	0.9386	0.9884	0.9040	0.9819	0.9492	**0.9952**
	Q_{MI}	0.4341	0.7543	0.3054	0.8573	0.7085	**0.9629**
	Q_{NICE}	0.8078	0.8228	0.8050	0.8277	0.8184	**0.8364**
	Q_{CB}	0.6882	0.7917	0.6505	0.7949	0.7481	**0.8103**
	Q_{SF}	0.0152	0.0255	0.0657	0.0262	0.0207	**0.0055**

Table 4.2: Average computation time (in seconds) comparison

Images	NSCTSR	GFF	MFGD	FFIF	CBF	MFRF
disk	44.6275	0.3572	1.0031	0.1000	52.3861	0.2887
lab	39.2177	0.3608	1.0045	0.1001	52.9350	0.2803
leaf	9.2184	0.0315	0.2165	0.0081	9.4504	0.0876
newsp	12.1355	0.0471	0.2245	0.0119	12.9385	0.1071
clock	10.8090	0.0411	0.0574	0.0104	11.4014	0.0950
temple	44.7760	0.2812	0.9639	0.0782	42.5466	0.2566

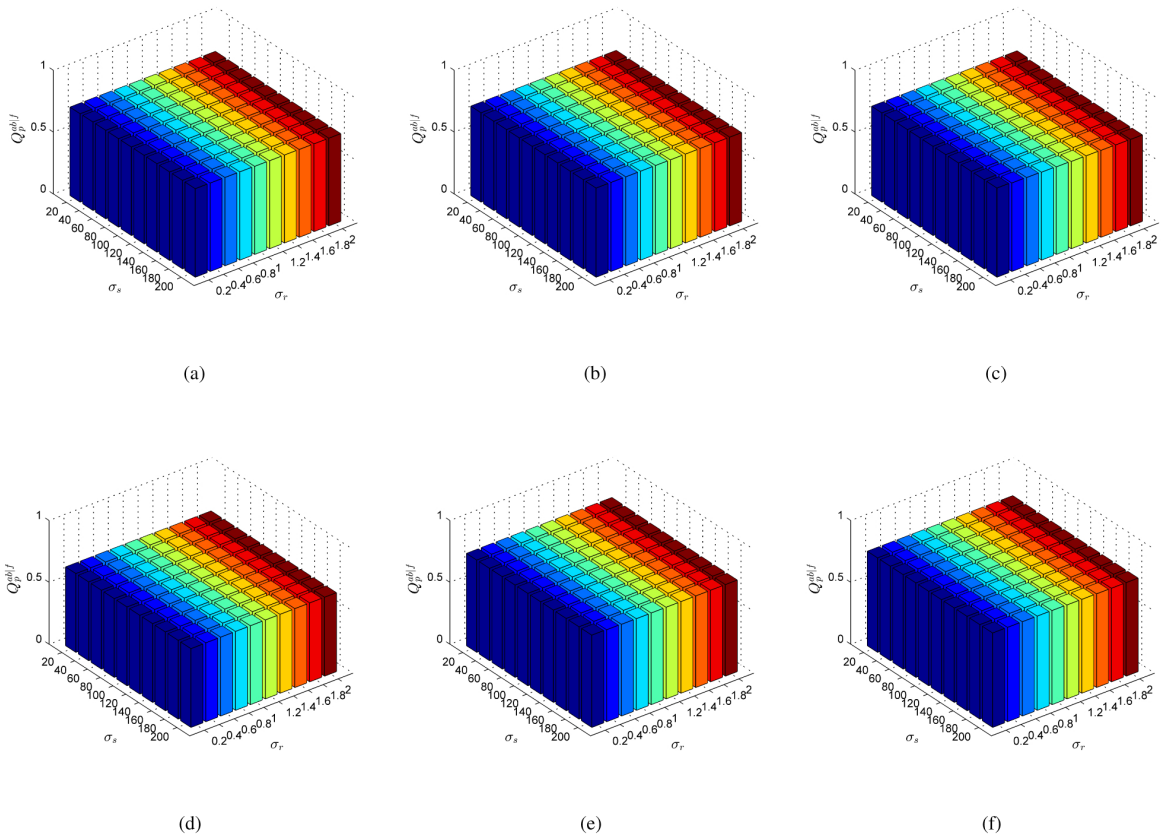


Figure 4.9: $Q_p^{ab|f}$ with respect to different σ_r and σ_s for six pairs multifocus images : (a) disk (b) lab (c) leaf (d) newspaper (e) clock (f) temple.

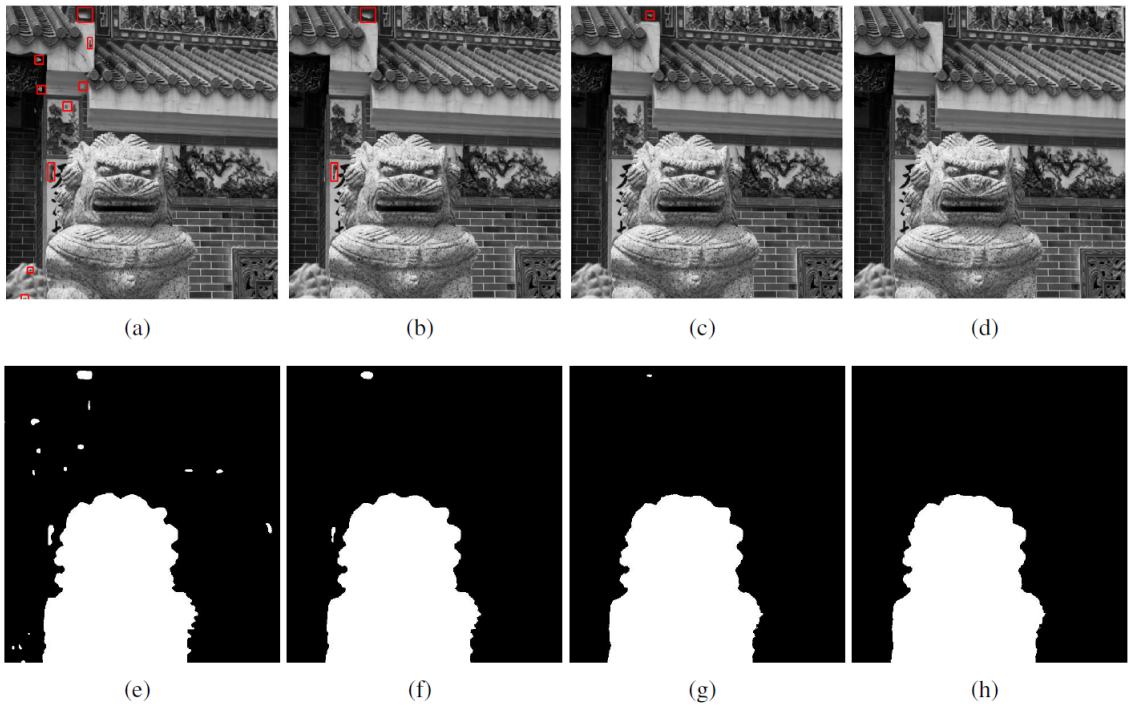


Figure 4.10: fusion results of different σ_s multifocus image "temple" and corresponding weight maps of them : (a) σ_s=5 (b) σ_s=10 (c) σ_s=20 (d) σ_s=40 (e) weight map of (a) (g) weight map of (b) (f)weight map of (c) (h) weight map of (d).

20 to 200 with intervals 20, respectively. Since the amplitudes of the column and row of Fig. 4.9 (a)-(f) change slightly, the objective performances of the MFRF are not sensitive to the altering of σ_r and σ_s. It can be seen that the performance of the proposed method is robust against the parameter variant in a wide range.

Table 4.3: Average performance in each data

Metrics	NSCTSR	GFF	MFGD	FFIF	CBF	MFRF
$Q_p^{ab\|f}$	0.7188	0.7359	0.6866	0.7330	0.7316	**0.7380**
$Q_w^{xy\|f}$	0.9555	0.9779	0.8386	0.9824	0.9540	**0.9861**
Q_{MI}	0.9752	1.1137	0.4989	1.1689	1.0202	**1.2247**
Q_{NICE}	0.8317	0.8400	0.8140	0.8438	0.8341	**0.8472**
Q_{CB}	0.7020	0.7811	0.6119	0.7871	0.7486	**0.7978**
Q_{SF}	0.0309	0.0352	0.1382	0.0565	0.0919	**0.0246**

Moreover, in order to show the subjective performance of the proposed method, we fix $\sigma_r = 0.2$ and test the fusion results with different values of σ_s in image pair "temple". Fig. 4.10 shows the fusion results of different σ_s by using the proposed method. Fig. 4.10 (a) - (d) show the fusion results of $\sigma_s = 5$, $\sigma_s = 10$, $\sigma_s = 20$ and $\sigma_s = 40$. The corresponding focus maps of them are shown in Fig. 4.10 (e) - (h). When $\sigma_s = 5$, the fusion result shown in Fig. 4.10 (a) has some artifacts which are noted by a red rectangle. The focus map of it is shown in Fig. 4.10 (e), and some focus regions are wrong. As the value of σ_s is increasing, the wrong focus regions are decreasing. When $\sigma_s = 40$, the focus map shown in Fig. 4.10 (h) is correct and the fusion result shown in Fig. 4.10 (d) is good. Additionally, as shown in Fig. 4.9 the value of σ_s from 40 continues to increase, the values of metric $Q_p^{ab|f}$ of six tested images are stable. The experiment results show that good fusion performance is obtained with these parameters.

