Multi-focus image fusion based on the neighbor distance

Hengjun Zhao, Zhaowei Shang, Yuan Yan Tang, Bin Fang

Abstract

The effective measurement of pixel's sharpness is a key factor in multi-focus image fusion. In this paper, a gray image is considered as a two-dimensional surface, and the neighbor distance deduced from the oriented distance in differential geometry is used as a measure of pixel's sharpness, where the smooth image surface is restored by kernel regression. Based on the deduced neighbor distance filter, we construct a multi-scale image analysis framework, and propose a multi-focus image fusion method based on the neighbor distance. The experiments demonstrate that the proposed method is superior to the conventional image fusion methods in terms of some objective evaluation indexes, such as spatial frequency, standard deviation, average gradient, etc.

1. Introduction

Due to the limited depth-of-focus of optical lenses in imaging camera, it is impossible to capture an image in which all containing objects appear sharp. Only the objects within the depth of field are sharp, while other objects are blurred. A popular way to get an image with every object in focus is image fusion, in which one can acquire a series of pictures with different focus settings and fuse them to create a new and improved image that contains a better description of the scene than any of the individual source images. Up to now, image fusion has been successfully applied to many fields, such as military affairs, medical imaging, remote sensing, digital camera, and so on [1].

As we know, the objective of multi-focus image fusion is to produce an image that contains all relevant objects in focus by extracting and synthesizing the focused objects of source images. The basic assumption of the multi-focus image fusion is that the focused object is sharper than the unfocused object, and the sharpness is linked to some easily computed information measures. During the last decade, a number of sharpness measures for multi-focus image fusion have been proposed. Basically, these measures can be categorized into two categories. The first category is the spatial domain-based measures, which directly estimate the sharpness by intensity values. The other category is the frequency domain-based measures under the assumption that the sharpness can be indicated by the high frequency sub-band coefficients of the source image's multi-scale decomposition.

The commonly used spatial domain-based measures include [2]: variance, energy of image gradient (EOG), Tenenbaum's algorithm, energy of lap (EOL), sum-modified-Laplacian (SML) and spatial frequency (SF). Huang and Jing [2] assessed these measures according to some objective standards. Their experiment results show that SML and EOL can provide better performance than other sharpness measures. The common scheme of multi-focus image fusion methods with the spatial domain-based sharpness measures is the block-based fusion scheme, which first divide the source images into blocks or regions, second compute the block's sharpness, and finally select the sharper blocks from source images by some selecting method, such as pulse-coupled neural network [3], artificial neural network [4], genetic algorithm [5], support vector machine [6] or simply copy the sharper blocks into the fused image [2]. In those methods, the sharpness's efficiency strongly depends on the block's size and segmented algorithm. If a divided block is partly clear and partly blurry, its sharpness is not precise and may produce block effects because the blurry part may be selected as part of the fused image when considering the integrity of the segmented part. The block effects will significantly compromise the quality of the fused Image. With a multi-resolution transform, an image can be decomposed into low frequency sub-band coefficients and high frequency sub-band coefficients. By the construction of the multi-resolution decomposition, the high frequency sub-band coefficients indirectly represent the gray value variation between a pixel and its neighbor. Hence, the high frequency sub-band coefficients can also indicate the pixel's sharpness. The basic scheme of multi-focus image fusion methods based on the multi-resolution
transform is to perform a multi-resolution decomposition on each source image, then integrates all these decompositions to form a composite representation, and finally reconstructs the fused image by performing an inverse multi-resolution transform. Relative to the block-based fusion methods, the multi-resolution transform-based fusion methods can successfully overcome the block effects mentioned above, because the high frequency coefficients are selected out to compose fused image, not pixels or blocks in spatial domain. The commonly used multi-scale decomposition transforms for image fusion include Laplacian pyramid (LAP) [7], filter subtract decimate hierarchical pyramid (FSD) [8], gradient pyramid (GRP) [9], ratio of low-pass pyramid (RAP) [10]. Due to the advantages over the pyramid transform, such as localization and direction, the discrete wavelet transform (DWT)-based fusion methods [11] are generally superior to the pyramid-based fusion methods. However, because of the underlying down-sampling process, the DWT is shift-variant. And consequently, the DWT-based fusion methods are also shift-dependent, which are undesirable since different fusion results are obtained once input images are mis-registered [12]. To overcome this disadvantage of DWT, Hill et al. [13] introduced the shift-invariant and directionally selection dual tree complex wavelet transform to image fusion. Zhang et al. [12] proposed a fusion algorithm based on the shiftable complex directional pyramid transform. Besides the above spatial and frequency domain-based clarity measures, Sheng et al. [14] used the support value of support vector machine to measure the clarity of multi-focus image, and proposed a support value transform (SVT)-based image fusion method.

