Steerable local frequency based multispectral multifocus image fusion

Vijay N. Gangapure, Sudipta Banerjee, Ananda S. Chowdhury*

Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata 700032, India

ARTICLE INFO

Article history:
Received 5 March 2014
Received in revised form 28 July 2014
Accepted 30 July 2014
Available online 8 August 2014

Keywords:
Multispectral focus measure
Multifocus image fusion
Oriented analytic image
Steerable local frequency

ABSTRACT

Design of a focus measure and a fusion algorithm which will perform well across different spectra remains an extremely challenging task. In this work, the problem of multispectral multifocus image fusion is addressed using the phase information of the source image pixels at different orientations. We make the local frequency, the spatial derivative of the local phase of the pixels, steerable to obtain a good novel focus measure. Oriented analytic image based on the theory of steerable filters is constructed for that purpose. A multifocus fusion algorithm is proposed next using this focus measure. Comprehensive experiments clearly demonstrate that our focus measure as well as the multifocus fusion algorithm yield promising results across the visual (VIS), the near-infrared (NIR) and the thermal (TH) spectra.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Image fusion with several important applications in diverse areas like medicine, surveillance and remote sensing has generated lot of interest among the researchers over the last decade [1]. Multifocus image fusion aims at integrating multiple images of the same scene captured at different focal settings into a single all-in-focus image. This all-in-focus image can be thought of as an ensemble of the best focused pixels extracted from the set of source images. According to Stathaki [2], multifocus image fusion methods can be broadly classified into spatial domain methods and transform domain methods. In spatial domain methods [3,4], the fusion rules are directly applied to the pixels or a region in an image. In contrast, in the transform domain methods, images are initially processed by discrete cosine, wavelet and similar other transforms before the fusion rules are applied [5].

The fundamental step behind the multifocus fusion lies in the determination of the focus quality of the source images [6,7]. Zukal et al. introduced the determination of focus measure via interest point detection [8]. An interest point in the image has significant local variations such as a corner or a junction. Very recently, detection of interest point based focus measure for multispectral images has gained popularity. A majority of these measures are based on intensity [9]. However, some of the interest point detection methods are found to use frequency and phase congruency [10,11]. In the present work, we compute local frequency of the pixels from their local phase at different orientations. The proposed steerable local frequency (local frequency considered at different orientations) based focus measure is shown to perform well in various spectra like visual (VIS), near-infrared (NIR) and thermal (TH). We further illustrate that this focus measure yields good performance for multispectral multifocus image fusion.

The rest of the paper is organized in the following manner: in Section 2, we discuss some popular focus measures and multifocus image fusion schemes with their limitations and highlight our contributions. In Section 3, we provide the necessary theoretical foundations. In Section 4, we describe the proposed method in details. In Section 5, we compare the performance of the proposed method with several existing approaches. Finally, the paper is concluded in Section 6 with an outline of directions for future research.

2. Related work

Multifocus image fusion consists of two major steps, namely, the focus measure computation and its application in image fusion. We first discuss certain limitations of the existing focus measures. Some problems in the current fusion schemes are analyzed next. We end this section highlighting our contributions.

Focus measure algorithms are classified into four broad categories, such as, derivative based, statistics based, histogram based and intuition based. In [12], Liu et al. evaluated the performance of eighteen focus measures, e.g., EOG (Energy of Gradient), SML (Sum of Laplacian), EOL (Energy of Laplacian), TEN (Tenengrad) for microscopic images from the above four categories. All the above focus measures are based on variations in pixel intensities only. These methods have several drawbacks, like, the performance variation with spectral content of the source images, insensitivity
to defocus, fluctuation with noise content and narrow effective range. To overcome these limitations, Minhas et al. [13] proposed a novel efficient focus measure for shape from focus (SFF) application. In another method, Tian and Chen [14], examined statistics of details wavelet coefficients to perform sharpness measurement in the input image. Zhao et al. [15] related the degree of gray level surface curvature to the sharpness of image region. But these attempts to measure focus are limited to the visual spectrum only. In other spectra like thermal and near-infrared, less focus measure works are reported due to unavailability of scene viewing in case of manual focusing, and limited resolution as well as lack of auto-focus features in the cameras. Faundez-Zanuy et al. [16] addressed the problem of determining the optimal focus position in the thermal images. Zukal et al. [9] proposed a focus measure based on interest points detection (IPD) to achieve uniform performance in multispectral imaging. The results show that this focus measure performs better than the standard focus measures for thermal images but lag behind in the visual and the near-infrared spectra.

Many fusion algorithms are available in the literature. These algorithms operate at pixel level or region level, and in spatial as well as transform domains. Spatial domain pixel level algorithms are popular due to their computational efficiency [2]. Multiresolution transform based algorithms are preferred nowadays due to their robust performance [17]. Within the transform domain, Discrete Wavelet Transform (DWT) based algorithms, though perform better than the Laplacian Pyramid Transform (LPT), have limited orientation selectivity [18]. More improved multiresolution transform techniques include Stationary Wavelet Transform (SWT), Curvelet Transform (CVT), Contourlet Transform (CT), Dual Tree Complex Wavelet Transform (DTCWT) and Non-Subsampled Contourlet Transform (NSCT) [17,18]. In [17], Li et al. has evaluated the performance of such multiresolution transforms for multifocus image fusion in the visual spectrum only. Benes et al. [19] proposed a new multifocus image fusion algorithm for thermal images where they employed pixel level weighted averaging based on modified EOL. But such a linear combination often fails to preserve the original information in source images leading to a degradation in the fusion performance.