Further Statistical Experiment

To carry out statistics valuation, more experiments are conducted on Zhang *et al.*'s dataset [109] and the Lytro dataset [61]. Excepting the images used in our experiments,

we choose 12 multi-focus images in the Zhang *et al.*'s dataset, and 20 images in Lytro. Totally, we have applied 32 images to demonstrate the effectiveness of MFRF.

We calculate the average performance of different methods on these 32 images. The result is listed in Table 4.3. As shown in Table 4.3, the MFRF method outperforms the other methods in terms of the six metrics.

4.1.4 Conclusion

In this paper, a fusion method has been proposed based on RF in the spatial domain. Particulary, the latest RF has been introduced as the structure-preserving filter. A focused region detection method has been presented to generate initial weight maps, which utilizes the absolute differences between source images and their corresponding smoothed images to detect saliency region. The RF has been used to refine the initial weight maps to obtain the refinement weight maps. The refinement weight maps can improve the spatial consistency and then the fusion results. Experimental results have shown that the proposed method presents superior performance in terms of both visual performance and objective metrics. Furthermore, the proposed method is not sensitive to the setting of the parameters and is promising for a real-time implementation.

4.2 Medical Image Fusion Using Segment Graph Fliter and Sparse Representation

4.2.1 Introduction

With the development of various imaging devices, multimodal medical image fusion has become an important research topic to obtain accurate clinical information to physicians for better diagnosis. The single modality medical image cannot fully satisfy the doctors to diagnose the patient's condition. For example, the CT image can only show high-resolution information such as bone structures and implants with less distortion while

MR image can only show normal and pathological soft tissues information such as flesh [5, 8, 84, 41]. Therefore, the technology of fusing different modality medical images into a single image has attracted many researchers attention [3].

Multimodal image fusion is generally merged in the spatial domain or transform domain [60, 47]. In the spatial domain, a variety of methods have been proposed, such as methods based on the principal component analysis (PCA) [65, 79], independent component analysis (ICA) [60]. However, these approaches are not fully suitable for application of medical image fusion since the features are sensitive to the human visual system existing in different scales [8, 95]. On the contrary, the multiscale or multiresolution analysis is more suitable for the medical fusion purpose [8, 102]. Multiscale transforms, such as Laplacian pyramid (LP) decompositions, wavelet transform (WT) and other multiscale transforms have applied to image fusion [73, 10, 74, 11, 46, 4]. Many edge-preserving filters have also been developed to prevent smoothing across structure while still smoothing texture, such as anisotropic diffusion [67], bilateral filter [75], weighted least squares filter [27], L_0-smoothing filter [88], and guided filter (GF) [33]. For the reason that the edge-preserving filters can be used to achieve multiscale decomposition like a multiscale transform and a Laplacian pyramid decomposition, many edge-preserving filters are applied to image fusion [102, 6, 47] like algorithms based on the bilateral filter (BF) and cross bilateral filter (CBF) [36, 41], algorithms based on the guided filter (GF) [44, 100]. The main weaknesses with these edge-preserving filters are that they may suffer from various problems, including "halo" artifacts, residual artifacts, "leak" problem and time-consuming shortcoming [103]. In order to solve these problems, an edge-preserving filter based on double weight average, i.e., segment graph filter (SGF), is proposed [103]. In this chapter, we will use edge-preserving decomposition based on SGF to decompose the source images into base images and detail images. In recent years, sparse representation (SR) has drawn significant interests in computer vision and image processing [63], such

as image denoising [23], face recognition [86], action recognition [30], and object tracking [97]. Yang *et al.* [92] first introduced the technology of SR into their multifocus image fusion method. As the SR can enhance performance in image fusion, SR-based fusion has emerged as a new active branch in image fusion and many researches have been proposed [92, 53, 106, 83, 61]. Zhang *et al.* [105] have pointed out that most of the SR-based image fusion methods also belong to the multiscale transform-based techniques. Further, they have noted that almost all SR-based multimodal images fusion methods are multiscale transform method [105].

In this paper, a novel method for medical image fusion based on segment graph filter and sparse representation is proposed. Making use of SGF to preserve the structure information, and the SR is utilized to improve the fusion performance. First, we apply the SGF to decompose the source multimodal medical images into base images and detail images. Second, the base images and detail images are treated separately: the base images are fused by a fusion rule based on Shannon, and the detail images are fused by the SR technology with a learned dictionary. Finally, reconstruct a fusion image by combining the fused base image and the fused detail image. Experiments on different multimodal images are conducted to validate the proposed fusion method. Comparing SGF with other structure filters demonstrates that the SGF preserves well the structure information. Performance comparison of the proposed method with the state-of-the-art methods shows the efficiency of the proposed fusion method.

4.2.2 Proposed Fusion Framework

In this section, the proposed fusion framework will be discussed in detail. Considering two perfectly registered source medical images A and B, the proposed image fusion approach consists of the following four steps:

Step 1 : Perform the SGF on the two source images A and B to obtain the base

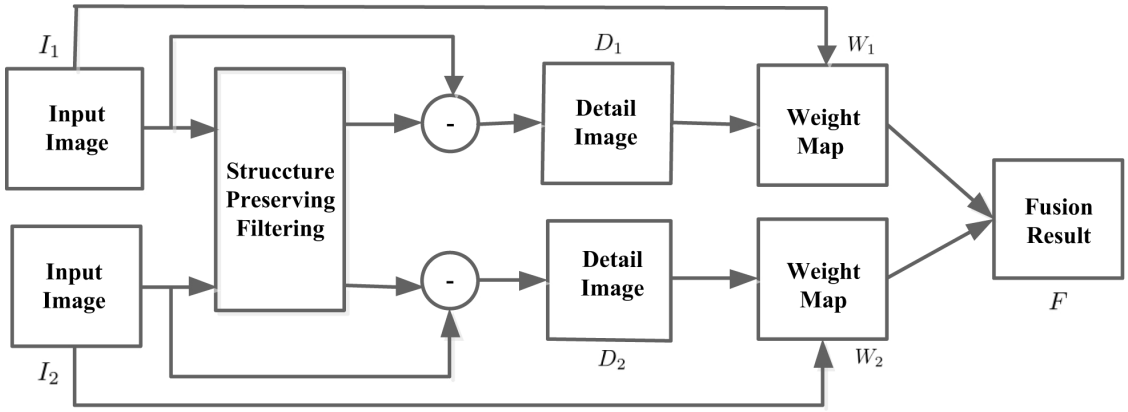


Figure 4.11: Framework for the proposed method.

images

$$A_b = SGF(A), \tag{4.8}$$

$$B_b = SGF(B). \tag{4.9}$$

And then the detail images are obtained by subtracting SGF output from the original images A and B

$$A_d = A - A_b, \tag{4.10}$$

$$B_d = B - B_b. \tag{4.11}$$

Step 2: A fusion criterion based on the activity measurement of base information is chosen to fuse the base image in the proposed method [8]. The advantage of the fusion rule is that it is well suitable for medical image and well retains the contrast in the fused image [8]. The fusion criterion is computed by as follows:

1. Use the normalized Shannon entropy to compute the activity measurement for the base image in a region R at point (x, y)

$$E_A(x, y) = \frac{1}{|R|} \sum_{i,j \in (R)} (A_b(i, j))^2 log(A_b(i, j)^2), \tag{4.12}$$

$$E_B(x,y) = \frac{1}{|R|} \sum_{i,j \in (R)} (B_b(i,j))^2 log(B_b(i,j)^2), \tag{4.13}$$

where $|R|$ is the size of the region, i.e., the total number of pixels contained in R.

2. Extract the salient information from base information of each image at a location (x, y) to obtain the corresponding weight

$$S_A(x,y) = \frac{E_A(x,y)}{E_A(x,y) + E_B(x,y)}, \tag{4.14}$$

$$S_B(x,y) = \frac{E_B(x,y)}{E_A(x,y) + E_B(x,y)}. \tag{4.15}$$

3. Fuse the base information as following

$$C_b^F(x,y) = S_A(x,y)A_b(x,y) + S_B(x,y)B_b(x,y). \tag{4.16}$$

Step 3: Apply the method based on SR with a learned dictionary to fuse detail images.