With the two spatial coordinates forming two of three dimensions and the gray value forming the third dimension, a gray image can be seen as a two-dimensional surface called as a gray image surface or image surface in the three-dimensional space. Fig. 1 shows two common test images with different focus settings, where the image (a) focuses on the big clock, and the image (b) focuses on the small clock. The image surfaces in image (a) and (b) correspond to the same block belonging to the big clock. As shown in Fig. 1, when the big clock is focused, the selected block is clearer and its corresponding surface is sharper; when the big clock is unfocused, the selected block is blurrier and its corresponding surface is flatter. Hence, the degree of surface curvature is related to the clarity of image block.

In differential geometry, the surface curvature—Gaussian curvature, mean curvature [15]—measure how much the surface curves at a particular point. The surface curvature has been used as a measure of image texture [16]. But the high computational complexity and the large value range (from $-\infty$ to $+\infty$) of curvature restrict its other applications in image processing. In differential geometry, the oriented distance between a point and its neighbor shown in Fig. 2 can measure how the surface is bending in the given direction $(\Delta u, \Delta v)$ at a particular point $(u, v)$. When $(\Delta u, \Delta v)$ are given, the larger the oriented distance, the more curving the surface in the direction $(\Delta u, \Delta v)$ at $(u, v)$. Therefore, it is rational to use the oriented distance to measure the image surface curvature at a point, and consequently, the oriented distance is suitable to measure the pixel’s clarity.

In an image, a pixel surrounded by eight pixels has eight neighbor directions, then the corresponding point in image surface has eight oriented distances. In this work, we name the sum of eight oriented distances as neighbor distance, and use it to measure the clarity of image pixels with different focus settings. Therefore, we can extract the focused objects of source images and integrate them into the fused image. The rest of this work is organized as follows. In Section 2, we introduce the definition of oriented distance and neighbor distance. The neighbor distance filter is also deduced in this section. In Section 3, we propose a neighbor distance transform, and describe a multi-focus image fusion method based on the neighbor distance transform. The experiment results are given in Section 4. Finally, conclusions are drawn in Section 5.

2. Neighbor distance and neighbor distance filter

This section is organized as follows. In the first section, we introduce the oriented distance and the neighbor distance defined on a smooth surface. Because the oriented distance is a property...
of smooth surface and the digit gray image is sampled from smooth image surface, we restore the smooth image surface by the non-parametric regression in the second subsection. The neighbor distance filter is also deduced in this section.

2.1. Oriented distance and neighbor distance

Suppose surface S is a smooth surface, and let \( S = \mathbb{Z}(u,v) \), then the oriented distance (OD) [17] from point \((u,v)\) to point \((u+\Delta u,v+\Delta v)\) is

\[
OD(u,v, (u+\Delta u,v+\Delta v)) = \frac{1}{2} L(u,v) \Delta u \Delta v + 2M(u,v) \Delta u \Delta v + N(u,v) \Delta v \Delta v
\]

where \( L(u,v), M(u,v), N(u,v) \) are, respectively,

\[
L(u,v) = \frac{Z_{uu}(u,v)}{\sqrt{1 + Z_{uu}^2(u,v) + Z_{uv}^2(u,v)}}
\]

\[
M(u,v) = \frac{Z_{uv}(u,v)}{\sqrt{1 + Z_{uu}^2(u,v) + Z_{uv}^2(u,v)}}
\]

\[
N(u,v) = \frac{Z_{vv}(u,v)}{\sqrt{1 + Z_{uu}^2(u,v) + Z_{uv}^2(u,v)}}
\]

The oriented distance is the distance from point \((u+\Delta u,v+\Delta v)\) to the tangent plane at point \((u,v)\). It indirectly measures the variation between the two points function value. As shown in Fig. 2, with a fixed \((\Delta u,\Delta v)\), the more curving the smooth surface S at point \((u,v)\), the larger the oriented distance from point \((u,v)\) to point \((u+\Delta u,v+\Delta v)\). So the oriented distance can describe how much the surface is bending in the given direction \((\Delta u,\Delta v)\) at the particular point \((u,v)\).

Definition. Let \((\{u_k,v_k\})_N\) is the set of all neighbor points around point \((u,v)\) on a smooth surface S, then the neighbor distance of point \((u,v)\) is the sum of oriented distances from \((u,v)\) to \((u_k,v_k)\), that is

\[
ND(u,v) = \sum_{k=1}^{N} OD(u,v,(u_k,v_k))
\]

If we consider a gray image as a smooth surface in three-dimensional space, then we can get the neighbor distance on the image with the definition of neighbor distance on the smooth surface. Denote the eight neighbors \((\{u_k,v_k\})_8 = \{i-j, i, i+j\} : i, j = 0, \pm 1\). With Eq. (3), the neighbor distance of point \((u_0,v_0)\) on image surface can be written as

\[
ND(u_0,v_0) = \sum_{k=1}^{8} OD(u_0,v_0, (u_k,v_k)) = 2L(u_0,v_0) + 2M(u_0,v_0) + 2N(u_0,v_0)
\]

where \( L(u_0,v_0) \) and \( N(u_0,v_0) \) are given by Eq. (2).

