Available online at www.sciencedirect.com
ScienceDirect
Information Fusion 8 (2007) 177–192
INFORMATION FUSION
www.elsevier.com/locate/inffus

A new metric based on extended spatial frequency and its application to DWT based fusion algorithms

Yufeng Zheng [a,*], Edward A. Essock [a,b], Bruce C. Hansen [a], Andrew M. Haun [a]

[a] Department of Psychological & Brain Sciences, University of Louisville, Louisville, KY 40292, USA
[b] Department of Ophthalmology & Visual Sciences, University of Louisville, Louisville, KY, 40292, USA

Received 12 May 2004; received in revised form 29 April 2005; accepted 29 April 2005
Available online 15 July 2005

Abstract

A new quantitative metric is proposed to objectively evaluate the quality of fused imagery. The measured value of the proposed metric is used as feedback to a fusion algorithm such that the image quality of the fused image can potentially be improved. This new metric, called the ratio of spatial frequency error (rSFe), is derived from the definition of a previous measure termed "spatial frequency" (SF) that reflects local intensity variation. In this work, (1) the concept of SF is first extended by adding two diagonal SFs, then, (2) a reference SF (SF_R) is computed from the input images, and finally, (3) the error SF (SF_E) (subtracting the fusion SF from the reference SF), or the ratio of SF error ($rSFe = SF_E/SF_R$), is used as a fusion quality metric. The rSFe (which can be positive or negative) indicates the direction of fusion error—over-fused (if $rSFe > 0$) or under-fused (if $rSFe < 0$). Thus, the rSFe value can be back propagated to the fusion algorithm (BP fusion), thereby directing further parameter adjustments in order to achieve a better-fused image. The accuracy of the rSFe is verified with other quantitative measurements such as the root mean square error (RMSE) and the image quality index (IQI), as well as with a qualitative perceptual evaluation based on a standard psychophysical paradigm. An advanced wavelet transform (*a*DWT) method that incorporates principal component analysis (PCA) and morphological processing into a regular DWT fusion algorithm is implemented with two adjustable parameters—the number of levels of DWT decompositions and the length of the selected wavelet. Results with *a*DWT were compared to those with a regular DWT and with a Laplacian pyramid. After analyzing several inhomogeneous image groups, experimental results showed that the proposed metric, rSFe, is consistent with RMSE and IQI, and is especially powerful and efficient for realizing the iterative BP fusion in order to achieve a better image quality. Human perceptual assessment was measured and found to strongly support the assertion that the *a*DWT offers a significant improvement over the DWT and pyramid methods.

Keywords: Back-propagation (BP); Fusion evaluation; Image fusion; Image pyramid; Wavelet transform (DWT); Spatial frequency

1. Introduction

Image fusion is a tool that serves to combine multiple-source imagery by using advanced image processing techniques. Specifically, it aims for the integration of disparate and complementary data in order to enhance the information apparent in the respective source images, as well as to increase the reliability of interpretation. This leads to more accurate data [1] and increased utility [2,3]. In addition, it has been stated that fused data provides for more robust aspects of operational performance such as increased confidence, reduced ambiguity, improved reliability and improved classification [2,4].

This paper focuses on the so-called "pixel-level" fusion process, where a composite image has to be built of several (typically two) input images. A general framework of image fusion can be found in [5]. In pixel-level

* Corresponding author. Tel.: +1 502 85 24056; fax: +1 502 85 28904.
E-mail address: yufeng.zheng@louisville.edu (Y. Zheng).

1566-2535/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.inffus.2005.04.003

image fusion, some general requirements [6] are imposed on the fusion result: (1) the fusion process should preserve all relevant information of the input imagery in the composite image (pattern conservation); and (2) the fusion scheme should not introduce any artifacts or inconsistencies which would distract the human observer or subsequent processing stages. Consequently, quantitative evaluation of the quality of fused imagery is considered very important for an objective comparison of the performance of the respective fusion algorithms. In addition, a quantitative metric may potentially be used as feedback to the fusion algorithm to further improve the fused image quality.

Through the practical applications of image fusion in medical imaging, remote sensing, nighttime operations and multi-spectral imaging, many fusion algorithms have been developed. Two common fusion methods are the discrete wavelet transform (DWT) [7–10] and various pyramids (such as Laplacian, contrast, gradient and morphological pyramids) [11,12]. However, only a few metrics are available for quantitative evaluation of the quality of fused imagery. For example, the root mean square error (RMSE) is the natural measure of image quality if there is a "ground truth" image available; however, for realistic image fusion problems there are no ground truths. Beauchemin et al. [13] presented a method using local variance for image fusion assessment that still requires a comparison with the measurement of ground truth. Leung et al. [14] proposed the image noise index (INI), based on entropy calculation, to measure fused image quality; this method requires the exact reverse process of an image fusion procedure, which is impractical for most fusion processes such as DWT or pyramid methods. Piella and Heijmans [15] recently presented a new metric for image fusion—the image quality index (IQI), which measures how similar the fused image is to both input images. The values of IQI are within $[0,1]$ and have an ideal value of 1 (if two input images are identical). The IQI metric has been shown to be consistent with other methods for evaluating fused image quality [16]. However, the IQI does not give the error direction, which would indicate whether the fused image is under- or over-fused.

An image measure termed "spatial frequency" (SF) [20] can indicate how much information is contained in an image and thus may be used as a fusion rule (determining which input should be selected in the fused image) [17], but it cannot directly be used to measure the fused image quality. From the definition of SF (given in Section 2.3), we know that the SF metric is sensitive to gray level changes, although it cannot distinguish useful information from noise or artifacts. Thus, to use the spatial frequency value as a fusion metric, a standard or reference value computed from the input images must be constructed for the purpose of comparison. The new image quality metric presented in this paper, termed as "the ratio of SF error (rSFe)", is a relative measurement regardless of the type of image to be analyzed. The accuracy of the rSFe metric can be verified with the currently used measurements of RMSE and IQI. In addition, the rSFe value can be further back propagated to the fusion algorithm (BP fusion) such that the fusion parameter adjustment is directed to perform the next loop of fusion, a process that is repeated until an optimized fusion result is achieved.