1. Divide the detail images A_d and B_d into image patches of the same size 8×8. In this step, a sliding window at a step length of a fixed number of pixels is often used to reduce block artifacts and improve robustness [105]. Suppose that there are N patches denoted as $\{p_A^i\}_{i=l}^N$ and $\{p_B^i\}_{i=l}^N$ for images A_d and B_d, respectively. And then learning a dictionary from the set of patches used by K-SVD dictionary learning algorithm [1].

2. For a pair of patch $\{p_A^i, p_B^i\}$ at position i, rearrange it into column vectors$\{v_A^i, v_B^i\}$.

3. Calculate the sparse coefficient vectors $\{\alpha_A^i, \alpha_B^i\}$ of $\{v_A^i, v_B^i\}$ using the orthogonal matching pursuit (OMP) algorithm [76]

$$\alpha_A^i = min_\alpha \|\alpha\|_0 \quad s.t. \|v_A^i - D\alpha\|_2 < \varepsilon, \tag{4.17}$$

$$\alpha_B^i = min_\alpha \|\alpha\|_0 \quad s.t. \|v_B^i - D\alpha\|_2 < \varepsilon, \tag{4.18}$$

where D is the learn dictionary used K-SVD algorithm.

4. Obtain the fused sparse vector α_F^i by using the "$max-l_1$" rule since it is appropriate for image fusion [93]

$$\alpha_F^i = \begin{cases} \alpha_A^i, & if \quad \|\alpha_A^i\|_1 > \|\alpha_B^i\|_1, \\ \alpha_B^i, & otherwise. \end{cases} \tag{4.19}$$

And then the fusion result of v_A^i and v_B^i is computed by

$$v_F^i = D\alpha_F^i. \tag{4.20}$$

Finally, the fused image D_F^i at position i can be reconstructed by v_F^i. Reshape the v_F^i into a patch p_F^i with size 8×8 and then plug it into its original position. For all the source image patches in $\{p_A^i\}_{i=1}^N$ and $\{p_B^i\}_{i=1}^N$, repeat the above process to obtain the detail fused image C_d^F.

Step 4: Once the base fused image C_b^F and detail fused image C_d^F have been obtained, the fused image F is reconstructed by

$$F = C_b^F + C_d^F. \tag{4.21}$$

According to the above description, the framework of the proposed method is represented in Fig. 4.11.

4.2.3 Experiment Results

Experimental Setting

There are three parameters that should be set in the SGF of the proposed fusion algorithm: r is the window size, σ and τ control speed of the attenuation in Eq.(2.10) and stop the aggregation from S_i and S_0 in Eq.(2.15), respectively. Under the range of the

three parameters described in the [103], we set three parameters are $r = 12$, $\sigma = 0.01$ and $\tau = 40/255$, respectively.

Performance Metrics

The fusion performances are evaluated by the four measures including feature-based metric $Q_p^{ab|f}$ [90], structure-based metric $Q_w^{xy|f}$ [94], the normalized mutual information Q_{MI} [35], and a nonlinear correlation information entropy Q_{NICE} [72, 56].

Comparing other Structure Preserving Filter

In order to validate that the segment graph filter (SGF) can preserve well structure information. We compare SGF with different edge-preserving filters, i.e., weighted least squares filter (WLS) [27], an isotropic diffusion [67], L_0-smoothing filter [88], Cross Bilateral Filter (CBF) [36, 41] and guided filter (GF) [33]. In general, therefore, we apply different edge-preserving filters into the same proposed method framework for a fair comparison. Experiments are carried on four pairs of testing images. And the images are divided into two groups. Group a contains Fig. 4.12 (a) and Fig. 4.12 (b), and they are CT and MRI images of the brain. Group b contains Fig. 4.13 (a) and Fig. 4.13 (b), and they are T1-weighted MR image (MR-T1) and MRA, respectively. Group c contains Fig. 4.14 - 4.15 (a) and Fig. 4.14 - 4.15 (b), and they are MR-T1 and MR-T2 images.

The experimental results are shown in Fig. 4.12-Fig. 4.15. The fusion results of L_0 -SR and GF-SR based method shown in Fig. 4.12 (e) - Fig. 4.15 (e) and Fig. 4.12 (g) - Fig. 4.15 (g) lead to the reduced contrast comparing with other fusion results. The fused images shown in Fig. 4.12 (c) - Fig. 4.15 (c), Fig. 4.12 (d) - Fig. 4.15 (d), Fig. 4.12 (f) - Fig. 4.15 (f) and Fig. 4.12 (g) - Fig. 4.15 (g) obtain the good fusion results in subjective assessment.

The objective evaluations of the fusion results of the test four pairs images are shown in Table. 4.4. In order to see easily, the highest values for all objective metrics have been

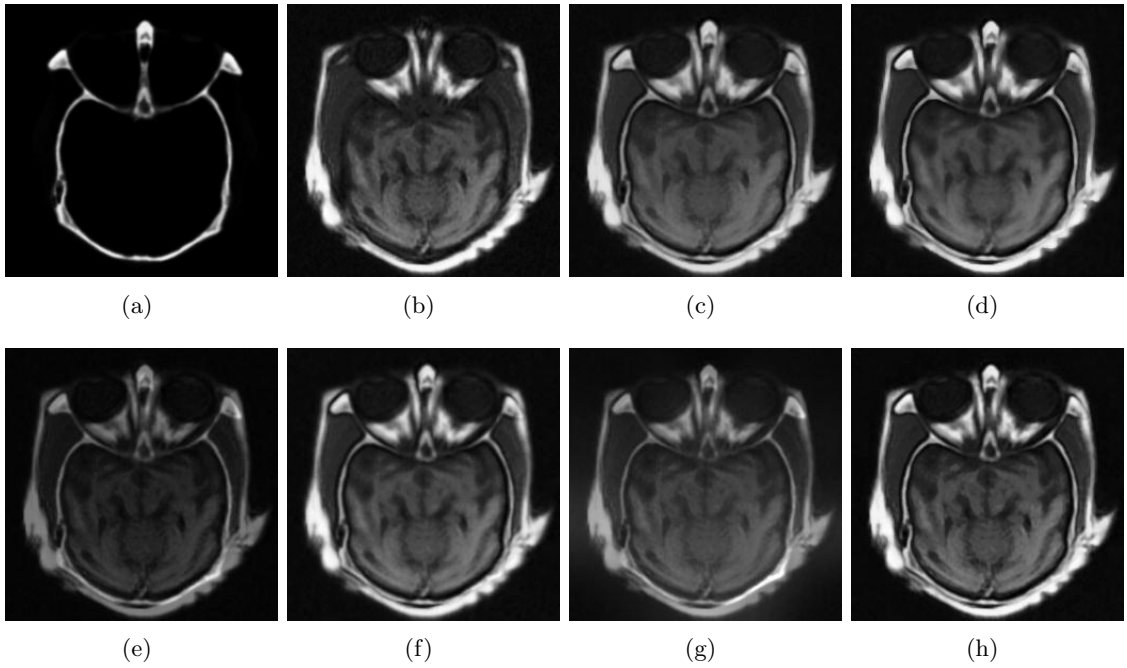
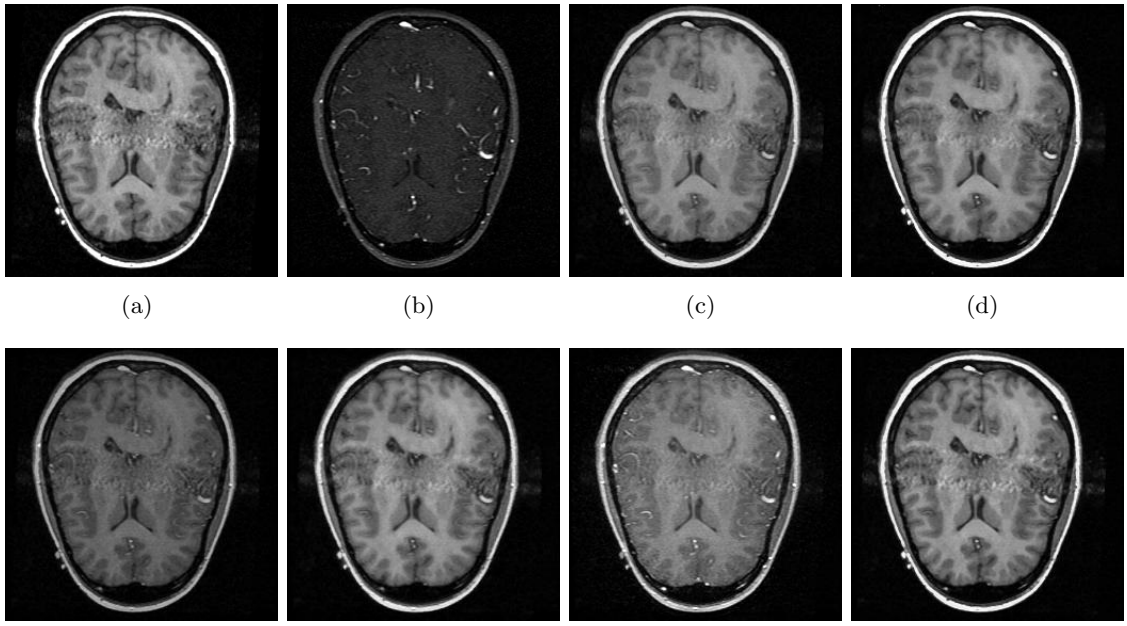


Figure 4.12: Fusion result of different structure preserving filter for image 1: (a) CT (b) MRI (c) WLS-SR (d) AD-SR (e) L0-SR (f) CBF-SR (g) GFF-SR (h) MSGF.

marked in bold. We can see clearly that the fusion result based on SGF obtain a better quantitative value than other structure-preserving filters. In a word, the SGF can keep the image structure better than other edge-preserving filters.