2.2. Neighbor distance in gray image

Because the oriented distance is defined on a smooth surface, we restore the smooth gray image surface with the Nadaraya-Waston kernel (NWK) estimator [18], which is a kernel regression method to estimate a non-linear relation between random variables. The purpose of using NWK estimator is to de-noise the image, and to decrease the neighbor distance’s sensitivity to noise, as the first- and second-order derivatives of \( L, M \) and \( N \) given by Eq. (2) are sensitive to noise. The kernel regression for image de-noising and image processing can be seen in Takeda’s papers [19,20].

For a two-dimensional image, let \( D \) be a symmetric 2-D point set centered at point \((0,0)\), so the set \( D \) with radius \( R \) and \( C \) can be presented as \( D = \{ |(i,j) : |i| \leq R, |j| \leq C \} \). Then, the point set centered at a point \((u_0,v_0)\) can be presented as \( (u_0,v_0) + D = \{ (u_0 + i, v_0 + j) : (i,j) \in D \} \). Denote the gray value of pixel \((u,v)\) as \( Z(u,v) \), then the gray image block centered at point \((u_0,v_0)\) can be represented as \( ([u_0,v_0],Z(u_0,v_0)) \) with \( (u,v) \in (u_0,v_0) + D \).

With the discrete image dataset \( ([u_0,v_0],Z(u_0,v_0)) \), the smooth image function \( S = \mathbb{Z}(u,v) \) estimated by NWK estimator [18] is

\[
\hat{Z}(u,v) = \sum_{(i,j) \in D} \frac{K(u,v) - (u_0-v_0)^2 + (v_0-v_0)^2}{-h^2} z(u_i,v_j)
\]

where \( (u,v) \in (u_0,v_0) + D, K(x) \) is a kernel function, and \( h \) is the bandwidth of NWK estimator.

Many functions are suitable to be the kernel function of NWK estimator. Some commonly used kernel functions are the boxcar kernel, the Gaussian kernel, the Epanechnikov kernel [21] and the Tricube kernel [22]. In this study, we only focus on the choice of the Gaussian function \( f(x) = \exp(-x^2) \) because of its good differential features. With the Gaussian kernel function, the estimated value \( \hat{Z}(u,v) \) at point \((u_0,v_0)\) can be rewritten as follows:

\[
\hat{Z}(u_0,v_0) = \sum_{i,j} \exp \left( \frac{-(u_i-u_0)^2 + (v_j-v_0)^2}{-h^2} \right) z(u_i,v_j)
\]

By Eq. (6), the \( \hat{Z}(u_0,v_0) \) is the linear combination of observed gray value \( Z(u_i,v_j) \), and the weights depend only on their spatial coordinates and the bandwidth \( h \), not on their gray value \( Z(u_i,v_j) \). Therefore, Eq. (6) can be rewritten as

\[
\hat{Z}(u_0,v_0) = \mathbb{E} \exp Z
\]

where \((\text{T})\) is the transposition operator, \( Z \) and \( \mathbb{E} \) are the expected value of matrix \( Z(i,j) \) and \( \mathbb{E}(i,j) \), whose element \( Z(i,j) \) and \( \mathbb{E}(i,j) \) are, respectively, given by

\[
\mathbb{E}(i,j) = \sum_{i,j} \exp \left( \frac{-(u_i-u_0)^2 + (v_j-v_0)^2}{-h^2} \right) \]

\[
Z(i,j) = Z(u_i,v_j)
\]

where \((i,j) \in D \), and \((u_i,v_j) \in (u_0,v_0) + D \).

Denote the first- and second-order derivatives of vector \( \mathbb{E} \) as \( \mathbb{E}_{uu}, \mathbb{E}_{ux}, \mathbb{E}_{uy}, \mathbb{E}_{vx}, \mathbb{E}_{vy}, \mathbb{E}_{vv} \). By Eq. (7), the first- and second-order derivatives of function \( \hat{Z}(u,v) \) at point \((u_0,v_0)\) can be presented as

\[
\hat{Z}_{u}(u_0,v_0) = (\mathbb{E}_{uu} Z, \mathbb{E}_{ux} Z, \mathbb{E}_{uy} Z)
\]

\[
\hat{Z}_{vv}(u_0,v_0) = (\mathbb{E}_{vx} Z, \mathbb{E}_{vy} Z, \mathbb{E}_{vv} Z)
\]