So, existing literature clearly suggests that designing a focus measure and applying it for fusion across different spectra still poses a considerable challenge. In this paper, we propose a novel focus measure using steerable local frequency based interest point detection. A recent work is reported where pixel intensities at different orientations are considered for obtaining a focus measure [13]. However, phase of a pixel carry more useful information than intensity [20]. To the best of our knowledge, this phase information has not been captured in different orientations earlier. In this paper, we make the local frequency of the pixels, the spatial derivative of the local phase, steerable to obtain a good focus measure. For this purpose, we suggest the construction of the oriented analytic image. The proposed focus measure captures all possible sharp image features in different orientations and hence perform well across all spectra. As a second contribution, we employ our focus measure for multispectral multifocus image fusion. Detailed experimentation reveal much improved multifocus image fusion performances across different spectra.

3. Theoretical foundations

In this section, we provide the theoretical foundations behind the proposed method. In particular, analytic image and steerable filters are discussed in details.

3.1. Analytic image

We start with the concept of analytic signal in 1D, which can be easily extended to higher dimensions [21]. Given a time domain signal $s(t)$ in 1D, its analytic signal is defined as:

$$s_a(t) = s(t) - j s_H(t)$$

where $s_H(t)$ is the Hilbert transform of $s(t)$. An image can be treated as a 2D spatial domain signal. Corresponding analytic image can be expressed as:

$$I_a(x,y) = I(x,y) - j I_H(x,y)$$

where $I_H(x,y)$ is the Hilbert transformation of $I(x,y)$. Argument of $I_a(x,y)$, defined in the spatial domain, is referred to as the local phase of $I(x,y)$. Khan et al. [10] have used the local frequency of an image to capture the dominant regions in an image (a dominant region will contain many pixels with high local frequencies). The local frequency can be determined easily as it is the spatial derivative of local phase. High value of local frequency at a particular pixel of an image indicates the presence of interest point at that location. Concept of quadrature pair of filters can be introduced in this context. Quadrature pair of filters has same frequency response but differ in phase by an angle of 90°, i.e., in effect they must be Hilbert transforms of each other [24].

3.2. Steerable Gaussian filter

Steerable filters can be defined as a special class of filters in which an arbitrarily oriented filter can be designed using a linear combination of a set of basis filters [22,23]. The direction derivative of a 2D Gaussian function is steerable because of its circular symmetry. In [24], Freeman and Adelson have shown that the first order $x$-derivative ($G_1^0$) of a Gaussian filter oriented at an arbitrary orientation $\theta$ can be expressed as a linear combination of $G_0^0$ and $G_0^0$ in the following manner:

$$G_1^\theta = \cos(\theta) G_0^\theta + \sin(\theta) G_0^\theta$$

In the above equation, $G_0^0$ and $G_0^0$ are the basis filters and the terms $\cos(\theta)$ and $\sin(\theta)$ are the interpolation functions. Thus an image filtered at any orientation can be expressed as a linear combination of the image convolved with the basis filters (convolution operation being linear).

Then, we can write:

$$R_1^\theta = G_1^\theta * I$$

$$R_0^\theta = G_0^0 * I$$

$$R_1^\theta = \cos(\theta) R_0^\theta + \sin(\theta) R_0^\theta$$

In the above equation, $R_1^\theta$ represents the image $I$ filtered using the basis filter at an arbitrary orientation $\theta$ and $\ast$ denotes the convolution operation.

4. Proposed method

Oppenheim and Lim [20] demonstrated the importance of phase in images through a series of experiments. The standard focus measures are mainly based on intensity. In this work, we explore the potential of local phase information of the pixels in the source images for determining the focus measure. Features in an image can be oriented at any angle $\theta$ (0° ≤ $\theta$ ≤ 180°) [25]. For each pixel, corresponding responses from the filter at different orientations need to be compared to get the maximum response. Local frequency map obtained from the analytic image does not include any knowledge of orientation. To capture orientation, we introduce the concept of steerable local frequency map. Oriented analytic image is used to build the steerable local frequency map. Hilbert transform, realized through the quadrature pair of filters ($G_0, H_0$), is used first to obtain the analytic image. Fourth order derivative of Gaussian ($G_2$) offers higher resolution analysis as it
has narrow frequency tuning. In [24], the approximation to the Hilbert transform of \( G_d \) denoted by \( H_d \) is obtained using the least squares fit of product of a 5th order polynomial with six basis functions and a radially symmetric Gaussian function. To obtain the oriented analytic image, we therefore require a steerable Hilbert kernel. However, since the Hilbert Transform itself cannot be made steerable in its present form, we apply the concept of steerable quadrature pair of filters \((G'_4, H'_4)\). The analytical expression of \( G'_4 \) is given by:

\[
G'_4 = (K_x(0)G_a + K_x(0)G_b + K_x(0)G_c + K_x(0)G_d + K_x(0)G_e)
\]

where \( G_a, G_b, G_c, G_d \) and \( G_e \) constitute the basis set functions and \( K_{x}(0), K_{y}(0), K_{x}(0), K_{y}(0) \) and \( K_{x}(0)K_{y}(0) \) are the interpolation functions. Similarly, the analytic expression for and \( H'_4 \) is given by:

\[
H'_4 = (K_x(0)H_a + K_x(0)H_b + K_x(0)H_c + K_x(0)H_d + K_x(0)H_e)
\]

where \( H_a, H_b, H_c, H_d, H_e \) and \( H_f \) are basis set functions and \( K_{x}(0), K_{y}(0), K_{x}(0), K_{y}(0) \) and \( K_{x}(0)K_{y}(0) \) are the interpolation functions. The equations for basis and interpolation functions are given in Appendix A. This steerable quadrature pair \( G'_4 \) and \( H'_4 \) is used to filter the original image \( I(x,y) \) to obtain the oriented analytic image \( I_{x,0}(x,y) \) at an arbitrary orientation \( \theta \). So, we can write:

\[
I_{x,\theta}(x,y) = I(x,y) + G'_4
\]

\[
I_{y,\theta}(x,y) = I(x,y) + H'_4
\]