Compare with Different State-of-the-art Methods

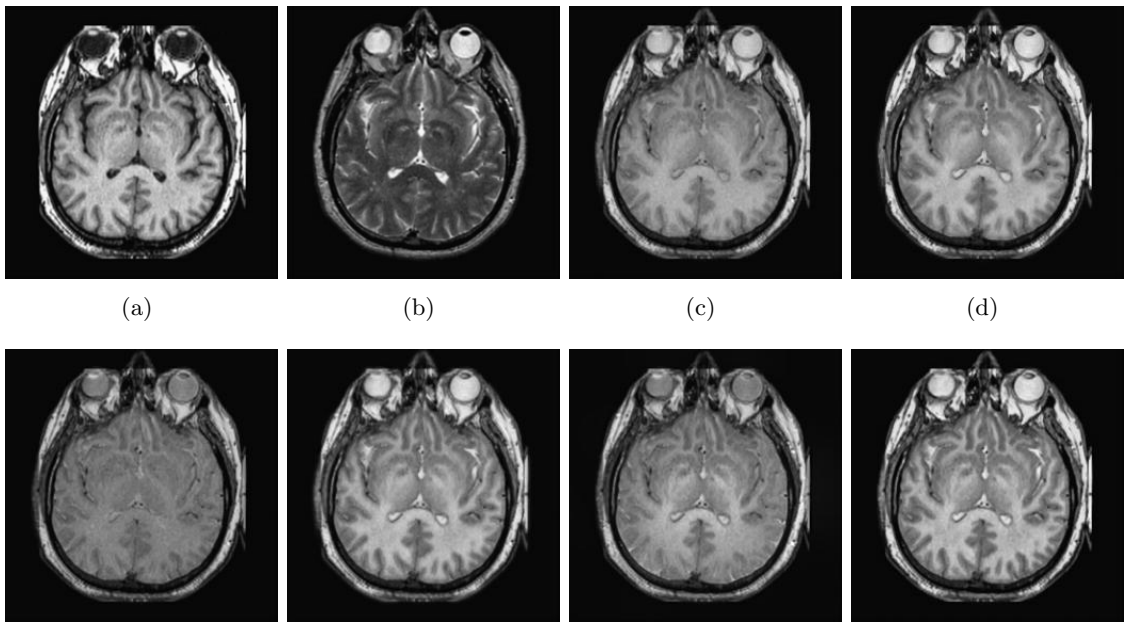
To show the effectiveness of the medical image fusion using segment graph filter (MSGF), we perform the experiment and compare with some state-of-the-art methods. Particularly, five fusion methods are chosen to compare the performance of the proposed algorithm: the fusion method based on nonsubsampled contourlet transform (NSCT) [53], image fusion method based on Cross Bilateral Filter proposed by Kumar (CBF) [41], GFF based on guided filtering proposed by Li *et al.* (GFF) [44], medical image fusion based on nonsubsampled contourlet transform and PCNN proposed by Das *et al.* (NFCD) [19] and the medical image fusion based guided image filter and image statistics (GFIS) [5].



(a) (b) (c) (d)

(e) (f) (g) (h)

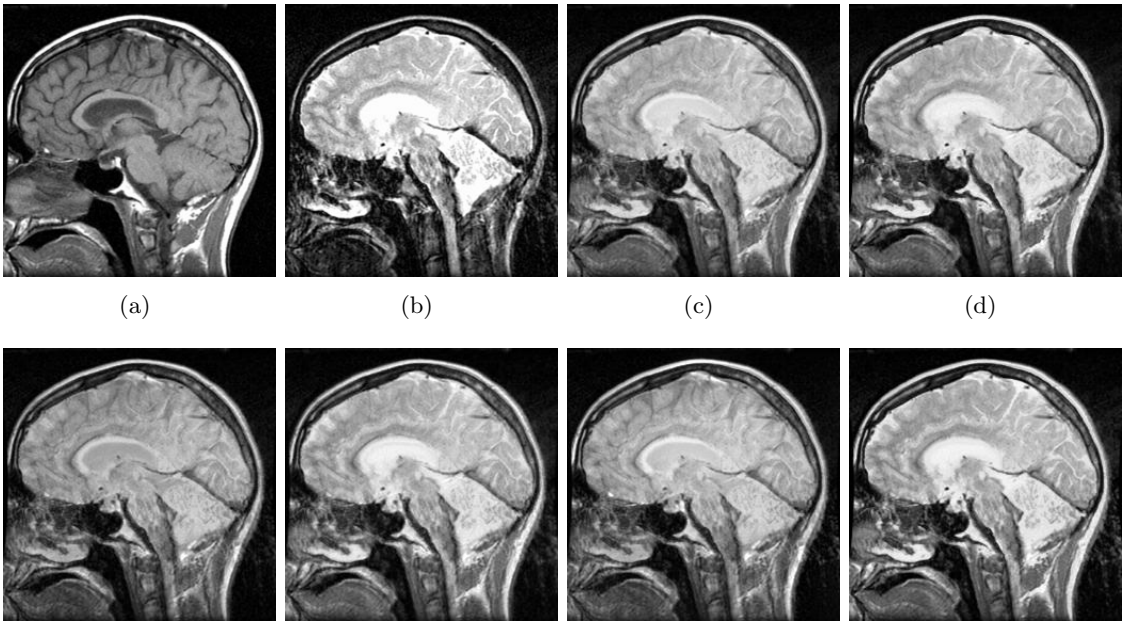
Figure 4.13: Fusion result of different structure preserving filter for image 2: (a) $MR-T_1$ (b) MRA (c) WLS-SR (d) AD-SR (e) L0-SR (f) CBF-SR (g) GFF-SR (h) MSGF.



(a) (b) (c) (d)

(e) (f) (g) (h)

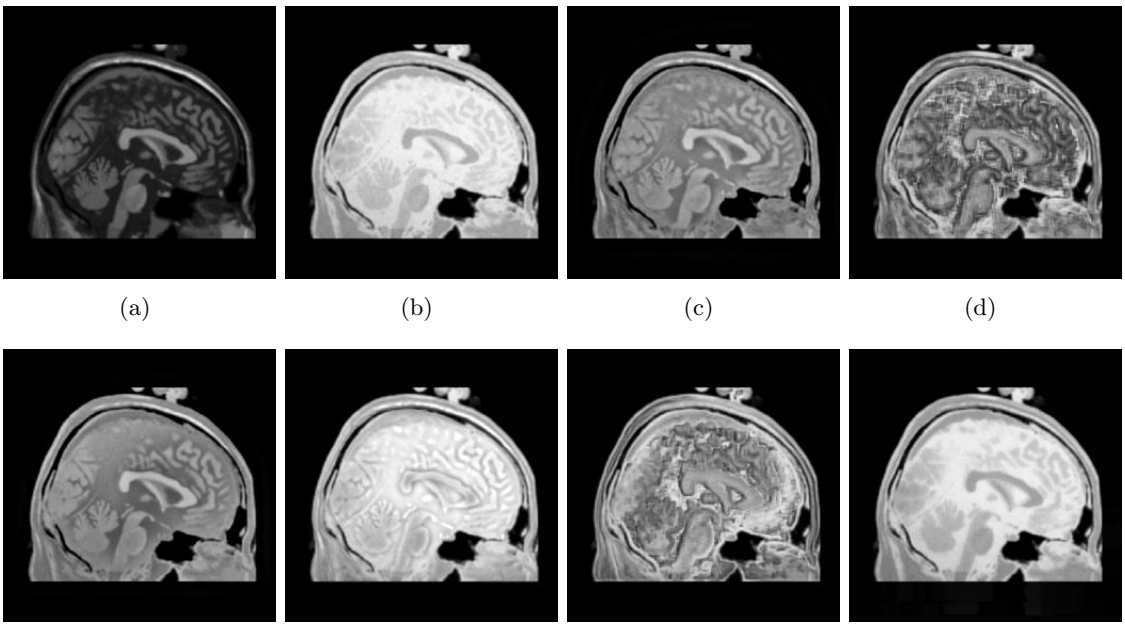
Figure 4.14: Fusion result of different structure preserving filter for image 3: (a) $MR-T_1$ (b) $MR-T_2$ (c) WLS-SR (d) AD-SR (e) L0-SR (f) CBF-SR (g) GFF-SR (h) MSGF.