An advanced wavelet transform (*a*DWT) method that incorporates principal component analysis (PCA) and morphological processing into a regular DWT fusion algorithm was recently presented [16]. Furthermore, in [16], experimental results showed an important relationship between the fused image quality and the wavelet properties, that is, a higher level of DWT decomposition (with smaller image resolution at higher scale) or a lower order of wavelets (with shorter length) usually resulted in a more sharpened fused image. This means that we can use the level of DWT decomposition and the length of a wavelet as control parameters of an iterative DWT-based fusion algorithm. Together with the rSFe metric, an iterative BP-*a*DWT can be realized. In the current experiments, a regular DWT and also a Laplacian pyramid (which has been shown to be better in fusing images when compared to other pyramids [16]) were also implemented with the BP-fusion procedure and both methods were compared to the results of BP-*a*DWT. To further verify the evaluation results of quantitative metrics, a qualitative experiment carried out with human observers was designed and analyzed.

The subsequent sections of this paper are organized as follows. The currently used metrics—RMSE, IQI and SF are described in Section 2, where the new metric, rSFe, is also introduced. Next is a full description of Laplacian pyramid method, the *a*DWT, and the iterative BP-*a*DWT directed by the rSFe metric. Lastly, the experimental results and discussion of both quantitative and qualitative analyses are presented, followed by conclusions.

2. Image quality metrics

As mentioned in the introduction, the general requirements of an image fusion process are that it should preserve all valid and useful pattern information from the source images, while at the same time it should not introduce artifacts that could interfere with subsequent analyses [18]. However, it is nearly impossible to combine images without introducing some form of distortion. In the current body of literature, image fusion results tend to be evaluated either subjectively (with human assessment) or objectively (with quantitative metrics) [3,4]. In this section we focus on quantitative

measures (i.e., objective) that could be carried out automatically by computers. Qualitative (subjective) evaluation perceived by human observers will be described in Section 4.2. Of the three commonly used performance measures described below, RMSE needs a reference (ground truth) image while the others do not. But it should be noted that ideally, the "best criterion" should always be linked appropriately with a specific application.

2.1. Root mean square error

The root mean square error (RMSE) between the reference image (ground truth) and the fused image is

$$\mathrm{RMSE} = \left(\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[I_{\mathrm{R}}(i,j) - I_{\mathrm{F}}(i,j)]^2}{M \times N} \right)^{\frac{1}{2}}, \tag{1}$$

where $I_{\mathrm{R}}(i,j)$ and $I_{\mathrm{F}}(i,j)$ are the image pixel values of the reference image and the fused image respectively. $M \times N$ is the image size.

2.2. Image quality index

The image quality index (IQI) was recently introduced by Wang and Bovik [19]. Given two sequences $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, let $\bar{x}$ denote the mean of x, and σ_x and σ_{xy} denote the variance of x and covariance of x and y, respectively. The global quality index of two vectors is defined as

$$Q_0(x,y) = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\bar{x}^2 + \bar{y}^2)(\sigma_x^2 + \sigma_y^2)} \tag{2}$$

which can be decomposed as

$$Q_0(x,y) = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \cdot \frac{2\bar{x}\bar{y}}{(\bar{x}^2 + \bar{y}^2)} \cdot \frac{2\sigma_x\sigma_y}{(\sigma_x^2 + \sigma_y^2)}. \tag{3}$$

Note that the first component in Eq. (3) is the correlation coefficient between x and y. This value is a measure for the similarity of the vectors x and y, and takes values between -1 and 1. Keep in mind that in this case (image quality evaluation), the values x_i, y_i are positive gray-scale values. The second component in Eq. (3) corresponds to the luminance distortion, which has a dynamic range of $[0,1]$. The third factor in Eq. (3) measures the contrast distortion and its range is also $[0,1]$. In summary, $Q_0 \in [0,1]$, and the maximum value $Q_0 = 1$ is achieved when x and y are identical.

Piella and Heijmans [15] introduced a weighting procedure into Q_0 calculation. The weight should reflect the local relevance of an input image that may depend on its local variance, contrast, sharpness, or entropy. Given the local salience (e.g., local variance) of two input images A and B, we compute a local weight λ indicating the relative importance of image A compared to image B: the larger λ, the more weight is given to image A. A typical choice for λ is

$$\lambda = S(I_{\mathrm{A}})/[S(I_{\mathrm{A}}) + S(I_{\mathrm{B}})], \tag{4}$$

where $S(I_{\mathrm{A}})$ and $S(I_{\mathrm{B}})$ denote the salience of input image A and B, respectively. Then, the weighted image quality index (IQI) can be defined as

$$Q_{\mathrm{w}} = \lambda Q_0(I_{\mathrm{A}}, I_{\mathrm{F}}) + (1-\lambda) Q_0(I_{\mathrm{B}}, I_{\mathrm{F}}). \tag{5}$$

In fact, the IQI metric measures the similarity between the fused image (I_{F}) and both of the input images (I_{A} and I_{B}) by assuming that an ideally fused image should resemble both original input images. Piella also suggested using local variance as the salience of an image, i.e., $S(I_{\mathrm{A}}) = \sigma_{\mathrm{A}}$. Since image signals are generally non-stationary, it is more appropriate to measure the weighted image quality index Q_{w} over local regions (e.g., by parsing the entire image into a set of 'blocks') and then combine these local results into a single measure as the global measure of the entire image. For example, in our experiments, each image (I_{A}, I_{B} or I_{F}) is divided into a certain number of blocks by designating the block size of 16×16 pixels, Q_0 is computed for each small block (i.e., a 16×16 window), then Q_0's are summed across all blocks and the average is taken.