\( I_{x,\theta}(x,y) \) and \( I_{y,\theta}(x,y) \) together constitute the steerable quadrature filtered response of the original image \( I(x,y) \). The steerable local phase \( \Phi(x,y) \) of the Gaussian filtered image can now be obtained using

\[
\Phi(x,y) = \text{abs}(\text{arctan}(I_{x,\theta}(x,y)/I_{y,\theta}(x,y)))
\]

To get rid of background noise or distortions, the mean of the steerable local phase map is subtracted from the phase value at each of the pixels to construct the modified phase map \( \Phi'_d(x,y) \) [25].

\[
\Phi'_d(x,y) = \Phi(x,y) - \overline{\Phi}
\]

In the above equation, \( \overline{\Phi} \) is mean of the steerable local phase map. Steerable local frequency map is obtained using the gradient of the modified local phase in the following manner:

\[
\text{Freq}_d(x,y) = \sqrt{\left\{ \left( \frac{\partial \Phi'_d(x,y)}{\partial x} \right)^2 + \left( \frac{\partial \Phi'_d(x,y)}{\partial y} \right)^2 \right\}}
\]

where \( \frac{\partial \Phi'_d(x,y)}{\partial x} = \Phi'_d(x+1,y) - \Phi'_d(x,y) \)

and \( \frac{\partial \Phi'_d(x,y)}{\partial y} = \Phi'_d(x,y+1) - \Phi'_d(x,y) \)

The local frequency maps obtained at different orientations are further max-pooled to obtain the resultant steerable local frequency map, \( \text{Freq}_{\text{max}} \):

\[
\text{Freq}_{\text{max}}(x,y) = \max(\text{Freq}_d(x,y), \text{Freq}_{d1}(x,y), \ldots, \text{Freq}_{d13}(x,y))
\]

In the above equation, \( \theta_1, \theta_2, \ldots, \theta_{13} \) denote 13 orientations covering the entire range \([0^\circ, 180^\circ]\) in steps of 15°. Number of orientations is chosen experimentally and this is described later (Section 5.2.2). We then choose the best suitable threshold \( T \) experimentally, for thresholding max-pooled local frequency map, \( \text{Freq}_{\text{max}} \) to compute the number of interest points in source image. Selection of the best performing threshold \( T \) is also discussed later (Section 5.2.1). Number of interest points \( n \) detected in the source image is given by:

\[
n = \sum_x \sum_y (\text{Freq}_{\text{max}}(x,y) \geq T)
\]

Thus, the proposed focus measure can be normalized in [0,1] using [9]:

\[
F_{\text{M proposed}} = \frac{n - n_{\text{min}}}{n_{\text{max}} - n_{\text{min}}}
\]

Here, \( n_{\text{max}} \) is the maximum number of interest points and \( n_{\text{min}} \) is the minimum number of interest points detected among all the source images in a set \( I \).

We now present Algorithm 1 where various steps to obtain the proposed focus measure are shown.

---

**Algorithm 1: Computation of Focus Measure (FM)**

**Input:** \( I \), An image from Visual spectrum (VIS), Near-Infrared spectrum (NIR) or Thermal spectrum (TH) image set.

**Output:** \( F_{\text{M proposed}} \), Image level Focus measure of the input image.

1. **// Obtain local frequency maps at different orientations:**
   - for \( \theta = 0^\circ \): 180° in step of 15° // \( \theta \) is the orientation angle.
   - Compute 7 × 7 quadrature pair kernels \( G'_4 \) and \( H'_4 \) at orientation \( \theta \).
   - Convolve image \( I \) with kernels \( G'_4 \) and \( H'_4 \) to obtain \( I_{x,\theta}(x,y) \) and \( I_{y,\theta}(x,y) \).
   - Compute Oriented analytic image \( I_{x,\theta}(x,y) = I_{x,\theta}(x,y) - jI_{y,\theta}(x,y) \).
   - 5. Obtain steerable local phase map in \( I \):
     \( \Phi(x,y) = \text{abs}(\text{arctan}(I_{x,\theta}(x,y)/I_{y,\theta}(x,y))) \)
   - 6. Obtain modified steerable local phase map in \( I \):
     \( \Phi'_d(x,y) = \Phi(x,y) - \overline{\Phi} \)
   - 7. Obtain steerable local frequency \( \text{Freq}_d(x,y) \) in \( I \): use gradient (\( \Phi'_d(x,y) \)), see Eqs. (12) and (13).

2. **// Obtain max-pooled steerable local frequency map \( \text{Freq}_{\text{max}} \) from steerable local frequency maps \( \text{Freq}_d \):**
   - for \( i = 1 : p \)
   - for \( j = 1 : q \) // where \( p \times q \) is the size of the max-pooled steerable local frequency map:
     - \( \text{Freq}_{\text{max}}(i,j) = \max(\text{Freq}_d(i,j), \text{Freq}_{d1}(i,j), \ldots, \text{Freq}_{d13}(i,j)) \)

3. **// Obtain interest points from resultant frequency map:**
   - Determine an experimental threshold \( T \) to yield interest points.
   - // for \( n = 0 \)
   - 13. for \( i = 1 : p \)
   - 14. for \( j = 1 : q \) // where \( p \times q \) is the size of the max-pooled steerable frequency map:
     - if \( \text{Freq}_{\text{max}}(i,j) \geq T \) // Threshold
     - 16. \( n = n + 1 \) // Since, \( (i,j) \) is an interest point.