(a) (b) (c) (d)

(e) (f) (g) (h)

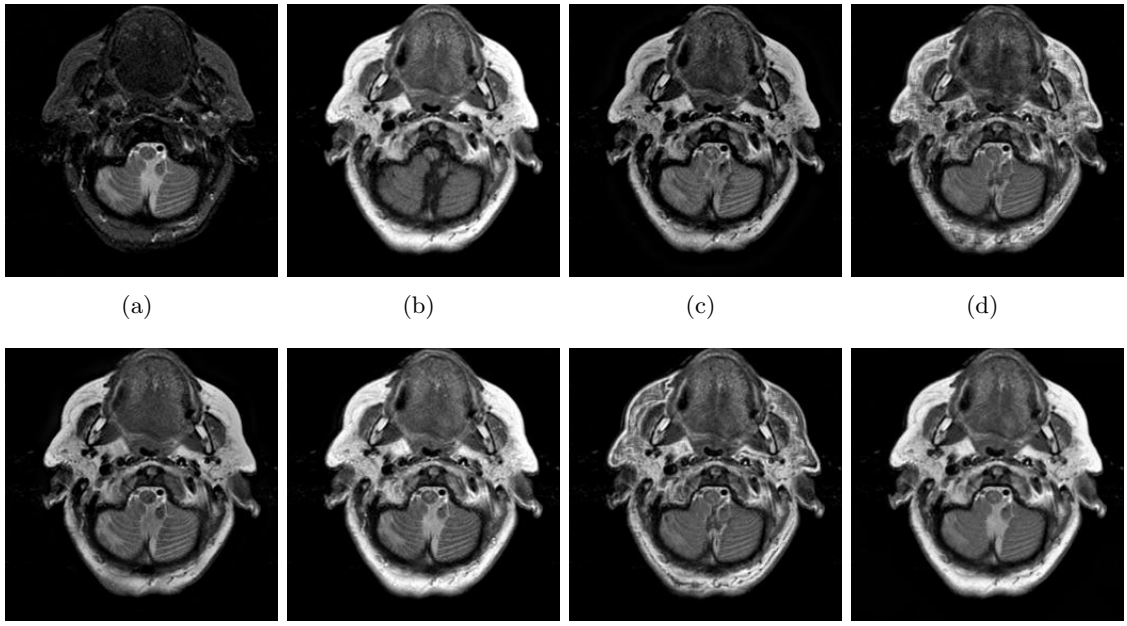
Figure 4.15: Fusion result of different structure preserving filter for image 4: (a) $MR-T_1$ (b) $MR-T_2$ (c) WLS-SR (d) AD-SR (e) L0-SR (f) CBF-SR (g) GFF-SR (h) MSGF.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4.16: Fusion result of different methods for different medical images for image 5 : (a) $MR-T_1$ (b) $MR-T_2$ (c) NSCT (d) CBF (e) GFF (f) NFCD (g) GFIS (h) MSGF.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4.17: Fusion result of different methods for different medical images for image 6: (a) $MR-T_1$ (b) $MR-T_2$ (c) NSCT (d) CBF (e) GFF (f) NFCD (g) GFIS (h) MSGF.

Table 4.4: Objective performance

Images	Metric	WLS-SR	AD-SR	L0-SR	CBF-SR	GF-SR	MSGF
Fig.4.12	$Q_p^{ab\|f}$	0.6777	0.6869	0.4144	0.7002	0.4337	**0.7510**
	$Q_p^{xy\|f}$	0.8543	0.8602	0.6472	0.8778	0.7044	**0.8954**
	Q_{MI}	0.5760	0.5925	0.9318	0.6226	0.3998	**0.7227**
	Q_{NICE}	0.8102	0.8109	0.8216	0.8119	0.8055	**0.8153**
Fig.4.13	$Q_p^{ab\|f}$	0.5686	0.5593	0.4929	0.5512	0.4753	**0.6141**
	$Q_p^{xy\|f}$	**0.8536**	0.8144	0.7227	0.8246	0.7513	0.8484
	Q_{MI}	0.7937	0.8000	1.1047	0.7783	0.7013	**0.8894**
	Q_{NICE}	0.8120	0.8121	0.8177	0.8117	0.8095	**0.8146**
Fig.4.14	$Q_p^{ab\|f}$	0.4292	0.4762	0.3459	0.4744	0.3848	**0.5463**
	$Q_p^{xy\|f}$	0.8092	0.8342	0.7258	0.8413	0.7174	**0.9154**
	Q_{MI}	0.7041	0.7213	0.6886	0.7219	0.6129	**0.7906**
	Q_{NICE}	0.8112	0.8117	0.8105	0.8118	0.8099	**0.8135**
Fig.4.15	$Q_p^{ab\|f}$	0.4477	0.4609	0.4025	0.4413	0.4231	**0.5111**
	$Q_p^{xy\|f}$	0.7870	0.7967	0.7208	0.7911	0.7321	**0.8629**
	Q_{MI}	0.5018	0.5183	0.5255	0.5118	0.4795	**0.5578**
	Q_{NICE}	0.8098	0.8104	0.8102	0.8102	0.8090	**0.8116**

For a fair comparison, we use the parameters reported by the authors to yield the best fusion results.

The experiments have been performed on two pairs of multimodal medical images as shown in Fig. 4.16 - 4.17 (a) and (b). From the fusion results of the method CBF and GFIS in Fig. 4.16 (d) - (g) and Fig. 4.17 (d) - (g), there are some artifacts in the two fusion images. Some detail information of source images isn't well integrated into the fused images. Fig. 4.16 (c) - (e) and Fig. 4.17 (c) - (e) are shown the fusion results of methods based on NSCT and GFF, and we can see clearly that the fusion results are reduced contrast and they are not well suitable for people's observation. It can be seen easily that the contrast of the fusion results of Fig. 4.16 (f) - (h) and Fig. 4.17 (f) - (h) retain better contrast than other fusion results.

The objective evaluations of the fusion results for the testing pairs images are shown in Table. 4.5. From the Table. 4.5, the fusion results of proposed method achieve the highest values (highlighted in bold) of many evaluation metrics. Therefore, the proposed method not only preserves the edge information but also achieves good fusion result when compared to existing methods.

Table 4.5: Objective performance

Images	Metric	NSCT	CBF	GFF	NFCD	GFIS	MSGF
Fig.4.16	$Q_p^{ab\|f}$	0.5905	0.4995	**0.5966**	0.5510	0.4781	0.5640
	$Q_p^{xy\|f}$	0.5966	0.7508	0.7232	0.7318	0.7287	**0.7547**
	Q_{MI}	0.6328	0.6885	0.6430	0.7263	0.6813	**0.8293**
	Q_{NICE}	0.8080	0.8087	0.8080	0.8093	0.8086	**0.8122**
Fig.4.17	$Q_p^{ab\|f}$	0.5559	0.4854	0.5795	0.5419	0.5421	**0.5925**
	$Q_p^{xy\|f}$	0.6804	0.8442	0.7755	0.8429	**0.8811**	0.7507
	Q_{MI}	0.5917	0.6908	0.6389	0.7248	0.7494	**0.7682**
	Q_{NICE}	0.8076	0.8089	0.8083	0.8098	0.8103	**0.8115**

Influence of parameters

We conduct the experiments on six pairs of images which study the influence of the parameters σ and r with respect to $Q_p^{ab|f}$ to reveal the proposed method robustness. The experiment result is shown in Fig. 4.18. As shown in Fig. 4.18, the influence of σ and r are varying from 0.1 to 0.15 with intervals 0.02 and 10 to 24 with intervals 2, respectively. From the Fig. 4.18, we can see that the proposed method is robust to different values of σ and r in a wide range.