Substitute the first- and second-order derivatives in Eqs. (4) and (2) with Eq. (9), we can get

\[
ND(u_0,v_0) = 2L(u_0,v_0) + 2M(u_0,v_0) + 2N(u_0,v_0)
\]
In Eq. (10), the computational complexity of numerator $\text{EXP}_{uu}^iZ$ is $O(n^2)$, and the computational complexity of denominator $\sqrt{1+(\text{EXP}_{uu}^iZ)^2+(\text{EXP}_{vv}^jZ)^2}$ is $O(n^6)$, then computational complexity of neighbor distance is $O(n^6)$. The square operation and square root operation of denominator tremendously increase the computational complexity of Eq. (10). However, the non-zero elements of vector $\text{EXP}_{uu}$ and $\text{EXP}_{vv}$ were very small and negligible relative to the elements of vector $\text{EXP}_{uu}$ and $\text{EXP}_{vv}$. Hence, omitting the first-order derivative of $L$ and $N$ do not cause a drastic change of neighbor distance value, but it can reduce the computation complexity of neighbor distance from $O(n^6)$ to $O(n^2)$. Therefore, we omit the first-order derivative of $L$ and $N$, and rewrite the neighbor distance formula (10) as following:

$$ND(u_0, v_0) = 2(\text{EXP}_{uu}^iZ + \text{EXP}_{vv}^jZ) = 2(\tilde{z}_{uu}(u_0, v_0) + \tilde{z}_{vv}(u_0, v_0))$$  

(11)

It is worth to note that the $ND = 2(\tilde{z}_{uu} + \tilde{z}_{vv})$ is similar to the Laplace operator $f_{uu} + f_{vv}$. But the neighbor distance is deducted and reduced from the oriented distance in differential geometry, and $\tilde{z}_{uu}$ and $\tilde{z}_{vv}$ are estimated by kernel regression to decrease their sensitivities to noise.

From Eq. (8), the vectors $\text{EXP}_{uu}$ and $\text{EXP}_{vv}$ are determined by the bandwidth $h$ and the input vectors defined over $K \times C$. They are constant for a specified block size and a specified bandwidth, and then they can be pre-calculated. For a fixed rectangular block, reshape the vector $\text{EXP}_{uu}$ and $\text{EXP}_{vv}$, and we can get two constant matrices $E_{uu}$ and $E_{vv}$, Eq. (11) shows that the neighbor distance $ND(u_0, v_0)$ is the linear combination of the gray value $\tilde{z}(u,v)$, whose weights depend only on the constant matrices $E_{uu}$ and $E_{vv}$, but not on their gray values. That is to say, the neighbor distance value of every image pixel can be computed by convolving the image with the second-order derivative matrix $E_{uu}$ and $E_{vv}$, that is

$$ND(\text{image}) = 2(E_{uu} \ast \text{image} + E_{vv} \ast \text{image}) = 2(E_{uu} + E_{vv}) \ast \text{image}$$  

(12)

From the definition of neighbor distance, the flatter the surface at a point, the smaller the neighbor distance of this point; the more curving the surface at a point, the larger the neighbor distance of this point. So the neighbor distance can describe the image surface curvature at this point, and then, it can measure the gray value variation between pixel and its neighbor when the gray image is viewed as a surface. Therefore, the neighbor distance is a high frequency characteristic of gray image. As the neighbor distance of image is obtained by convolving the image with the filter $2(E_{uu} + E_{vv})$ by Eq. (12), this filter can be viewed as a high frequency filter of image. Based on the above analysis, we normalize the filter $2(E_{uu} + E_{vv})$ by dividing it by its maximum to construct a standard image analysis filter, and name the obtained filter as neighbor distance filter.

The construction procedure of neighbor distance filter can be summarized as follows:

1. Give the rectangular block size $R$ and $C$, the bandwidth $h$ of non-parametric regression.
2. Calculate the matrices $E_{uu}, E_{vv}$, where $E_{uu}(i,j) = \text{EXP}_{uu}(i,j)$, $E_{vv}(i,j) = \text{EXP}_{vv}(i,j)$ is given by Eq. (8).
3. Add the matrix $E_{uu}$ to $E_{vv}$, and normalize the obtained matrix by dividing it by its maximum.

If the size of block is set to be $5 \times 5$, and the square of bandwidth $h$ is set to be 0.4, then the deduced neighbor distance filter is given as following:

$$[\begin{array}{ccccccc}
-0.0004 & -0.0103 & -0.0274 & -0.0103 & -0.0004 \\
-0.0103 & -0.1256 & -0.0760 & -0.1256 & -0.0103 \\
-0.0274 & -0.0760 & 1.0000 & -0.0760 & -0.0274 \\
-0.0103 & -0.1256 & -0.0760 & -0.1256 & -0.0103 \\
-0.0004 & -0.0103 & -0.0274 & -0.0103 & -0.0004 \\
\end{array}]$$  

(13)

With the neighbor distance filter, we can quickly obtain the every pixel’s sharpness by convolving the image with this filter. Hence, the neighbor distance’s computation complexity is very low. To show the neighbor distance’s effectiveness in measuring sharpness, we blurred the Lena by Gaussian blur with standard deviation from 0.5 to 3, and computed the blurry image’s neighbor distance with filter in (13). We also decomposed these images by Laplacian transform and stationary wavelet transform, and summed their high frequency coefficients as the images’ sharpness. According to the experiment results reported in [2].