4. **// Compute focus measure:**
   - \( F_{\text{M proposed}} = (n - n_{\text{min}})/(n_{\text{max}} - n_{\text{min}}) \) // \( n_{\text{max}} \) is the maximum number of interest points and \( n_{\text{min}} \) is the \( n_{\text{min}} \) // minimum number of interest points detected among all the images in a set to which source image \( I \) belongs.
The proposed multifocus image fusion begins with the assumption that the source images to be fused are pre-registered. The resultant fused image \( F \) contains pixels from the source image having highest max-pooled local frequency value for that pixel. Further, a \( 3 \times 3 \) majority filter is applied for consistency verification [15]. This step ensures that a pixel in the fused image is not allowed to come from a source image if majority of its neighbors in the two images (fused and the source) are different. This measure plays an important role in characterizing the performance of image fusion algorithms.

We, next, present Algorithm 2 where various steps for proposed multifocus image fusion are given.

Algorithm 2: Procedure for Multifocus Image Fusion

| Input: \( I_{Nt} \) Multiple registered source images from VIS, NIR or TH spectrum. |
| Output: \( F \) All-in-focus image. |

1. Obtain max-pooled steerable local frequency map of each source image. For each of the source image using Algorithm 1 (Step 1 to Step 10).
2. Obtain the final fused image \( F \) by selecting the pixels \((i, j)\) from the set of source images which yields maximum local frequency for the respective pixels.
3. Perform consistency verification using \( 3 \times 3 \) majority filter to obtain the output.

5. Experimental results

In this section, we first mention the datasets used for various experimentations along with the performance evaluation criteria. We next discuss how certain parameters are chosen experimentally. We then show the comparative performance analysis of the proposed focus measure. Finally, we demonstrate the improvements in fusion results using our focus measure.

5.1. Datasets and performance metrics

For the evaluation of focus measure performance we use three multispectral datasets, one each from the visual (VIS), near-infrared (NIR) and thermal (TH) spectrum [9]. Each dataset in turn consists of seven sets of images. Some sample image sets are shown in Fig. 1. For the evaluation of the multifocus image fusion in the visual spectrum, we use the same image sets as in [17] (see Fig. 2). In the near-infrared, only one image set was found to be suitable from the available NIR dataset [9] (see Fig. 3). We also use a multimodal medical image set to evaluate proposed fusion method. This consists of CT and MRI modality image set of human brain and the images are shown in Fig. 4. For the thermal spectrum, we experiment with the reduced set of multifocus thermal image datasets developed by Benes et al. [19]. The original thermal image database consists of five multifocus image sets with 96 images in each set. All the sets contain a scene image with two objects but with different backgrounds, varying temperatures and different object distances. A reduced set of 10 images for each dataset is derived from the original pool of 96 images using EOL based activity level measurement [19]. The reduced image sets for the mobile-interface and the two bulbs are shown in Figs. 5 and 6.

The performance evaluation measures are now briefly discussed below. The focus measure is evaluated based on different criteria such as monotonicity, magnitude of slope and smoothness. For this we employ the \( Q \) (Quality factor) and \( P \) (Peak of focus curve) performance metrics [9].

1. \( Q \) (Quality factor): The quality factor is computed from the focus curve. The focus curve is the plot between image index \( N \) and focus measure \( FM_{\text{proposed}} \). The formula for \( Q \) is given below:

\[
Q = \frac{1}{N_{\text{max}} - N_{\text{min}} + 1}
\]

\[
C_i[N] \geq 0.7079 \quad \text{For} \quad N_{\text{min}}, \ldots N_{i}, \ldots N_{\text{max}}
\]

\( C_i[N] \) in Eq. (18) is the focus curve normalized in the range \([0,1]\). Number of focus curve samples higher than 0.7079 are used to measure the \( Q \) factor. A narrow peak in the focus curve with a high \( Q \)-factor is favorable.

2. \( P \) (Peak of focus curve): \( P \) represents the image having highest focus evaluated from the focus curve, \( C_i[N] \). For the evaluation of multifocus image fusion in the visual and near-infrared spectrum, we use \( MI \) (Mutual Information), \( Q^{AB}\) and \( Q_0 \). They are described below:

\[ MI = I_{AF} + I_{BF} \] (19)

In Eq. (19) \( I_{AF} \) is the mutual information between the source image \( A \) and the fused image \( F \) whereas \( I_{BF} \) is the mutual information between the source image \( B \) and the fused image \( F \). A high value of \( MI \) indicates better result.

\[
Q^{AB} = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} [Q^{AF}(n,m)w^A(n,m) + Q^{BF}(n,m)w^B(n,m)]}{\sum_{n=1}^{N} \sum_{m=1}^{M} (w^A(n,m) + w^B(n,m))}
\]

In Eq. (20), \( A \) and \( B \) denotes the source images and \( f \) denotes the final fused image. \( Q^{AF} \) and \( Q^{BF} \) represents amount of edge information preserved in \( F \) from image \( A \) and that from image \( B \) respectively. \( w^A \) and \( w^B \) are weights derived by convolving sobel operator with images \( A \) and \( B \). \( Q^{AB} \) varies in the range \([0,1]\) where a value of 1 corresponds to the best performance.