2.3. Spatial frequency

The metric "spatial frequency" (SF) [17,20] is used to measure the overall activity level of an image, which is defined as follows:

$$\mathrm{SF} = \sqrt{(\mathrm{RF})^2 + (\mathrm{CF})^2}, \tag{6}$$

where RF and CF are row frequency and column frequency respectively.

$$\mathrm{RF} = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}[I(i,j) - I(i,j-1)]^2}, \tag{7a}$$

$$\mathrm{CF} = \sqrt{\frac{1}{MN}\sum_{j=1}^{N}\sum_{i=2}^{M}[I(i,j) - I(i-1,j)]^2}, \tag{7b}$$

Notice that the term "spatial frequency" that is computed in the spatial domain as defined in Eqs. (6) and (7), does not correspond with the homonymous term found in the context of the Fourier transform, where spatial frequency is measured in the frequency domain in units of "cycles per degree" or "cycles per millimeter".

2.4. The ratio of SF error

Similar to the definitions of RF and CF, spatial frequency along two diagonal directions (see Fig. 1b),

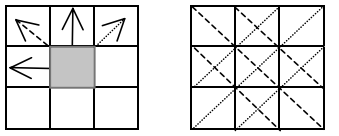

Fig. 1. Illustration of spatial frequency (SF) calculation: each small square denotes a pixel. (a) Four directional SFs: arrows are pointed to the gradient directions. (b) Two diagonal directions: the dashed lines stand for main diagonal directions, while the dotted lines represent secondary diagonal directions.

termed as main diagonal SF (MDF) and secondary diagonal SF (SDF), can be defined as below

$$\mathrm{MDF} = \sqrt{w_{\mathrm{d}} \cdot \frac{1}{MN} \sum_{i=2}^{M} \sum_{j=2}^{N} [I(i,j) - I(i-1,j-1)]^2}, \tag{8a}$$

$$\mathrm{SDF} = \sqrt{w_{\mathrm{d}} \cdot \frac{1}{MN} \sum_{j=1}^{N-1} \sum_{i=2}^{M} [I(i,j) - I(i-1,j+1)]^2}, \tag{8b}$$

where $w_{\mathrm{d}} = 1/\sqrt{2}$ is a distance weight; similarly it can be considered that $w_{\mathrm{d}} = 1$ in Eqs. (7).

Then the overall spatial frequency of an image becomes

$$\mathrm{SF} = \sqrt{(\mathrm{RF})^2 + (\mathrm{CF})^2 + (\mathrm{MDF})^2 + (\mathrm{SDF})^2} \tag{9}$$

which is a combination of four directional SFs (see Fig. 1a).

With Eq. (9) we can calculate the SFs of input images ($\mathrm{SF_A}$ and $\mathrm{SF_B}$) or of the fused image ($\mathrm{SF_F}$). Now we determine how to calculate a reference SF ($\mathrm{SF_R}$) with which the $\mathrm{SF_F}$ can be compared. The four differences (inside square brackets) defined in Eqs. (7) and (8) are actually the four first-order *gradients* along four directions (Fig. 1a) at that pixel, denoted as $\mathrm{Grad}(I(i,j))$. The four *reference gradients* can be obtained by taking the maximum of absolute gradient values between input image A and B along four directions:

$$\begin{aligned} &\mathrm{Grad^D}(I_{\mathrm{R}}(i,j)) \\ &\quad = \max\{\mathrm{abs}[\mathrm{Grad^D}(I_{\mathrm{A}}(i,j))], \mathrm{abs}[\mathrm{Grad^D}(I_{\mathrm{B}}(i,j))]\}, \\ &\quad \text{for each of four directions, i.e., } \mathrm{D} = \{\mathrm{H, V, MD, SD}\}, \end{aligned} \tag{10}$$

where the superscript 'D' denotes one of four directions (Horizontal, Vertical, Main Diagonal, and Secondary Diagonal). Substituting the differences (defined inside square brackets) in Eqs. (7) and (8) with $\mathrm{Grad^D}(I_{\mathrm{R}}(i,j))$, four directional reference SFs (i.e., $\mathrm{RF_R}$, $\mathrm{CF_R}$, $\mathrm{MDF_R}$ and $\mathrm{SDF_R}$) can be calculated. For example, the *reference row frequency* can be calculated as follows:

$$\mathrm{RF_R} = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=2}^{N} [\mathrm{Grad^H}(I_{\mathrm{R}}(i,j))]^2}. \tag{11}$$

Then, with Eq. (9), the $\mathrm{SF_R}$ can be computed. Note that the notation of "$\mathrm{Grad^H}(I_{\mathrm{R}}(i,j))$" is interpreted as "the horizontal reference gradient at point (i,j)", and no reference image is needed to compute the $\mathrm{SF_R}$ value.

Finally, the ratio of SF error (rSFe) is defined as follows:

$$\mathrm{rSFe} = (\mathrm{SF_F} - \mathrm{SF_R})/\mathrm{SF_R}. \tag{12}$$

Clearly, an ideal fusion has $\mathrm{rSFe} = 0$; that is, the smaller rSFe's absolute value, the better the fused image. Furthermore, $\mathrm{rSFe} > 0$ means that an over-fused image, with some distortion or noise introduced, has resulted; $\mathrm{rSFe} < 0$ denotes that an under-fused image, with some meaningful information lost, has been produced.

3. Image fusion methods

3.1. Laplacian pyramid

Image pyramids have been described for multi-resolution image analysis and have been proposed as a model for binocular fusion in human vision [18,21]. An image pyramid can be described as collection of low- or band-pass copies of an original image in which both the band-limit and sample density are reduced in regular steps. The basic strategy of image fusion based on pyramids is to use a feature selection rule to construct a fused pyramid representation from the pyramid representations of the original images. The composite image is obtained by taking the inverse pyramid transform. Several pyramid-based fusion schemes have been proposed in recent years [21–32]. Only the *Laplacian pyramid* is considered here as it has been shown to perform best among all pyramids [16].

A set of band-pass copies of an image is referred to as the Laplacian pyramid due to the similarity to a Laplacian operator. Each level of the Laplacian pyramid is recursively constructed from its lower level by applying the following four basic steps: blurring (low-pass filtering); subsampling (reduce size); interpolation (expand); and differencing (to subtract two images pixel by pixel) [22]. In the Laplacian pyramid, the lowest level of the pyramid is constructed from the original image. The Laplacian pyramid was first introduced as a model for binocular fusion in human stereo vision [18,21], where the implementation used a Laplacian pyramid and a maximum selection rule at each point of the pyramid transform.