![Fig. 3. The effectiveness of different sharpness measures. The σ of Gaussian blur to Lena from up to down is 0.5, 1, 2 and 3, respectively.](image-url)
the Sum-modified-Laplace (SML) and Energy of Laplacian of image (EOL) can provide better performance than other four common sharpness measures: variance, energy of image gradient, Tenenbaum’s algorithm and spatial frequency, we only compared the blurry images’ sharpness measured by SML and EOL. The experiment results are shown in Fig. 3. From this figure, these sharpness measures all satisfy the following requirements [2]: (1) monotonic with respect to blur; (2) the sharpness measure must be unimodal, that is, it has one and only one maximum value; (3) large variation in value with respect to the degree of blurring; (4) minimal computation complexity. This figure also shows that the tail of ND curve is steeper than the other four tails. It means that the neighbor distance can provide better distinction than EOL, SML, LAP and SWT when the image is very blurry.

3. Multi-focus image fusion based on neighbor distance filter

Similar to the variance, EOG, SML etc, the neighbor distance is constructed in spatial domain, and we can choose the block-based fusion scheme as our fusion scheme. However, the block-based fusion methods have block effects and their performances strongly depend on the block size, and choosing a suitable block size for fusion is not a trivial question. Hence, we choose the multi-resolution transform-based fusion scheme as our fusion scheme. By the derivation of neighbor distance, the normalized neighbor distance filter is a high frequency filter, so the corresponding neighbor distance image can be considered as a high frequency component of the original image. Similar to the Laplacian pyramid [7] and SVT transform [14], the low-frequency component can be obtained by subtracting the neighbor distance image from the original image. Therefore, we construct a multi-scale neighbor distance analysis framework based on the neighbor distance filter in this section.

The existing multi-scale analysis tools can be divided into two basic schemes: pyramid and parallelepiped [14]. The pyramid framework is to sub-sample or decimate the image while keeping filter constant, and is widely used in the existing image fusion methods, such as the LAP and the DWT methods. The parallelepiped framework is to over-sample or fill the filter with zeros while keeping the image size unchanged, and is used in various un-decimated wavelet transforms, such as the a trous wavelet transform and the support value transform. Since the pyramid scheme includes a sub-sampling or decimating process, it may cause that the fusion result is shift-variant, which appears obvious in the standard discrete wavelet transform. For example, when there is a slight camera/object movement or there is mis-registration of the source images, the performance of the pyramid scheme-based fusion method will thus quickly deteriorate [23]. However, because of the parallelepiped framework is to over-sample or fill the filter.

![Fig. 4. The decompositions of Lena. (a–d) The high frequency component; (e) the low frequency component; (f–i) the histograms of sub-image (a–d); (j) the original image Lena.](image)

![Fig. 5. The five group multi-focus test images. The upper row: images focus on the objects far-away from the camera, and the lower row: images focus on the closer objects. Columns 1–5: book, clock, desk, lab and pesi.](image)
with zeros while keeping the image size unchanged, the parallelepiped framework is superior to the pyramid scheme in shift-invariant and it is suitable to the multi-focus image fusion [24,25].

With the parallelepiped scheme, a series of multi-focus neighbor distance filters can be obtained by filling the basic neighbor distance filter with zeros. And then multi-scale neighbor distance analysis is available. The multi-scale neighbor distance analysis is an un-decimated dyadic transform and isotropic. It is shift-invariant, hence, it does not create artifacts.

Given an image \( I \), the sequence of its neighbor distance images obtained by multi-scale decomposition is \( \{D_1, D_2, \ldots, D_r\} \). The sequence of its approximations is the differences of the source image and its neighbor distance image. That is

\[
D_i = ND_i \ast I_i \\
I_{i+1} = I_i - D_i, \quad i = 1, \ldots, r, \quad I_1 = I
\]

where \( r \) refers to the decomposition level, \( \ast \) present the convolution operator and \( ND_i \) is the neighbor distance filter, which is obtained by filling the basic neighbor distance filter with \( 2^{i-1} \) zeros in order to match the resolution of desired level as the atrous wavelet and the support value transform.

The reconstruction formula can be written as

\[
I = I_{r+1} + \sum_{i=1}^{r} D_i
\]

Fig. 4 is the image Lena and its decompositions. The decomposition level is 4, the basic distance filter is same as (13). The coefficient histograms (f)–(i) of Fig. 4 are characterized by a very sharp peak at zero amplitude and extended tails to both sides of peak. This implies that the neighbor distance images are very sparse, as the majority of coefficients have amplitudes close to zero.