4.2.4 Conclusion

In this paper, a new fusion algorithm which is based on SGF and SR is proposed to fuse multimodal medical image fusion. Source medical images are filtered by SGF to obtain base images and detail images. Two different rules are implemented to fuse the base images and detail images. The base images are fused using a fusion rule based on normalized Shannon entropy for retaining contrast whereas the technology SR with a learned dictionary is used to fuse detail images for extracting features from source images. The SGF can better preserve edge than other edge-preserving filters, which is verified in the experimental results. Experimental results have also shown that the proposed method based on SGF and SR obtain the fusion results of state-of-the-art fusion algorithms.

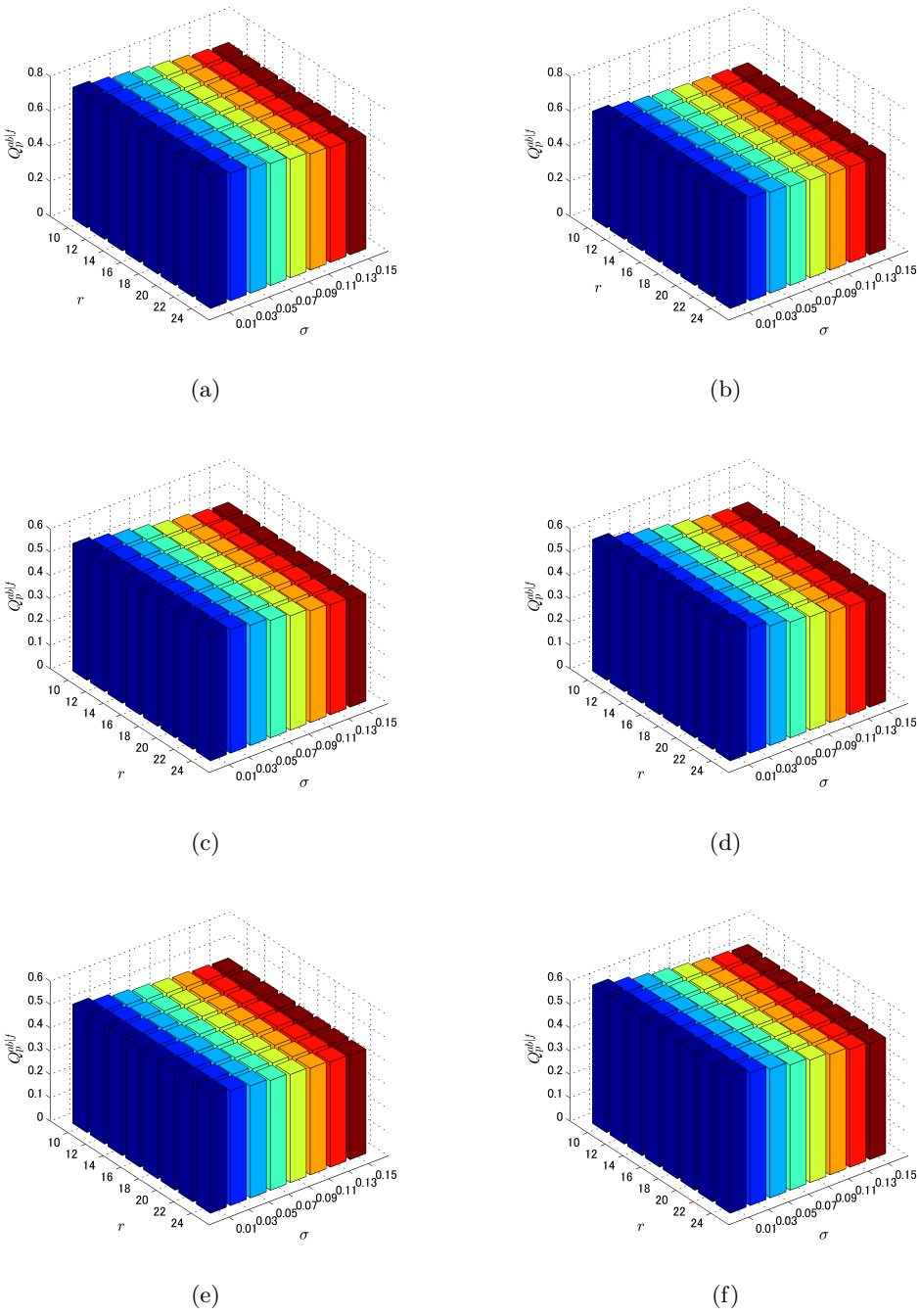


Figure 4.18: $Q_p^{ab|f}$ with respect to different r and σ for six pairs multimodal images:(a-f) are the results of input images of Fig.(4.12-4.16)

Chapter 5

Conclusions

This thesis have focused on the two aspects of image processing: image inpainting and image fusion. The similarity of image inpainting and image fusion are both to get a clear image for human visual system and machine processing. The related dictionary and histogram dictionary have been proposed for image inpainting based on sparse representation. Thesis two types of dictionary can avoid the irrelevant patches, and then the proposed image inpainting method based on the two dictionaries can get the better inpainting result. For image fusion, the structure-preserving filter has been applied to refine the focus map and preserve the structure information of the image. Specifically, the main contributions are drawn as follows.

- Two image inpainting methods based on sparse representation have been proposed in Chapter 3. Specifically, two types of dictionary, i.e., the related dictionary and histogram dictionary, have been proposed by using the similarity comparison methods based on histogram. And then based on thesis two types of dictionary, two new inpainting methods have been proposed by using sparse representation.

- A multifocus image fusion method using structure-preserving filter has been proposed in Chapter 4. The proposed fusion method is belongs to the spatial domain method, so the original intensities of source images can be preserved in fusion results.

Moreover, the proposed method also use the structure-preserving filter to smooth the detail information while preserving the structures of focus maps. And then, the refinement focus maps are obtained for image fusion. This refinement focus maps are more accurate than the original focus maps which can improve the fusion result.

- A medical image fusion using segment graph filter and sparse representation has also been proposed in Chapter 4. Different modalities are used to capture different informations. Medical image fusion is applied to fuse different modalities medical images to obtain a fused medical image for better diagnosis and treatment. The base images are extracted from each source image by using structure-preserving filter, and then detail images are obtained source images by subtracting its base images. The base images and the detail images are fused by using different fusion rules. The fused image is obtained by merging the fused base image and the fused detail image.

Bibliography

[1] Michal Aharon, Michael Elad, Alfred Bruckstein, et al. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311, 2006.

[2] Bruno Alfano, Mario Ciampi, and Giuseppe De Pietro. A wavelet-based algorithm for multimodal medical image fusion. In *International Conference on Semantic and Digital Media Technologies*, pages 117–120. Springer, 2007.

[3] Vincent Barra and Jean-Yves Boire. A general framework for the fusion of anatomical and functional medical images. *NeuroImage*, 13(3):410–424, 2001.

[4] Durga Prasad Bavirisetti and Ravindra Dhuli. Multi-focus image fusion using multi-scale image decomposition and saliency detection. *Ain Shams Engineering Journal*, 2016.

[5] Durga Prasad Bavirisetti, Vijayakumar Kollu, Xiao Gang, and Ravindra Dhuli. Fusion of mri and ct images using guided image filter and image statistics. *International Journal of Imaging Systems & Technology*, 27(3):227C237, 2017.

[6] Durga Prasad Bavirisetti, Gang Xiao, and Gang Liu. Multi-sensor image fusion based on fourth order partial differential equations. In *International Conference on Information Fusion*, 2017.

[7] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424. ACM Press/Addison-Wesley Publishing Co., 2000.

[8] Gaurav Bhatnagar, QM Jonathan Wu, and Zheng Liu. A new contrast based multimodal medical image fusion framework. *Neurocomputing*, 157:143–152, 2015.

[9] Aurélie Bugeau, Marcelo Bertalmío, Vicent Caselles, and Guillermo Sapiro. A comprehensive framework for image inpainting. *IEEE Transactions on Image Processing*, 19(10):2634–2645, 2010.

[10] Peter J Burt and Edward H Adelson. Merging images through pattern decomposition. In *Applications of Digital Image Processing VIII*, volume 575, pages 173–182. International Society for Optics and Photonics, 1985.

[11] Peter J Burt and Raymond J Kolczynski. Enhanced image capture through fusion. In *Computer Vision, 1993. Proceedings., Fourth International Conference on*, pages 173–182. IEEE, 1993.

[12] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[13] T Chan and J Shen. Local inpainting models and tv inpainting. *SIAM J. Appl. Math*, 62(3):1019–1043, 2001.

[14] Wen Chen and Xudong Chen. Focal-plane detection and object reconstruction in the noninterferometric phase imaging. *JOSA A*, 29(4):585–592, 2012.

[15] Yin Chen and Rick S Blum. A new automated quality assessment algorithm for image fusion. *Image and vision computing*, 27(10):1421–1432, 2009.

[16] Antonio Criminisi, Patrick Perez, and Kentaro Toyama. Object removal by exemplar-based inpainting. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE, 2003.

[17] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on image processing*, 13(9):1200–1212, 2004.