3.2. The DWT and advanced DWT methods

As with a pyramid method, the regular DWT method is also a multi-scale analysis method. In a regular DWT fusion process, DWT coefficients from two input images are fused (pixel-by-pixel) by choosing the average of the approximation coefficients at the highest transform scale, and the larger absolute value of the detail coefficients at each transform scale. Then an inverse DWT is performed to obtain the fused image.

In the advanced DWT (*a*DWT) method, *principal component analysis* (PCA) [33] and *morphological processing* were incorporated into the DWT fusion algorithm. Specifically, at each DWT scale of a particular image, the DWT coefficients of a 2-D image consist of four parts: approximation, horizontal detail, vertical detail and diagonal detail. We apply PCA to two input images' approximation coefficients at the highest transform scale, fusing them by using the *principal eigenvector* (corresponding to the larger eigenvalue) derived from the two original images, as described in Eq. (13) below

$$C_F = (a_1 \cdot C_A + a_2 \cdot C_B)/(a_1 + a_2), \tag{13}$$

where C_A and C_B are approximation coefficients (at the highest transform scale) transformed from input images A and B. C_F represents the fused coefficients; a_1 and a_2 are the elements of principal eigenvector, which were computed by analyzing the original input images (Note: C_A and C_B are not analyzed because their sizes at the highest transform scale are too small to conduct an accurate result). In practice, we have to convert 2-D images into 1-D vectors by simply stacking each image column-by-column so that the principal component can be computed. Note that the denominator in Eq. (13) is used for normalization so that the fused image has the same energy distribution as original input images.

To show the superiority of applying PCA to the combination of two approximation coefficients (as defined in Eq. (13)), two examples are given below. (1) For the clock pair (as shown in Fig. 7 in Section 4.1.1), we have $a_1 = 0.69$ (49%), $a_2 = 0.72$ (51%), and (2) for the medical pair (CT and MRI, see Fig. 8 in Section 4.1.1), $a_1 = 0.066$ (6%), $a_2 = 0.998$ (94%), where the normalized weights are listed inside the parentheses. It is clear that a larger weight (94% after normalization) was given to the MRI image in fusing the medical image pair, which is reasonable because the MRI image contains more useful information than the CT image does. On the other hand, if two input images are equally important like in the clock image pair two approximately equal weights were assigned to both images (i.e., 49% and 51%). In the sense of weighting the combination according to the amount of information contained in the original images, PCA is superior to the averaging process in combining two approximation coefficients otherwise used in a regular DWT fusion procedure.

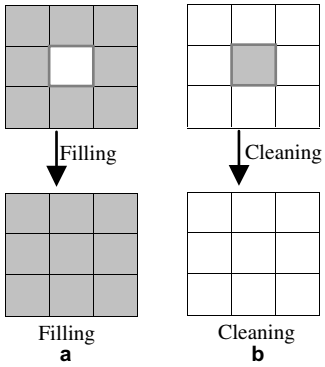

Fig. 2. Morphological processing (of a 3×3 region): (a) filling and (b) cleaning. Shaded pixels were chosen from image A, and white pixels were selected from image B.

For the detail coefficients (the other three quarters of the coefficients) at each transform scale, the larger absolute values were selected and followed by a neighborhood *morphological processing*, which served to verify the selected pixels by using a "filling" and "cleaning" operation (i.e., the operation fills or removes isolated pixels locally as shown in Fig. 2). Such an operation (similar to smoothing) can increase the consistency of coefficient selection thereby reducing the distortion in the fused image.

It is important to recognize that the highest DWT level (L_{dmax}) that can be decomposed depends on the input image size. However, the size of the smallest transformed image should not be less than 2×2. Thus we have

$$L_{dmax} = \text{int}\{\log_2[\min(M, N)/\min(m_0, n_0)]\}, \tag{14}$$

where the image size is $M \times N$, $m_0 \times n_0$ represents for the size of transformed image by DWT at the highest scale, "min" means taking the minimal value, and "int" stands for taking the integral part.

3.3. Iterative BP-aDWT directed by the rSFe

As shown in Fig. 3, we first propose a general iterative *back-propagation* (BP) model for any image fusion procedure. The term of back-propagation is borrowed from artificial neural network (ANN), where the output error is back propagated to the network (control unit) and used to adjust the neuron weights [34]. To make this model work properly, some requirements are needed: (a) the fusion process must have some adjustable control parameters that affect the metric value (i.e., with predictable variance); (b) there must be an ideal value for the metric that is used for judging the iteration convergence; (c) the error of the metric values (subtracting the current metric value from the ideal metric value) should indicate the error direction, which is used to direct the parameter adjustment (i.e., increment or decrement). Requirements

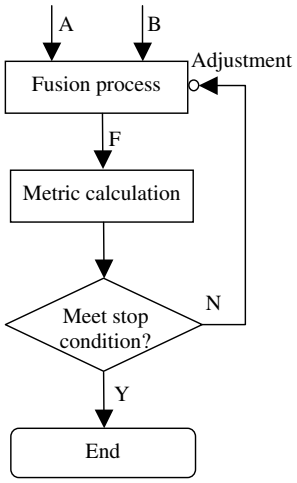

Fig. 3. Diagram of a general iterative back-propagation fusion process: A and B are input images, F is the fused image.

(a) and (b) are essential, (c) is optional but will expedite the convergence of iteration.

Previous experiments [16] have shown an important relationship between the fused image quality and the wavelet properties, specifically, that a higher-level DWT decomposition (with smaller image resolution at higher scale) or a lower order of wavelets (with shorter length) usually resulted in a more sharpened fused image. If measured with the rSFe metric, its value tends to be more positive for a more sharpened image. This means that we can use the level of DWT decomposition (L_d) and the length of a wavelet (L_w) as adjustable parameters of an iterative BP-*a*DWT algorithm. With the definition of rSFe, we know that it has an ideal value, 0; in addition, the sign of rSFe indicates an over-fused (if positive) or under-fused (if negative) status of the currently fused image. Thus an iterative BP-*a*DWT can be realized.