Similar to the traditional wavelet transform-based and support value transform-based image fusion, the general steps of the neighbor distance-based multi-focus image fusion approach can be listed as follows:

- **Step 1**: Co-register the source images and re-sample them to make their pixel size the same as one another.
- **Step 2**: Apply the neighbor distance multi-scale decomposition to each of the co-registered images and obtain a low-frequency component image and the neighbor distance image sequences.
- **Step 3**: Combine the multiple sets of low-frequency/neighbor distance components together. The most popular method is to select the components with largest activity level at each pixel location [choose-max (CM)]. Alternatively, one can also use salience/match measure with threshold [as proposed by Burt and Koczynski [9]] or choose maximum with consistency check [as proposed by Li et al. [26]] to select the components.
- **Step 4**: Optionally, perform consistency verification, which ensures that a fused coefficient does not come from a different source image from most of its neighbors. Usually, this is implemented by using a small majority filter.
- **Step 5**: Use the reconstruction formula (15) to recover the fused image.

### 4. Experimental results

In this section, we test the proposed image fusion method based on the neighbor distance filter on the popular multi-focus image fusion and compare it with other methods. The performance of the proposed method is evaluated based on objective and subjective criteria.

![Fig. 4](image1.png)

**Fig. 4**: The image Lena and its decompositions. The decomposition level is 4, the basic distance filter is same as (13). The coefficient histograms (f)–(i) of Fig. 4 are characterized by a very sharp peak at zero amplitude and extended tails to both sides of peak. This implies that the neighbor distance images are very sparse, as the majority of coefficients have amplitudes close to zero.

![Fig. 6](image2.png)

**Fig. 6**: The fused image obtained by ND with different high frequency coefficients selections. Low frequency: average; high frequency coefficients selection: row 1: choose-max; row 2: salience match (th. = 0.75); row 3: choose-max with consistency check.

<table>
<thead>
<tr>
<th>Test image</th>
<th>Book</th>
<th>Clock</th>
<th>Desk</th>
<th>Lab</th>
<th>Pesi</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>61.8573</td>
<td>40.9559</td>
<td>47.1712</td>
<td>47.237</td>
<td>45.3743</td>
</tr>
<tr>
<td>AV</td>
<td>9.4344</td>
<td>3.4596</td>
<td>5.8436</td>
<td>4.6179</td>
<td>5.5375</td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>62.5365</td>
<td>41.0765</td>
<td>47.4128</td>
<td>47.5362</td>
<td>45.4557</td>
</tr>
<tr>
<td>AV</td>
<td>9.6696</td>
<td>3.5477</td>
<td>5.9141</td>
<td>4.6834</td>
<td>5.802</td>
</tr>
<tr>
<td>(c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>62.0901</td>
<td>41.0107</td>
<td>47.19</td>
<td>47.355</td>
<td>45.4302</td>
</tr>
<tr>
<td>AV</td>
<td>9.4942</td>
<td>3.4872</td>
<td>5.83</td>
<td>4.6233</td>
<td>5.5869</td>
</tr>
<tr>
<td>SF</td>
<td>28.5588</td>
<td>8.6064</td>
<td>15.7419</td>
<td>13.1722</td>
<td>14.0253</td>
</tr>
</tbody>
</table>
test images (book, clock, desk, lab and pesi) shown in Fig. 5 Rockinger’s Matlab toolbox\(^1\) is used as the reference for the implementation of LAP method, FSD method, GRP method and RAP method. The non-subsampled contourlet toolbox\(^2\) is used as the reference for non-subsampled contourlet transform (NSCT). The Daubechies wavelet function “db4” is used in the discrete wavelet transform (DWT) and the stationary wavelet transform (SWT). We also compared the fused results with results obtained by the support value transform (SVT). The window size of salience computation and the consistency check is set to be 3\(^2\). For the proposed fusion method, the basic neighbor distance filter in (13) is used.

Evaluation of image fusion is a non-trivial tasks especially as it is often difficult to say which of two fused images is better without reference specific tasks [27]. Although there have been many attempts to quantitative evaluation, as yet, no universally accepted standard has been developed. In this experiment, we used the standard deviation (SD), average gradient (AG) and spatial frequency (SF) to quantitatively evaluate the performance of the proposed fusion method. The standard deviation is defined as follows:

$$SD = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (F(m,n) - \mu)^2$$

where \(F\) is the fused image, the \(\mu\) is the image’s mean, and 
$$\mu = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} F(m,n)$$

The larger the standard deviation, the sharper the image.