3. \( Q_0 \) [28]: This metric is designed by modeling any image distortion as a combination of three factors, namely, loss of correlation, luminance distortion, and contrast distortion. The value of \( Q_0 \) between source images \( A \) and \( B \) fused image \( F \) is expressed as:

\[
Q_0(A, B, F) = \frac{1}{2} \left[ Q_0(A, F) + Q_0(B, F) \right]
\]

\[ Q_0(A, F) = \frac{\sigma_d}{\sigma_d + \sigma_f} \cdot \frac{2a_f}{a^2 + f^2} = \frac{2\sigma_a}{\sigma_a + \sigma_f}
\]

where \( Q_0(A, F) \) is defined as:

\[
Q_0(A, F) = \frac{\sigma_d}{\sigma_d + \sigma_f} \cdot \frac{2a_f}{a^2 + f^2} = \frac{2\sigma_a}{\sigma_a + \sigma_f}
\]
Here, $\sigma_a$ and $\sigma_f$ are standard deviations of input image $A$ and fused image $F$; $\sigma_{af}$ denotes the covariance between $A$ and $F$. The dynamic range of $Q_0(A, B, F)$ is $[-1, 1]$ with best possible value as 1.

Three metrics RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and CC (Cross Correlation) as in [19], are employed to evaluate the performance of the proposed fusion method in thermal spectrum and are described below.

4. RMSE: The Root Means Square Error between the fused image $F$ and reference ground truth image $R$ is given by.

$$RMSE = \sqrt{\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (R(i,j) - F(i,j))^2}$$  \hspace{1cm} (23)

Here, $NM$ is the number of pixels in the image. Lower the value of RMSE, better is the performance of fusion.
5. MAE: The Mean Absolute Error between the fused image $F$ and reference ground truth image $R$ is given by.

$$MAE = \sqrt{\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} |R(i,j) - F(i,j)|}$$

(24)

Lower the value of MAE, better is the performance of fusion.

6. CC: The Cross Correlation between the fused image $F$ and reference ground truth image $R$ can be expressed as.

$$CC = \frac{2\sum_{i=1}^{N} \sum_{j=1}^{M} R(i,j)F(i,j)}{\sum_{i=1}^{N} \sum_{j=1}^{M} R(i,j)^2 + \sum_{i=1}^{N} \sum_{j=1}^{M} F(i,j)^2}$$

(25)

The dynamic range of $CC$ is $[0,1]$ with best possible value as 1.
Fig. 7. Specimen focus curves for ‘Loudspeaker’ and ‘Mixer’ image sets in VIS spectrum for different threshold values.

Fig. 8. Specimen focus curves for ‘Head’ and ‘Office desk’ image sets in NIR spectrum for different threshold values.

Fig. 9. Specimen focus curves for ‘Circuit breakers’ and ‘Circuit’ image sets in TH spectrum for different threshold values.
Fig. 10. Specimen focus curves for ‘Loudspeaker’ and ‘Mixer’ image sets in VIS spectrum for different number of orientations.

Fig. 11. Specimen focus curves for ‘Head’ and ‘Office desk’ image sets in NIR spectrum for different number of orientations.

Fig. 12. Specimen focus curves for ‘Circuit breakers’ and ‘Circuit’ image sets in TH spectrum for different number of orientations.
Fig. 13. Specimen focus curves for 'Loudspeaker' and 'Mixer' image sets in visual spectrum.

Fig. 14. Specimen focus curves for 'Head' and 'Office desk' image sets in near-infrared spectrum.

Fig. 15. Specimen focus curves for 'Circuit breakers' and 'Circuit' image sets in thermal spectrum.
5.2. Selection of threshold and number of orientations

We perform experiments to judiciously select the number of orientations \(O\) and the threshold \(T\) in the proposed method.

5.2.1. Selection of threshold

We have experimentally obtained the best performing value of threshold parameter \(T\). For the range of \(T\), we used \([\text{min}, \text{max}]\) of max-pooled local frequency map. \(T\) is set from the above range based on the performance of the focus curves \(Cs[N]\) in terms of Accuracy, Width at 50% maximum and Number of local maxima. Some sample focus curves obtained \((Cs[N])\) for five different threshold values \((T1, T2, T3, T4, T5)\) in this range are shown in Figs. 7–9. The focus measure curves reveal an interesting trend. The curves are almost comparable with varying values of threshold indicating the robustness of the proposed focus measure. However, in terms of Accuracy and Width at 50% maximum, the experimentally selected value of \(T3 = 0.0607\) emerges as an optimal choice for the threshold.

5.2.2. Selection of Number of orientations

Features in an image can be oriented at any angle \(h\) within the range \(0–180^\circ\) \([25]\). Selection of number of intervals for orientations influences the detection of oriented features in the input image. Note that less number of intervals may fail to capture the finer oriented features present in the image. On other hand, use of large number of intervals for orientations can be unreliable (sensitive to noise) in addition to increasing the computational overhead.