As indicated in Fig. 3, termination conditions to stop the fusion iteration are required. The fusion iteration will stop when one of the following conditions is met: (1) rSFe has converged at the ideal value (i.e., rSFe = 0)—the absolute value of (rSFe-0), abs(rSFe-0), is smaller than a designated tolerance error (say 0.001); (2) rSFe has passed the zero point—rSFe changes its sign in two successive iterations; (3) rSFe has converged at a fixed value other than the ideal value—there is no significant change of abs(rSFe) between two adjacent iterations (e.g., the change < 0.0005); (4) rSFe has not converged—abs(rSFe) is increasing or fluctuating but overall increasing for subsequent iterations; and (5) rSFe has not converged—parameters' boundaries are reached. In implementing the iteration of a BP-fusion procedure, legal boundaries of varying parameters should be designated based on the definition of parameters and the context. The details of implementation are depicted in the next section.

4. Experimental results and discussion

The following experiments were performed to investigate: (1) whether the metric of rSFe is consistent with other metrics such as RMSE and IQI; (2) whether there is an optimal value of rSFe among the results of varying two parameters (L_d and L_w); (3) whether the optimal value of rSFe can be easily achieved by separately varying the two parameters; (4) when judged with the metric IQI, how the iterative BP-*a*DWT performs compared to a regular DWT algorithm or Laplacian pyramid; and (5) whether the quantitative (objective) evaluation results are consistent with qualitative (subjective) evaluations based on human perceptual assessment. In the current experiments, we tested the three fusion algorithms and the three image quality measures mentioned above by using three types of imagery (already well registered and with 256 gray levels) which were as follows: (a) one simulated image pair with ground truth available, (b) three frequently used samples, and (c) three pairs of night-vision images.

4.1. Quantitative analyses and discussion

For imagery where the ground truth image is available, RMSE is the natural measure to indicate how close the fused image is to the ground truth. Since the potential artificial distortion during the fusion process can also increase the spatial frequency value of the fused image, thus it is not reasonable to directly measure or compare the spatial frequency values between input images and the fused image (refer to Table 3). For imagery without ground truths, however, the rSFe and IQI are relatively faithful, where the signed rSFe indicates how far away from the ideal fusion (rSFe = 0) and an under- or over-fusion status, while IQI gives a quantity (between 0 and 1) regarding the similarity between the fused image and the two input images.

In the iterative *a*DWT or regular DWT algorithms, "Symlets" were selected as the wavelet with adjustable parameters of the DWT decomposition level (L_d) and the wavelet length (L_w). Of course, for Laplacian pyramid, only the decomposition level (L_d) can be varied. The extreme values of two parameters are given as: $1 \leqslant L_d \leqslant L_{dmax}$ and $2 \leqslant L_w \leqslant 10$. The current set of quantitative experiments showed that functions rSFe(L_d) and rSFe(L_w) monotonically increased and decreased, respectively. In addition, varying L_d had a bigger impact on the fused image quality than changing L_w (see Figs. 4 and 5). Therefore, in the iterative BP-fusion algorithm, parameters L_d and L_w were varied separately, which dramatically reduced the computation loads (refer to the number of "Iterations" in Table 1) comparing with jointly varying parameters (for instance, Iterations $= 8 \times 9 = 72$ for a 512×512 image). Specifically, the procedure was as follows, given the initialization

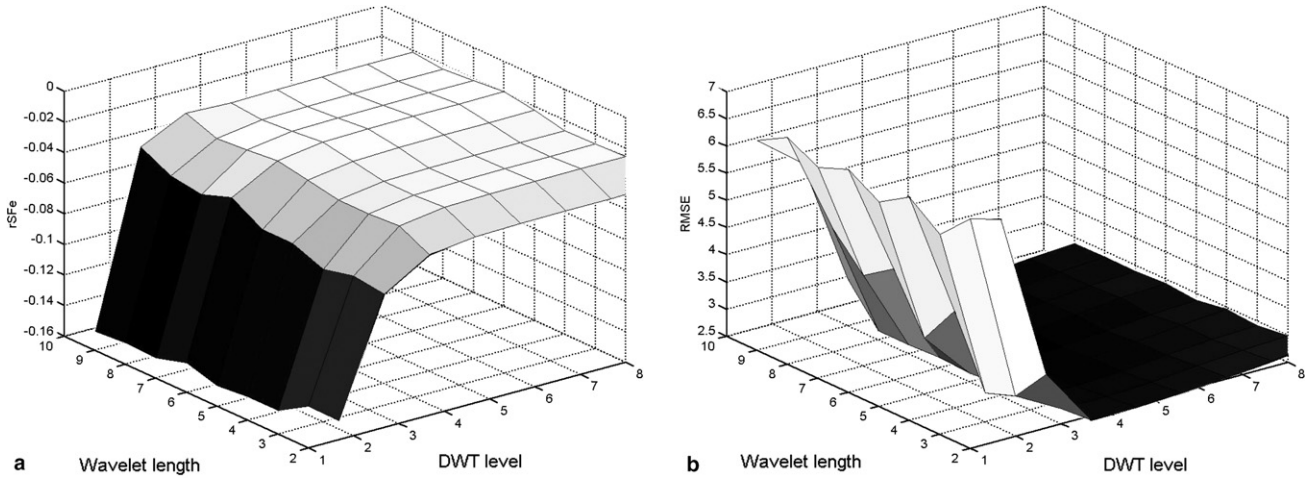

Fig. 4. Fusion metric distributions with varying parameters of DWT level (L_d) and wavelet length (L_w) while fusing the simulated pair from image Lena. (a) Distribution of rSFe(L_d, L_w). (b) Distribution of RMSE(L_d, L_w). (a) Shows a monotonically increasing function of rSFe(L_d). By comparing (a) and (b), it is seen that two metrics (rSFe and RMSE) are very consistent, which means that a small abs(rSFe) corresponds a small RMSE.