The average gradient is given by

$$AG = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sqrt{\Delta F_x^2(m,n) + \Delta F_y^2(m,n)}$$

where \(\Delta F_x\) and \(\Delta F_y\) are the differences in \(x\) and \(y\) direction of the fused image. The larger the average gradient, the sharper the image.

The spatial frequency is defined as

$$SF = \sqrt{RF^2 + CF^2}$$

where \(RF\) and \(CF\) are the row frequency

$$RF = \sqrt{\frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [F(m,n) - F(m,n-1)]^2}$$

---

\(^1\) [http://www.metapix.de/download.htm](http://www.metapix.de/download.htm)

\(^2\) [http://www.ifp.illinois.edu/minhdo/software/](http://www.ifp.illinois.edu/minhdo/software/)
and column frequency

\[
RF = \sqrt{\frac{1}{MN} \sum_{m=1}^{M-1} \sum_{n=0}^{N-1} (F(m,n) - F(m-1,n))^2}
\]

respectively, and \( F \) is the fused image.

Fig. 6 gives the fused images obtained by the neighbor distance method with different high frequency coefficients selection. Although these images with different coefficients selection have the similar visual effect, their values of objective evaluation indexes given in Table 1 are different. The salience match and the choose-max with consistency check, can improve the fusion results in terms of the SD, AV and SF. For example, for the book, the AV value is increased from 9.4344 to 9.6696, when the coefficient selection is changed from choose-max to salience match. Similar results can be found for other test images.

Fig. 7 is the fused image obtained by the GRP, LAP, DWT, SWT, and neighbor distance, respectively. If we carefully observe these images, it can be found that the visual difference between the fused image from the row 1 to row 5 is not obvious. The proposed fused method provides a very close visual performance to the popular fused methods on multi-focus images.

Table 2 gives the values of standard deviation, average gradient and spatial frequency for the image fusion approaches including DWT, FSD, GRP, LAP, RAP, SVT, SWT, NSCT and the proposed neighbor distance image fusion schemes, where the

<table>
<thead>
<tr>
<th>Method</th>
<th>DWT</th>
<th>FSD</th>
<th>GRP</th>
<th>LAP</th>
<th>RAP</th>
<th>SVT</th>
<th>SWT</th>
<th>NSCT</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>47.9772</td>
<td>45.7757</td>
<td>45.7819</td>
<td>48.4938</td>
<td>46.5032</td>
<td>47.8688</td>
<td>47.8807</td>
<td>47.9175</td>
<td>48.5191</td>
</tr>
<tr>
<td>AV</td>
<td>5.604</td>
<td>4.6507</td>
<td>4.6196</td>
<td>5.6156</td>
<td>3.8209</td>
<td>5.1648</td>
<td>5.5690</td>
<td>5.6668</td>
<td>5.7791</td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>48.2211</td>
<td>46.035</td>
<td>46.0466</td>
<td>48.8001</td>
<td>47.3647</td>
<td>48.0495</td>
<td>47.9570</td>
<td>48.0050</td>
<td>48.8036</td>
</tr>
<tr>
<td>AV</td>
<td>5.8505</td>
<td>4.737</td>
<td>4.6974</td>
<td>5.7275</td>
<td>4.2158</td>
<td>5.2583</td>
<td>5.6770</td>
<td>5.7620</td>
<td>5.9234</td>
</tr>
<tr>
<td>(c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>48.0087</td>
<td>45.7855</td>
<td>45.7917</td>
<td>48.564</td>
<td>48.3013</td>
<td>47.9736</td>
<td>47.9523</td>
<td>47.9764</td>
<td>48.6152</td>
</tr>
<tr>
<td>AV</td>
<td>5.626</td>
<td>4.6702</td>
<td>4.6419</td>
<td>5.6569</td>
<td>5.7034</td>
<td>5.2803</td>
<td>5.6121</td>
<td>5.7133</td>
<td>5.8043</td>
</tr>
</tbody>
</table>

Table A2
The performance of different fusion methods on test image: clock.

<table>
<thead>
<tr>
<th>Method</th>
<th>DWT</th>
<th>FSD</th>
<th>GRP</th>
<th>LAP</th>
<th>RAP</th>
<th>SVT</th>
<th>SWT</th>
<th>NSCT</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>40.3966</td>
<td>39.1287</td>
<td>39.1341</td>
<td>40.9321</td>
<td>39.4398</td>
<td>39.1341</td>
<td>40.4739</td>
<td>40.3139</td>
<td>40.3616</td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>40.5197</td>
<td>39.3515</td>
<td>39.3559</td>
<td>41.0353</td>
<td>40.9788</td>
<td>40.5579</td>
<td>40.3193</td>
<td>40.3679</td>
<td>41.0765</td>
</tr>
<tr>
<td>(c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>40.4079</td>
<td>39.1029</td>
<td>39.1031</td>
<td>40.0059</td>
<td>40.7092</td>
<td>40.4902</td>
<td>40.3242</td>
<td>40.3667</td>
<td>41.0107</td>
</tr>
</tbody>
</table>
decomposition level is set to be three. All the reported values are the average results of the performance of five group test images shown in Tables A1–A5.