[18] Sudeb Das and Malay Kumar Kundu. Nsct-based multimodal medical image fusion using pulse-coupled neural network and modified spatial frequency. *Medical & biological engineering & computing*, 50(10):1105–1114, 2012.

[19] Sudeb Das and Malay Kumar Kundu. A neuro-fuzzy approach for medical image fusion. *IEEE transactions on biomedical engineering*, 60(12):3347–3353, 2013.

[20] Liang-Jian Deng, Ting-Zhu Huang, and Xi-Le Zhao. Exemplar-based image inpainting using a modified priority definition. *PloS one*, 10(10):e0141199, 2015.

[21] David L Donoho and Yaakov Tsaig. Fast solution of -norm minimization problems when the solution may be sparse. *IEEE Transactions on Information Theory*, 54(11):4789–4812, 2008.

[22] Alexei A Efros and Thomas K Leung. Texture synthesis by non-parametric sampling. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1033–1038. IEEE, 1999.

[23] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.

[24] Michael Elad, J-L Starck, Philippe Querre, and David L Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Applied and Computational Harmonic Analysis*, 19(3):340–358, 2005.

[25] Gilberte Emile-Male. *The restorer's handbook of easel painting*, volume 31. Van Nostrand Reinhold New York, 1976.

[26] Mohamed-Jalal Fadili, J-L Starck, and Fionn Murtagh. Inpainting and zooming using sparse representations. *The Computer Journal*, 52(1):64–79, 2007.

[27] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)*, volume 27, page 67. ACM, 2008.

[28] Muhammad Shahid Farid, Arif Mahmood, and Somaya Ali Al-Maadeed. Multi-focus image fusion using content adaptive blurring. *Information Fusion*, 45:96–112, 2019.

[29] Eduardo SL Gastal and Manuel M Oliveira. Domain transform for edge-aware image and video processing. In *ACM Transactions on Graphics (ToG)*, volume 30, page 69. ACM, 2011.

[30] Tanaya Guha and Rabab K Ward. Learning sparse representations for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(8):1576–1588, 2012.

[31] Christine Guillemot and Olivier Le Meur. Image inpainting: Overview and recent advances. *IEEE signal processing magazine*, 31(1):127–144, 2014.

[32] Mohammad Bagher Akbari Haghighat, Ali Aghagolzadeh, and Hadi Seyedarabi. Multi-focus image fusion for visual sensor networks in dct domain. *Computers & Electrical Engineering*, 37(5):789–797, 2011.

[33] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. In *European conference on computer vision*, pages 1–14. Springer, 2010.

[34] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369. IEEE, 2010.

[35] M Hossny, S Nahavandi, and D Creighton. Comments on'information measure for performance of image fusion'. *Electronics letters*, 44(18):1066–1067, 2008.

[36] Jianwen Hu and Shutao Li. The multiscale directional bilateral filter and its application to multisensor image fusion. *Information Fusion*, 13(3):196–206, 2012.

[37] Xiaoli Huan, Beddhu Murali, and Adel L Ali. Image restoration based on the fast marching method and block based sampling. *Computer Vision and Image Understanding*, 114(8):847–856, 2010.

[38] Rizwan Ahmed Khan, Hubert Konik, and Éric Dinet. Enhanced image saliency model based on blur identification. pages 1–7, 2010.

[39] Minjae Kim, David K Han, and Hanseok Ko. Joint patch clustering-based dictionary learning for multimodal image fusion. *Information fusion*, 27:198–214, 2016.

[40] BK Shreyamsha Kumar. Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform. *Signal, Image and Video Processing*, 7(6):1125–1143, 2013.

[41] BK Shreyamsha Kumar. Image fusion based on pixel significance using cross bilateral filter. *Signal, image and video processing*, 9(5):1193–1204, 2015.

[42] John J Lewis, Robert J O' Callaghan, Stavri G Nikolov, David R Bull, and Nishan Canagarajah. Pixel-and region-based image fusion with complex wavelets. *Information fusion*, 8(2):119–130, 2007.

[43] Shutao Li, Xudong Kang, Leyuan Fang, Jianwen Hu, and Haitao Yin. Pixel-level image fusion: A survey of the state of the art. *Information Fusion*, 33:100–112, 2017.

[44] Shutao Li, Xudong Kang, and Jianwen Hu. Image fusion with guided filtering. *IEEE Transactions on Image Processing*, 22(7):2864–2875, 2013.

[45] Shutao Li and Bin Yang. Multifocus image fusion using region segmentation and spatial frequency. *Image and vision computing*, 26(7):971–979, 2008.

[46] Tianjie Li and Yuanyuan Wang. Biological image fusion using a nsct based variable-weight method. *Information Fusion*, 12(2):85–92, 2011.

[47] Wen Li, Yuange Xie, Haole Zhou, Ying Han, and Kun Zhan. Structure-aware image fusion. *Optik*, 2018.

[48] Lin Liang, Ce Liu, Ying-Qing Xu, Baining Guo, and Heung-Yeung Shum. Real-time texture synthesis by patch-based sampling. *ACM Transactions on Graphics (ToG)*, 20(3):127–150, 2001.

[49] Shuaiqi Liu, Tao Zhang, Hailiang Li, Jie Zhao, and Huiya Li. Medical image fusion based on nuclear norm minimization. *International Journal of Imaging Systems & Technology*, 25(4):310–316, 2016.

[50] Shuaiqi Liu, Jie Zhao, and Mingzhu Shi. Medical image fusion based on improved sum-modified-laplacian. *International Journal of Imaging Systems and Technology*, 25(3):206–212, 2015.

[51] Wei Liu, Xiaogang Chen, Chunhua Shen, Jingyi Yu, Qiang Wu, and Jie Yang. Robust guided image filtering. *arXiv preprint arXiv:1703.09379*, 2017.

[52] Yu Liu, Shuping Liu, and Zengfu Wang. Medical image fusion by combining non-subsampled contourlet transform and sparse representation. In *Chinese Conference on Pattern Recognition*, pages 372–381. Springer, 2014.

[53] Yu Liu, Shuping Liu, and Zengfu Wang. A general framework for image fusion based on multi-scale transform and sparse representation. *Information Fusion*, 24:147–164, 2015.

[54] Yu Liu and Zengfu Wang. Multi-focus image fusion based on wavelet transform and adaptive block. *Journal of image and graphics*, 18(11):1435–1444, 2013.

[55] Yu Liu and Zengfu Wang. Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Processing*, 9(5):347–357, 2014.

[56] Zheng Liu, Erik Blasch, Zhiyun Xue, Jiying Zhao, Robert Laganiere, and Wei Wu. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. *IEEE transactions on pattern analysis and machine intelligence*, 34(1):94–109, 2012.

[57] Julien Mairal, Michael Elad, and Guillermo Sapiro. Sparse representation for color image restoration. *IEEE Transactions on image processing*, 17(1):53–69, 2008.

[58] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, 41(12):3397–3415, 1993.

[59] Simon Masnou and J-M Morel. Level lines based disocclusion. In *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269)*, pages 259–263. IEEE, 1998.

[60] Nikolaos Mitianoudis and Tania Stathaki. Pixel-based and region-based image fusion schemes using ica bases. *Information Fusion*, 8(2):131–142, 2007.

[61] Mansour Nejati, Shadrokh Samavi, and Shahram Shirani. Multi-focus image fusion using dictionary-based sparse representation. *Information Fusion*, 25:72–84, 2015.

[62] Filippo Nencini, Andrea Garzelli, Stefano Baronti, and Luciano Alparone. Remote sensing image fusion using the curvelet transform. *Information fusion*, 8(2):143–156, 2007.

[63] Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607, 1996.

[64] Vishal M Patel and Rama Chellappa. Sparse representations, compressive sensing and dictionaries for pattern recognition. In *Pattern Recognition (ACPR), 2011 First Asian Conference on*, pages 325–329. IEEE, 2011.

[65] Ujwala Patil and Uma Mudengudi. Image fusion using hierarchical pca. In *image Information Processing (ICIIP), 2011 International Conference on*, pages 1–6. IEEE, 2011.

[66] Sujoy Paul, Ioana S Sevcenco, and Panajotis Agathoklis. Multi-exposure and multi-focus image fusion in gradient domain. *Journal of Circuits, Systems and Computers*, 25(10):1650123, 2016.

[67] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639, 1990.

[68] Gemma Piella. A general framework for multiresolution image fusion: from pixels to regions. *Information fusion*, 4(4):259–280, 2003.

[69] Ron Rubinstein, Alfred M Bruckstein, and Michael Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010.

[70] Bin Shen, Wei Hu, Yimin Zhang, and Yu-Jin Zhang. Image inpainting via sparse representation. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 697–700. IEEE, 2009.