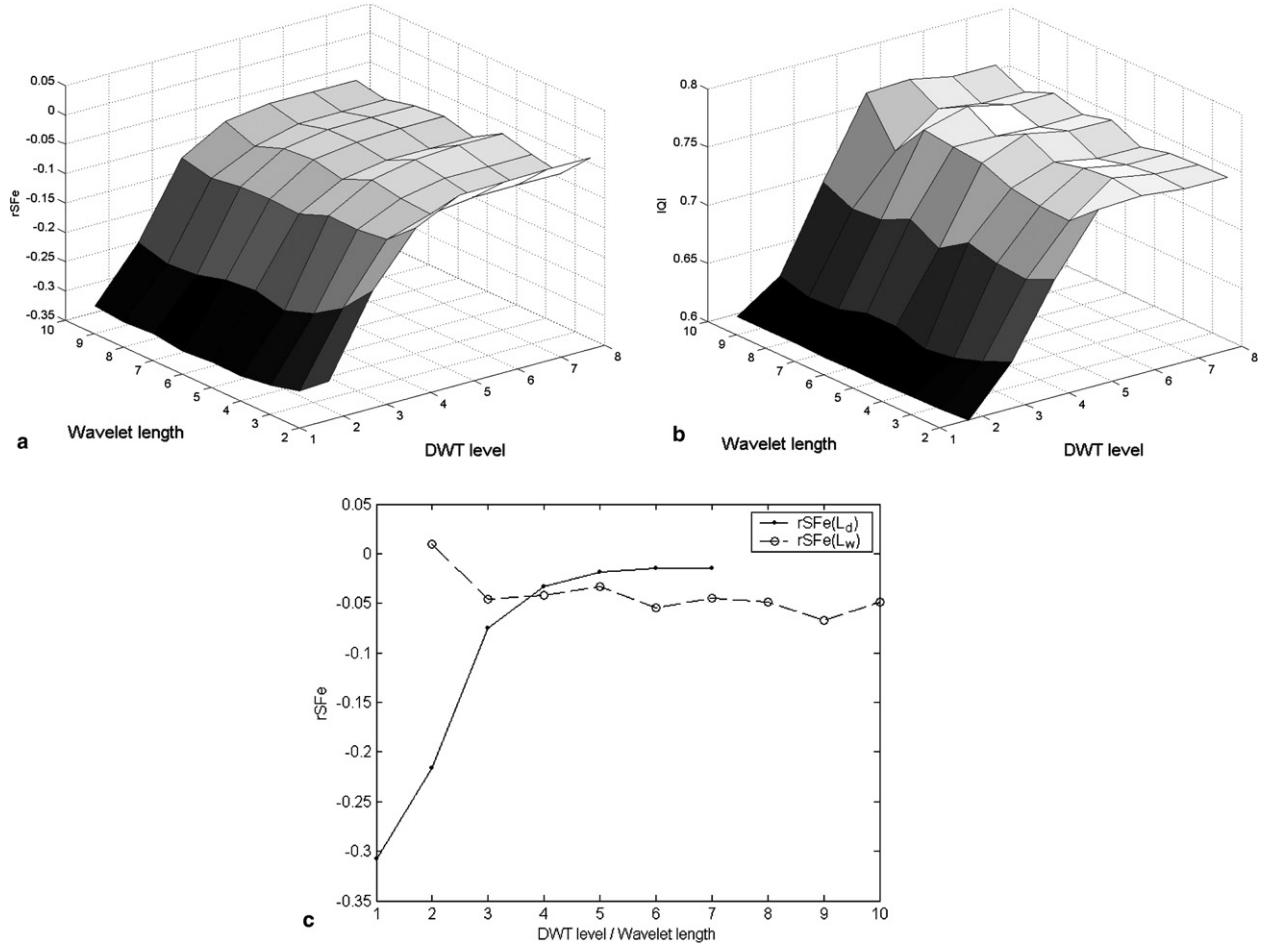

Fig. 5. Fusion metric distributions with varying parameters of DWT level (L_d) and wavelet length (L_w) while fusing the medical image pair. (a) Distribution of rSFe(L_d, L_w). (b) Distribution of IQI(L_d, L_w). (c) Distributions of rSFe(L_d) and rSFe(L_w)—two slices of (a) at $L_w = 4$ and $L_d = 4$, respectively. (c) Clearly shows a monotonically increasing function of rSFe(L_d), while rSFe(L_w) is overall a decreasing function although it is locally fluctuating. By comparing (a) and (b), it is seen that two metrics (rSFe and IQI) are very consistent, which means that a small abs(rSFe) corresponds a large IQI.

Table 1
The fused image measurements with two metrics (rSFe and IQI) over three iterative fusion methods that were directed by the back-propagated rSFe

Method	Metric	Lena	Clock	Medical	Remote	NV0140	NV1812	NV7415
*a*DWT	rSFe	−0.0256	−0.0326	0.0112	−0.0401	0.0006	0.0035	−0.0023
	IQI	0.9747	0.9222	0.7712	0.8629	0.8503	0.7082	0.7081
	(L_d, L_w)	(6,3)	(7,3)	(5,2)	(4,2)	(7,2)	(5,3)	(6,4)
	Iterations	5	6	5	4	6	4	5
Laplacian pyramid	rSFe	−0.0439	−0.2288	−0.2870	−0.3110	−0.2742	−0.1124	−0.1464
	IQI	0.9755	0.8887	0.7653	0.8526	0.8167	0.7389	0.7457
	L_d	4	4	7	8	5	4	4
	Iterations	2	2	4	5	2	2	2
Regular DWT	rSFe	−0.0522	−0.1624	−0.4272	−0.4141	−0.3317	−0.1999	−0.0289
	IQI	0.9758	0.8933	0.5625	0.7728	0.6744	0.7067	0.7219
	(L_d, L_w)	(5,4)	(4,2)	(4,2)	(4,4)	(4,3)	(6,3)	(5,4)
	Iterations	4	4	4	3	4	6	4

The converged parameters of DWT level (L_d) and wavelet length (L_w) and the number undergone iterations are also given in the table. The RMSEs of fusing the Lena pair with three methods were 2.8228, 3.0480, and 2.0041, respectively.

values of $L_d = 4$ and $L_w = 4$, begin by first varying L_d until rSFe passes the zero point, then with this fixed L_d value, change L_w until rSFe converges to the ideal value or the iteration meets one of the stop conditions. The current experiments demonstrated that this method of changing the parameters separately works very well for finding the optimal rSFe value (Table 2).