From Table 2, we can easily note that the proposed neighbor distance-based fusion method provides the best fusion performance in terms of resulting values of the quantitative evaluation indexes including $SD$, $AV$ and $SF$. For example, the $AV$ value of proposed ND is 5.7791 and is greater than 5.604, 4.6507, 4.6196, 5.6156, 3.8209, 5.1648, 5.5690 and 5.6668, respectively provided by the DWT, FSD, GRP, LAP, RAP, SWT, NSCT when the high frequency coefficients selection method is choose-max. This means that the proposed neighbor distance-based fusion method provides the sharper fused image. The other quantitative evaluation indexes—$SD$ and $SF$—also show the similar results when the high frequency coefficients selection method is salience match or choose-max with consistency check. Although the $AV$ and $SF$ values of NSCT in Tables A3 and A4 are greater than the proposed ND $AV$ and $SF$ values, it is worth to note that the proposed ND performs better than NSCT in terms of $SD$, $AV$ and $SF$ shown in other Tables A1, A2, A5 and the average values Table 2. And the $SD$ values of proposed ND in Tables A3 and A4 are also greater than the $SD$ values of NSCT.

Based on the above analysis, we can see that the neighbor distance can effectively describe the image’s clarity and neighbor distance filter-based multi-focus image fusion method is effective and it is superior to the conventional image fusion methods including the DWT, FSD, GRP, LAP, RAP, SWT and NSCT methods, in term of the related quantitative fusion evaluation indexes including $SD$, $AV$ and $SF$, on the multi-focus image fusion test.

<table>
<thead>
<tr>
<th>Table A3</th>
<th>The performance of different fusion methods on test image: desk.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>DWT</td>
</tr>
<tr>
<td>(a)</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>AV</td>
</tr>
<tr>
<td>(b)</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>AV</td>
</tr>
<tr>
<td>(c)</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>AV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table A4</th>
<th>The performance of different fusion methods on test image: lab.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>DWT</td>
</tr>
<tr>
<td>(a)</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>AV</td>
</tr>
<tr>
<td>(b)</td>
<td>SD</td>
</tr>
<tr>
<td>(c)</td>
<td>SD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table A5</th>
<th>The performance of different fusion methods on test image: pesi.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>DWT</td>
</tr>
<tr>
<td>(a)</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>AV</td>
</tr>
<tr>
<td>(b)</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>AV</td>
</tr>
<tr>
<td>(c)</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>AV</td>
</tr>
</tbody>
</table>
5. Conclusion

This work presents a new multi-focus image fusion method based on the neighbor distance. The neighbor distance is deduced and reduced from the oriented distance in differential geometry. Based on the experiments presented in Section 4, we can draw a conclusion that the neighbor distance can effectively measure the clarity of multi-focus image's pixel and the proposed image fusion method can extract the clearer pixels and integrate them into the resulting image.

Acknowledgments

We thank the associate editor and anonymous reviewers for their helpful and constructive suggestions. This work is supported by the Fundamental Research Funds for the Central University (CDJXS11182240) and the National Natural Science Foundation of China (61173129, 61173130, 61273244).

Appendix A. The performance of different fusion methods on five group test images

References


Hengjun Zhao received his B.Sc. and M.Sc. degrees in Applied Mathematics, from Southwest University in 2004 and 2007, respectively. Since 2009, he is currently a doctoral student in the Department of Computer Science at Chongqing University. His research interests include pattern recognition and image processing.

Zhaowei Shang received his Ph.D. degree in Computer Science from Xi’an Jiaotong University, China. He is presently a professor in the Department of Computer Science at Chongqing University. His research interests include wavelet theory and image processing.

Yuanyan Tang received the Ph.D. degree in Computer Science from Concordia University, Montreal, Canada. He is presently a professor in the Department of Computer Science at Chongqing University and a chair professor in the Faculty of Science and Technology, University of Macau. His current interests include wavelet theory and applications, pattern recognition, image processing, document processing, etc. He is the Founder and Editor-in-Chief of International Journal on Wavelets, Multiresolution, and Information Processing (IJWMIP). Professor Y.Y. Tang is an IAPR fellow and IEEE fellow.

Bin Fang received the B.E. degree in Electrical Engineering from Xi’an Jiaotong University, Xi’an, China, the M.S degree in Electrical Engineering from Sichuan University, Chengdu, China, and the Ph.D degree in Electrical Engineering from the University of Hong Kong, Hong Kong, China. He is currently a professor in the Department of Computer Science at the Chongqing University. His research interests include computer vision, pattern recognition, medical image processing, biometrics applications, and document analysis.