### Table 1

<table>
<thead>
<tr>
<th>Spectrum</th>
<th>Object</th>
<th>Metric</th>
<th>Proposed method</th>
<th>FH</th>
<th>FAST</th>
<th>HL</th>
<th>EOL</th>
<th>SML</th>
<th>SF</th>
<th>(P) (subjective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIS</td>
<td>Guitar</td>
<td>(P)</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Head-Phones</td>
<td>(P)</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Keyboard</td>
<td>(P)</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Keys</td>
<td>(P)</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Loudspeaker</td>
<td>(P)</td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Mixer</td>
<td>(P)</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>11</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Sunglass</td>
<td>(P)</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Spectrum</th>
<th>Object</th>
<th>Metric</th>
<th>Proposed method</th>
<th>FH</th>
<th>FAST</th>
<th>HL</th>
<th>EOL</th>
<th>SML</th>
<th>SF</th>
<th>(P) (subjective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIR</td>
<td>Building</td>
<td>(P)</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>20</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Car</td>
<td>(P)</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>20</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Corridor</td>
<td>(P)</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>20</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Head</td>
<td>(P)</td>
<td>5</td>
<td>4</td>
<td>19</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Keyboard</td>
<td>(P)</td>
<td>5</td>
<td>4</td>
<td>16</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Office desk</td>
<td>(P)</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Pens</td>
<td>(P)</td>
<td>8</td>
<td>7</td>
<td>18</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Spectrum</th>
<th>Object</th>
<th>Metric</th>
<th>Proposed method</th>
<th>FH</th>
<th>FAST</th>
<th>HL</th>
<th>EOL</th>
<th>SML</th>
<th>SF</th>
<th>(P) (subjective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TH</td>
<td>Circuit Breakers</td>
<td>(P)</td>
<td>19</td>
<td>17</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Building</td>
<td>(P)</td>
<td>26</td>
<td>25</td>
<td>3</td>
<td>23</td>
<td>12</td>
<td>12</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Circuit</td>
<td>(P)</td>
<td>7</td>
<td>6</td>
<td>26</td>
<td>5</td>
<td>4</td>
<td>26</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Engine</td>
<td>(P)</td>
<td>15</td>
<td>15</td>
<td>0</td>
<td>16</td>
<td>14</td>
<td>14</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Printer</td>
<td>(P)</td>
<td>25</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Server</td>
<td>(P)</td>
<td>20</td>
<td>21</td>
<td>3</td>
<td>18</td>
<td>20</td>
<td>6</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Tube</td>
<td>(P)</td>
<td>6</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>
So, as a trade-off, five intermediate choices O1 (7 orientations in step of 30°), O2 (10 orientations in step of 20°), O3 (13 orientations in step of 15°), O4 (16 orientations in step of 12°), O5 (19 orientations in step of 10°) are used with a fixed threshold (T).

These focus measure curves are evaluated in terms of Accuracy, Width at 50% maximum and Number of local maxima of the focus curves obtained ($C_{176}$) [12]. In the visible (VIS) spectrum the performance is uniform with respect to Accuracy but in terms of Width at 50% maximum O3 yields better performance. As can be seen from the focus curves in the NIR spectrum, O3 performs better. For example, in case of ‘Office Desk’ image set O3 produces

---

**Table 4**

Fusion results of proposed method: MI, $Q_{AB/f}$ and $Q_0$ values with and without Consistency Verification (CV).

<table>
<thead>
<tr>
<th>Images</th>
<th>MI Without CV</th>
<th>MI With CV</th>
<th>$Q_{AB/f}$ Without CV</th>
<th>$Q_{AB/f}$ With CV</th>
<th>$Q_0$ Without CV</th>
<th>$Q_0$ With CV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multifocus (VIS spectrum)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clock</td>
<td>8.5045</td>
<td>8.5563</td>
<td>0.6179</td>
<td>0.6701</td>
<td>0.9783</td>
<td>0.9785</td>
</tr>
<tr>
<td>Desk</td>
<td>7.9716</td>
<td>8.0246</td>
<td>0.5979</td>
<td>0.6697</td>
<td>0.9585</td>
<td>0.9586</td>
</tr>
<tr>
<td>Lab</td>
<td>8.3728</td>
<td>8.4551</td>
<td>0.6160</td>
<td>0.6838</td>
<td>0.9758</td>
<td>0.9759</td>
</tr>
<tr>
<td>Pepsi</td>
<td>8.2292</td>
<td>8.2621</td>
<td>0.6344</td>
<td>0.6838</td>
<td>0.9810</td>
<td>0.9810</td>
</tr>
<tr>
<td><strong>Multifocus (NIR spectrum)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keyboard</td>
<td>7.9217</td>
<td>7.9342</td>
<td>0.6561</td>
<td>0.6853</td>
<td>0.9907</td>
<td>0.9908</td>
</tr>
<tr>
<td>Medical</td>
<td>7.0279</td>
<td>7.0278</td>
<td>0.6677</td>
<td>0.6870</td>
<td>0.5010</td>
<td>0.5028</td>
</tr>
</tbody>
</table>

---

**Table 5**

Multifocus image fusion: Performance comparison with (a) FH IPD based method and (b) best results of multi-resolution based fusion methods.

<table>
<thead>
<tr>
<th>Spectrum</th>
<th>Method</th>
<th>MI (mean)</th>
<th>$Q_{AB/f}$ (mean)</th>
<th>$Q_0$ (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual (VIS spectrum)</td>
<td>Proposed method</td>
<td>8.3245</td>
<td>0.6768</td>
<td>0.9735</td>
</tr>
<tr>
<td></td>
<td>FH</td>
<td>8.2493</td>
<td>0.5804</td>
<td>0.9746</td>
</tr>
<tr>
<td></td>
<td>DWT</td>
<td>2.4126</td>
<td>0.6866</td>
<td>0.7206</td>
</tr>
<tr>
<td></td>
<td>SWT</td>
<td>2.4510</td>
<td>0.7140</td>
<td>0.7555</td>
</tr>
<tr>
<td></td>
<td>DTCWT</td>
<td>2.4814</td>
<td>0.7231</td>
<td>0.7650</td>
</tr>
<tr>
<td></td>
<td>CVT</td>
<td>2.4387</td>
<td>0.7075</td>
<td>0.7421</td>
</tr>
<tr>
<td></td>
<td>CT</td>
<td>2.3978</td>
<td>0.6700</td>
<td>0.7076</td>
</tr>
<tr>
<td></td>
<td>NSCT</td>
<td>2.4804</td>
<td>0.7219</td>
<td>0.7799</td>
</tr>
<tr>
<td>Near-infrared (NIR spectrum)</td>
<td>Proposed method</td>
<td>7.9342</td>
<td>0.6853</td>
<td>0.9908</td>
</tr>
<tr>
<td></td>
<td>FH</td>
<td>8.0198</td>
<td>0.7105</td>
<td>0.9912</td>
</tr>
<tr>
<td></td>
<td>DWT</td>
<td>5.9485</td>
<td>0.5135</td>
<td>0.9061</td>
</tr>
<tr>
<td></td>
<td>DTCWT</td>
<td>7.3575</td>
<td>0.7082</td>
<td>0.9902</td>
</tr>
</tbody>
</table>