The iterative BP-*a*DWT always did better than the other two methods according to the rSFe measurement for all three types of imagery (see Table 1). So in the following analysis, all comparisons were focused on the IQI or RMSE metric. To simplify the notation, the three fusion methods were denoted by *a*DWT, Laplacian pyramid and regular DWT, and each should be interpreted with a prefix "iterative BP-" unless otherwise indicated.

4.1.1. Experimental results

4.1.1.1. One simulated image pair. Two input images were generated by filtering a given image, 'Lena' of resolution 512×512 pixels, (i.e., the ground truth as shown in Fig. 6a) with a 5×5 Gaussian window. The Lena image (served as the reference image, I_R) was divided into 16 blocks, eight of which were randomly chosen to be blurred as one of the input images, referred as "Image A" (I_A). The other eight blocks were blurred to form "Image B" (I_B). The orthogonal stripes (observed at the borders between blocks), as shown in Fig. 6b and c, were caused by sliding the window within each block. Three fused images (I_F) are illustrated in Fig. 6d and f. (Notice that through visual inspection, it is fairly difficult to discriminate between the three fused images, especially with the scaled hardcopies.) The image quality evaluation results of the fused images by three methods are given in Table 1. The RMSEs (not given in Table 1) of three fused images with *a*DWT, Laplacian pyramid and regular DWT were 2.8228, 3.0480, and 2.0041, respectively. Judged on RMSE or IQI, the best method is the regular DWT, while *a*DWT did better than Laplacian pyramid according to RMSE.

Table 2
The optimization traces of two metrics while fusing the NV7415 pair with the iterative BP-*a*DWT

(L_d, L_w)	(4,4)	(5,4)	*(6,4)*	(7,4)	(6,3)
rSFe	−0.0083	−0.0037	*−0.0023*	−0.0039	0.0105
IQI	0.6950	0.7065	*0.7081*	0.6990	0.7086

The BP-*a*DWT algorithm tried the fusion with parameters (7,4) and (6,3), but finally the best result given by (6,4) was selected (as shown in Table 1).

Fig. 4 shows two distributions of two functions—rSFe(L_d, L_w) and RMSE(L_d, L_w) while fusing this simulated pair from image Lena. Fig. 4a clearly shows a monotonically increasing function of rSFe(L_d). By comparing Fig. 4a and b, it is seen that two metrics (rSFe and RMSE) are very consistent, which means that a small abs(rSFe) corresponds a small RMSE.

4.1.1.2. Frequently used samples. Three frequently used image samples were tested: a pair of clock images obtained with two different focal planes; two different types of medical imagery, CT and MRI; and a pair of remote sensing images (infrared & low-light "visible" sensors). The quantitative assessments of fused images are listed in Table 1. The *a*DWT's performance was the best according to both metrics (rSFe and IQI). The fused images are shown in Figs. 7–9. Note that there was no post-processing imposed on these fused images.

To illustrate the relationship between rSFe and IQI, two functions, rSFe(L_d, L_w) and IQI(L_d, L_w) are provided in Fig. 5 (using the medical image pair as shown in Fig. 8). Fig. 5c shows two separate distributions of rSFe(L_d) and rSFe(L_w)—i.e., two slices of Fig. 5a at $L_w = 4$ and $L_d = 4$, respectively. Fig. 5c clearly shows a monotonically increasing function of rSFe(L_d), while

Fig. 6. Image fusion with the simulated pair from image Lena. The original image size is 512×512 pixels. All images were scaled to fit in the paper space. (a) Lena image (reference or ground truth), (b) input image A, (c) input image B, (d) fused by *a*DWT, (e) fused by Laplacian pyramid (f) fused by regular DWT.

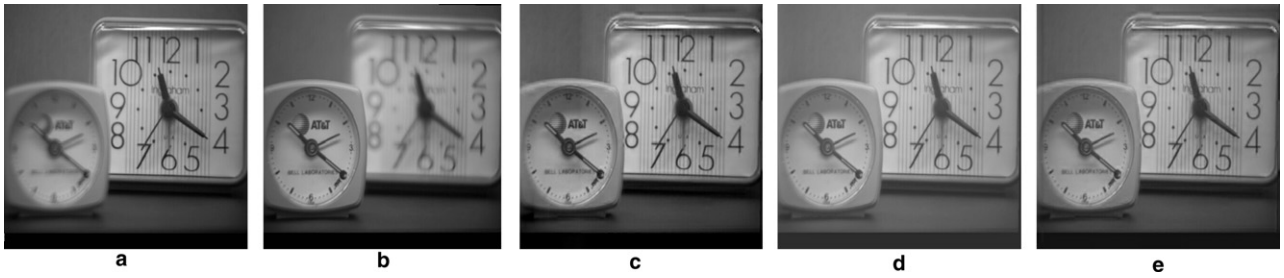

Fig. 7. Image fusion with off-focus images (512×512 pixels). (a) Input image A, (b) input image B, (c) fused by *a*DWT, (d) fused by Laplacian pyramid, and (e) fused by regular DWT.

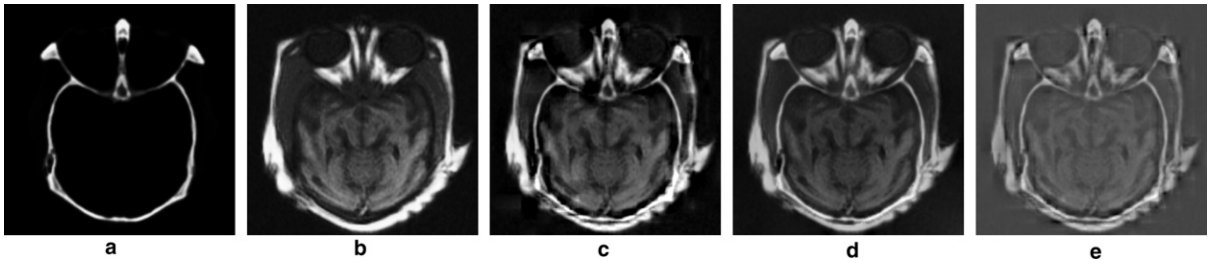

Fig. 8. Image fusion with medical images (256×256 pixels). (a) Image A (CT), (b) image B (MRI), (c) fused by *a*DWT, (d) fused by Laplacian pyramid, and (e) fused by regular DWT.

rSFe(L_w), in general, is a decreasing function exhibiting local fluctuations. By comparing Fig. 5a and b, it is observed that the two metrics (rSFe and IQI) are consistent, that is, a small abs(rSFe) corresponds to a large IQI.