[71] Julius Orion Smith. *Introduction to digital filters: with audio applications*, volume 2. Julius Smith, 2007.

[72] Tania Stathaki. *Image fusion: algorithms and applications*. Elsevier, 2011.

[73] Kumaravel Subramaniam Tamilselvan and Govindasamy Murugesan. Survey and analysis of various image fusion techniques for clinical ct and mri images. *International Journal of Imaging Systems and Technology*, 24(2):193–202, 2014.

[74] Alexander Toet, Lodewik J Van Ruyven, and J Mathee Valeton. Merging thermal and visual images by a contrast pyramid. *Optical engineering*, 28(7):287789, 1989.

[75] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision*, pages 839–846. IEEE, 1998.

[76] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*, 53(12):4655–4666, 2007.

[77] Joel A Tropp, Anna C Gilbert, and Martin J Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.

[78] Sarah Walden. *The ravished image: or how to ruin masterpieces by restoration*. St.Martin's Press, New York, 1985.

[79] Tao Wan, Chenchen Zhu, and Zengchang Qin. Multifocus image fusion based on

robust principal component analysis. *Pattern Recognition Letters*, 34(9):1001–1008, 2013.

[80] Qiang Wang, Yi Shen, and Jing Jin. *Performance evaluation of image fusion techniques*. 2008.

[81] Zhou Wang and Alan C Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117, 2009.

[82] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[83] Qi Wei, José Bioucas-Dias, Nicolas Dobigeon, and Jean-Yves Tourneret. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7):3658–3668, 2015.

[84] Alexander Wong and William Bishop. Efficient least squares fusion of mri and ct images using a phase congruency model. *Pattern Recognition Letters*, 29(3):173–180, 2008.

[85] Alexander Wong and Jeff Orchard. A nonlocal-means approach to exemplar-based inpainting. In *2008 15th IEEE International Conference on Image Processing*, pages 2600–2603. IEEE, 2008.

[86] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):210–227, 2009.

[87] Kai Xu, Nannan Wang, and Xinbo Gao. Image inpainting based on sparse representation with dictionary pre-clustering. In *Chinese Conference on Pattern Recognition*, pages 245–258. Springer, 2016.

[88] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia. Image smoothing via l 0 gradient minimization. In *ACM Transactions on Graphics (TOG)*, volume 30, page 174. ACM, 2011.

[89] Zongben Xu and Jian Sun. Image inpainting by patch propagation using patch sparsity. *IEEE transactions on image processing*, 19(5):1153–1165, 2010.

[90] CS Xydeas, , and V Petrovic. Objective image fusion performance measure. *Electronics letters*, 36(4):308–309, 2000.

[91] Mehrdad Yaghoobi, Di Wu, and Mike E Davies. Fast non-negative orthogonal matching pursuit. *IEEE Signal Processing Letters*, 22(9):1229–1233, 2015.

[92] Bin Yang and Shutao Li. Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation and Measurement*, 59(4):884–892, 2010.

[93] Bin Yang and Shutao Li. Pixel-level image fusion with simultaneous orthogonal matching pursuit. *Information fusion*, 13(1):10–19, 2012.

[94] Cui Yang, Jian-Qi Zhang, Xiao-Rui Wang, and Xin Liu. A novel similarity based quality metric for image fusion. *Information Fusion*, 9(2):156–160, 2008.

[95] L Yang, BL Guo, and W Ni. Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform. *Neurocomputing*, 72(1-3):203–211, 2008.

[96] Ming Yin, Xiaoning Liu, Yu Liu, and Xun Chen. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. *IEEE Transactions on Instrumentation and Measurement*, (99):1–16, 2018.

[97] Xiao-Tong Yuan, Xiaobai Liu, and Shuicheng Yan. Visual classification with multitask joint sparse representation. *IEEE Transactions on Image Processing*, 21(10):4349–4360, 2012.

[98] Kun Zhan, Qiaoqiao Li, Jicai Teng, Mingying Wang, and Jinhui Shi. Multifocus image fusion using phase congruency. *Journal of Electronic Imaging*, 24(3):033014, 2015.

[99] Kun Zhan, Qiaoqiao Li, Jicai Teng, Mingying Wang, and Jinhui Shi. Multifocus image fusion using phase congruency. *Journal of Electronic Imaging*, 24(3):033014, 2015.

[100] Kun Zhan, Jicai Teng, Qiaoqiao Li, Jinhui Shi, et al. A novel explicit multi-focus image fusion method. *Journal of Information Hiding and Multimedia Signal Processing*, 6(3):600–612, 2015.

[101] Kun Zhan, Haibo Wang, Yuange Xie, Chutong Zhang, and Yufang Min. Albedo recovery for hyperspectral image classification. *Journal of Electronic Imaging*, 26(4):043010, 2017.

[102] Kun Zhan, Yuange Xie, Haibo Wang, and Yufang Min. Fast filtering image fusion. *Journal of Electronic Imaging*, 26(6):063004, 2017.

[103] Feihu Zhang, Longquan Dai, Shiming Xiang, and Xiaopeng Zhang. Segment graph based image filtering: fast structure-preserving smoothing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 361–369, 2015.

[104] Qiang Zhang and Bao-long Guo. Multifocus image fusion using the nonsubsampled contourlet transform. *Signal processing*, 89(7):1334–1346, 2009.

[105] Qiang Zhang, Yi Liu, Rick S Blum, Jungong Han, and Dacheng Tao. Sparse repre-

sentation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Information Fusion*, 40:57–75, 2018.

[106] Qiong Zhang and Xavier Maldague. An adaptive fusion approach for infrared and visible images based on nsct and compressed sensing. *Infrared Physics & Technology*, 74:11–20, 2016.

[107] Xianquan Zhang, Feng Ding, Zhenjun Tang, and Chunqiang Yu. Salt and pepper noise removal with image inpainting. *AEU-International Journal of Electronics and Communications*, 69(1):307–313, 2015.

[108] Xiaoli Zhang, Xiongfei Li, and Yuncong Feng. A new multifocus image fusion based on spectrum comparison. *Signal Processing*, 123:127–142, 2016.

[109] Yu Zhang, Xiangzhi Bai, and Tao Wang. Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure. *Information fusion*, 35:81–101, 2017.

[110] Jufeng Zhao, Huajun Feng, Zhihai Xu, Qi Li, and Tao Liu. Detail enhanced multi-source fusion using visual weight map extraction based on multi scale edge preserving decomposition. *Optics Communications*, 287:45–52, 2013.

[111] Yufeng Zheng, Edward A Essock, Bruce C Hansen, and Andrew M Haun. A new metric based on extended spatial frequency and its application to dwt based fusion algorithms. *Information Fusion*, 8(2):177–192, 2007.

[112] Jing-jing Zong and Tian-shuang Qiu. Medical image fusion based on sparse representation of classified image patches. *Biomedical Signal Processing and Control*, 34:195–205, 2017.

Publication List

Journal Papers:

1. **Qiaoqiao Li** *, Guoyue Chen, Xingguo Zhang, Kazuki Saruta, Yuki Terata. (2019) Structure Extraction with Total Variation for Hyperspectral Image Classification. IEEE ACCESS, 2019, 7: 91019-91033.

2. **Qiaoqiao Li**, Guoyue Chen*, Kun Zhan Xingguo Zhang, Kazuki Saruta, Yuki Terata. (2019). Multi-focus Image Fusion using Structure-Preserving Filter. Journal of Electronic Imaging. 28(2), 023005. (SCI)

3. **Qiaoqiao Li***, Guoyue Chen, Xingguo Zhang, Kazuki Saruta, Yuki Terata. (2018). Image Inpainting Based on Related Dictionary Constructed by Histogram. International Journal of Machine Learning and Computing. 8(5), 477-482.

4. **Qiaoqiao Li***, Guoyue Chen, Xingguo Zhang, Kazuki Saruta, Yuki Terata. (2018). Image Inpainting Based on Sparse Representation with Histogram Dictionary. Journal of Computers. 13(10), 1145-1155.

5. **Qiaoqiao Li**, Guoyue Chen*, Xingguo Zhang, Kazuki Saruta, Yuki Terata. (2019) Medical image fusion using segment graph filter and sparse representation. Optik - International Journal for Light and Electron Optics. (SCI, Under Review)

International Conference Papers:

1. **Qiaoqiao Li***, Guoyue Chen, Xingguo Zhang, Kazuki Saruta, Yuki Terata. (2018). MRI and CT image fusion based structure-preserving filter.8th International Conference On Computer Science, Engineering And Applications (ICCSEA 2018)

2. Zhan, Kun and Shi, Jinhui and, **Qiaoqiao Li** and Teng Jicai and Wang, Mingying. (2015). Image segmentation using fast linking SCM Neural Networks (IJCNN), 2015 International Joint Conference onC (pp.1-8). (EI)