---

![Fig. 16. Fused images obtained using the proposed method in the visual spectrum for four datasets performed without consistency verification (FI) and with consistency verification (FI-CV). (a) ‘Clock’ FI, (b) ‘Clock’ FI-CV; (c) ‘Desk’ FI, (d) ‘Desk’ FI-CV; (e) ‘Lab’ FI, (f) ‘Lab’ FI-CV; (g) ‘Pepsi’ FI, and (h) ‘Pepsi’ FI-CV.](image-url)
peak at image index 7 which is nearest to the index of image having highest focus from subjective assessment test given in [8]. In the TH spectrum there are a number of local maxima for each number of orientations due to limited resolution. But based on Width at 50% maximum and Accuracy it is clear that O3 performs better. So, we have used O3 (13 orientations in step of 15°) for our work.

5.3. Performance analysis of the multispectral focus measure

We compare our proposed focus measure with other such measures from two categories, namely, (i) the standard intensity driven focus measures and (ii) the recently developed IPD based focus measures. The first category includes Energy of Laplacian (EOL), Sum Modified Laplacian (SML) and Spatial Frequency (SF). From the second category we compare with Fast Hessian (FH), Harris-Laplace (HL) and Features from Accelerated Segment Test (FAST) [9]. The standard focus measures are known to yield good results. EOL and its modified adaptation SML are high performing derivative based focus measures. On the other hand IPD based focus measures perform relatively well when applied to multispectral images.

We now show focus curves for the proposed focus measure for datasets from the different spectra. Please see Fig. 13 for the focus curves of the ‘Loudspeaker’ and the ‘Mixer’ datasets.
in the visual spectrum, Fig. 14 for the focus curves of the ‘Head’ and the ‘Office desk’ datasets in the near-infrared spectrum, and, Fig. 15 for the focus curves of the ‘Circuit breakers’ and the ‘Circuit’ datasets in the thermal spectrum. A good focus measure possesses the characteristics of unimodality, monotonicity and is sensitive to defocus [9]. Our method exhibits all the desirable characteristics warranted of a good focus measure. Focus curves evaluated for the proposed focus measure reach a global maximum and decrease monotonically as the defocus increases on either side. However, a few false maxima and minima are observed in the focus curves of thermal images because of poor resolution due to limited focal length. It is reported in [9] that the interest point detectors perform dismally in the visual spectrum as compared to the standard focus measures. However, their performance is improved substantially in other spectra (thermal, near-infrared). Our interest point based focus measure shows decent performance across all the spectra. Comparative results for the visual spectra are shown in Table 1, EOL, SML and SF perform well for most of the datasets. Compared

![Fused images obtained in the near-infrared (NIR) spectrum performed without consistency verification (FI) and with consistency verification (FI-CV).](image)

**Fig. 20.** Fused images obtained in the near-infrared (NIR) spectrum performed without consistency verification (FI) and with consistency verification (FI-CV). (a) Proposed method FI, (b) proposed method FI-CV, (c) FH IPD based method FI, (d) FH IPD based method FI-CV, (e) DWT method and (f) DTCWT method.
Fig. 21. Ground truth (GT) and fused images (FI) obtained using the proposed method in the thermal spectrum for five datasets. (a) Mobile_RS232 GT, (b) Mobile_RS232 FI; (c) Bulbs Set 1 GT, (d) Bulbs Set 1 FI; (e) Bulbs Set 2 GT, (f) Bulbs Set 2 FI; (g) Bulbs Set 3 GT, (h) Bulbs Set 3 FI; (i) Bulbs Set 4 GT, and (j) Bulbs Set 4 FI.
to other interest point detectors, the proposed IPD based focus measure outperforms FAST, FH and HL in most of the cases. The proposed method is comparable with subjective analysis in terms of the $P$ metric as reported in [9]. Results for the near-infrared spectrum in Table 2 show significant improvements in the performance over other interest point detector based focus measures as well as standard focus measures. The Keyboard dataset belonging to near-infrared spectrum reveals that the proposed focus measure yields a $Q$ value of 0.5 whereas the reported $Q$ values of FH, FAST and HL are 0.25, 0.10, and 0.1250 respectively. Our $Q$ value is comparable to that of EOL and is better than SML and SF with reported values as 0.0830 and 0.3333. We outperform SML, SF in most of the cases. We have outdone FH in all the cases and perform much better compared to FAST and HL. Comparative analysis reveals that the performance of the proposed focus measure is best for the thermal images as shown in Table 3. Our detector performs better than SML, SF, FAST and HL and is comparable to EOL and FH. In the 'Circuit breakers' dataset, the best performing interest point detector based focus measure is FH and that from the standard focus measures is EOL, each having a $Q$ value of 0.5. The proposed focus measure having $Q$ value of 1.0 easily surpasses both. The superior performance of our focus measure as compared to the standard interest point detector based focus measures is due to use of phase information at various orientations.

Average execution time of obtaining the focus measure using the proposed method is 2 s. on a desktop PC with 3.4 GHz Intel Core CPU and 8 GB RAM. Our method tends to be slower compared to some of the other focus measures because we have to compute the local frequency map at thirteen different orientations. However, please note that Minhas et al. in [13] have reported an average execution time for their orientation-based focus measure to be 3.5 s for the same window size of $7 \times 7$ as ours.

5.4. Performance analysis of multifocus fusion results

In regards to multifocus fusion of images in the visual spectrum, we compare our method with the highly accurate multiresolution transform domain methods such as Discrete Wavelet Transform (DWT), Stationary Wavelet Transform (SWT), Curvelet Transform (CVT), Contourlet Transform (CT), Dual Tree Complex Wavelet

Fig. 22. Fused images obtained using the Fast Hessian (FH) IPD based method in the thermal spectrum for five datasets. (a) Mobile_RS232, (b) Bulbs Set 1, (c) Bulbs Set 2, (d) Bulbs Set 3, and (e) Bulbs Set 4.
Transform (DTCWT) and Non-Subsampled Contourlet Transform (NSCT) [17,18,29]. Since our fusion method is essentially based on interest point detection, we also compare our method with best performing Fast Hesian (FH) based fusion scheme from the same category [9]. For thermal fusion, we choose a recently reported EOL based method [19] for comparison. In addition, we also compare with FH based fusion scheme. Overall, we provide an extensive comparison with several recent and well-known spatial and transform domain based fusion methods.

Table 4 shows that the performances are improved with the inclusion of Consistency Verification (CV) for the proposed method. Fig. 16 qualitatively demonstrates the same results. Since only one image set pertaining to medical database is available, we just specify the result of the proposed method without any comparison. Next, in Table 5, we show that our method performs better compared to all the multi resolution transform based methods [17], in terms of a much higher MI, slightly higher Qo and comparable with Q40f (only marginally lower). Comparison to FH based fusion scheme reveals an improvement in terms of MI and Q40f values.

In VIS spectrum, for perceptual quality evaluation of fused images, we incorporate fused images obtained using FH, DWT and DTCWT methods (see Figs. 17–19). DWT is a basic method in VIS spectrum, for perceptual quality evaluation of fused images, we incorporate fused images obtained using FH, DWT and DTCWT methods (see Figs. 17–19). DWT is a basic method while it is comparable in case of

Next, in Table 5, we show that our method performs better compared to all the multi resolution transform based methods (see Figs. 17–19). DWT is a basic method in VIS spectrum, for perceptual quality evaluation of fused images, we incorporate fused images obtained using FH, DWT and DTCWT methods (see Figs. 17–19). DWT is a basic method while it is comparable in case of

Table 2 shows that the performances are improved with the inclusion of Consistency Verification (CV) for the proposed method. Fig. 16 qualitatively demonstrates the same results. Since only one image set pertaining to medical database is available, we just specify the result of the proposed method without any comparison. Next, in Table 5, we show that our method performs better compared to all the multi resolution transform based methods [17], in terms of a much higher MI, slightly higher Qo and comparable with Q40f (only marginally lower). Comparison to FH based fusion scheme reveals an improvement in terms of MI and Q40f values.

In VIS spectrum, for perceptual quality evaluation of fused images, we incorporate fused images obtained using FH, DWT and DTCWT methods (see Figs. 17–19). DWT is a basic method in VIS spectrum, for perceptual quality evaluation of fused images, we incorporate fused images obtained using FH, DWT and DTCWT methods (see Figs. 17–19). DWT is a basic method while it is comparable in case of

Table A2

Table A2 – Y-Separable basis set and interpolation functions for fourth derivatives of Gaussian.

| Hf  | = 0.3975/7.1838 – 7.5013 x + 3.7 x 2 | Ke(θ) = cos 3(θ) | Kd(θ) = sin 3(θ) |
| Hxf | = 0.3975/1.438 – 4.5013 y + 3.7 y 2 | Ke(θ) = −5cos 3(θ)sin(θ) | Kd(θ) = 10cos 3(θ)sin 3(θ) |
| Hyc | = 0.3975 x 2 – 2.2257x y + 0.6638ye 4(−x 3 ) | Ke(θ) = 3cos 3(θ)sin 4(θ) | Kd(θ) = 10cos 3(θ)sin 3(θ) |
| Hyc | = 0.3975/1.438 – 4.5013 y 2 + 3.7 ye 4(−x 3 ) | Ke(θ) = 5cos(θ)sin 3(θ) | Kd(θ) = cos(θ)sin 3(θ) |
| Hyc | = 0.3975/7.1838 – 7.5013 x + 3.7 x 2 | Ke(θ) = −sin 3(θ) | Kd(θ) = sin 3(θ) |

Acknowledgements

The authors would like to thank Zukal M., Melkysa J., Cika P., Smekal Z. of Brno University of Technology, Czech Republic for providing access to the multispectral database (http://splab.cz/en/download/database/multispec). Authors are also grateful to Espino-osa-Duró V. from EUP Mataró, Barcelona, Spain, for granting access to the thermal multifocus image datasets (http://splab.cz/en/download/database multi-focus-thermal-image-database).

Appendix A

We have included Tables A1 and A2 from [24] which were used for the computation of the oriented analytic image.

References