4.1.1.3. Night-vision images. Three pairs of image intensified (II) and infrared (IR) night-vision images taken outdoors (obtained from TNO Human Factors) were used. The source images were moderately noisy due to collection conditions. Therefore, to suppress the noise, a 2-D median filter (with a 3×3 slide window) was first applied to all images before any fusion operation [37,38]. This is a common preprocessing practice in fusing night time imagery; moreover denoised images were

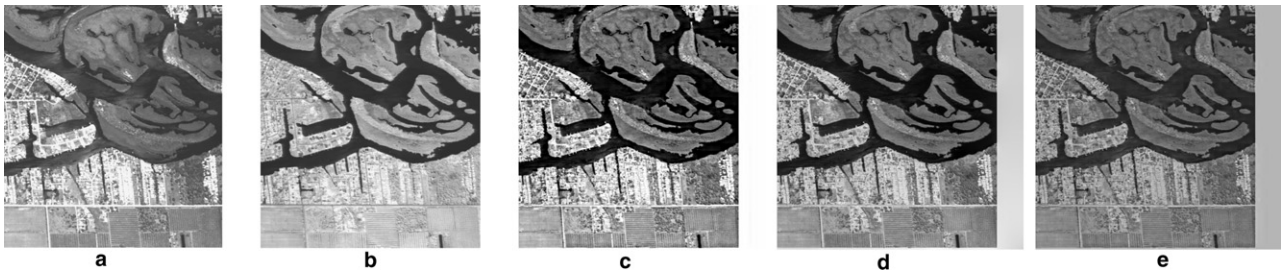

Fig. 9. Image fusion with remote sensing images (512×512 pixels). (a) Image A (infrared), (b) image B (low-light), (c) fused by *a*DWT, (d) fused by Laplacian pyramid, and (e) fused by regular DWT.

used for all three algorithms (to make a fair comparison of the fused images). The filtered images are not shown here, but this filtering operation was not detrimental to the fused image quality (e.g., see Fig. 10c–e). The complete assessments of the three algorithms along with the two metric evaluations are listed in Table 1. Judging from the IQI values, the *a*DWT was the best in fusing NV0140, but the Laplacian pyramid did the best for the other two image pairs—NV1812 and NV7415. However, some overshot was observed on the surroundings of the person in Fig. 11d (resulted by the Laplacian pyramid) although the IQI evaluation was the best. The IQI evaluations of the *a*DWT fusion for NV1812 and NV7415 are still comparable with that of regular DWT. The fused images are illustrated in Figs. 10–12.

The optimization traces of the two metrics while fusing the NV7415 pair with the iterative BP-*a*DWT algorithm were recorded in Table 2. The BP-*a*DWT algorithm tried the fusion with parameters (7, 4) and (6, 3), but finally the result (in the sense of rSFe) given by $(L_d, L_w) = (6, 4)$ was selected as the best (as shown in Table 1). To verify whether rSFe(6, 4) is the extreme value among the entire plane of (L_d, L_w), exhaustive fusion was done by using all combinations of $L_d \times L_w = 7 \times 9 = 63$. Then we found $\min\{\text{abs}[\text{rSFe}(L_d, L_w)]\} = 0.0023$ at $(L_d, L_w) = (6, 4)$ where $\text{IQI} = 0.7081$, which is the one located by separately varying L_d and L_w with only five iterations. On the other hand, we have $\max[\text{IQI}(L_d, L_w)) = 0.7095$ at $(L_d, L_w) = (6, 6)$, where $\text{rSFe} = -0.0087$. According to the IQI metric, $(L_d, L_w) = (6, 4)$ are not the best parameters for this particular fusion, but the difference of two IQIs is very small with only $(0.7095–0.7081)/1 = 0.0014 = 0.14\%$.

4.1.2. Further discussion with the quantitative analyses

The metric rSFe is defined by a spatial frequency (SF) comparison with a reference SF (SF_R) as shown in Eq. (12). To provide an intuitive impression of this important calculation, spatial frequency values of all images were given in Table 3. In each column of the table (each type of imagery), the SF of input images A and B (using denoised images for night-vision imagery), the fused image F (by the BP-*a*DWT algorithm with shown parameters in Table 1), computed reference (not the ground truth) SF, and the ratio of SF error (rSFe) were listed.

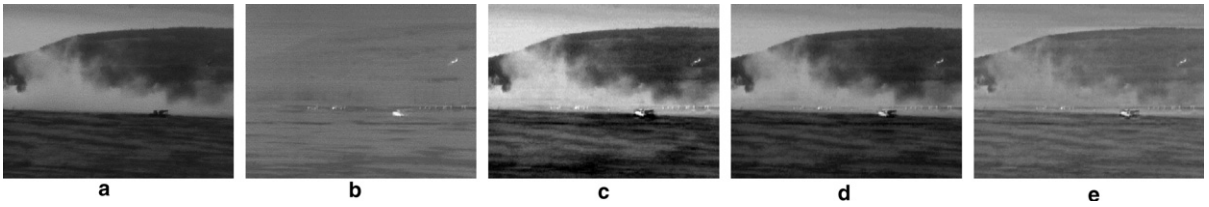

Fig. 10. Image fusion with night-vision image pair NV0140 (360×270 pixels). (a) Image A (II), (b) image B (IR), (c) fused by *a*DWT, (d) fused by Laplacian pyramid, and (e) fused by regular DWT.

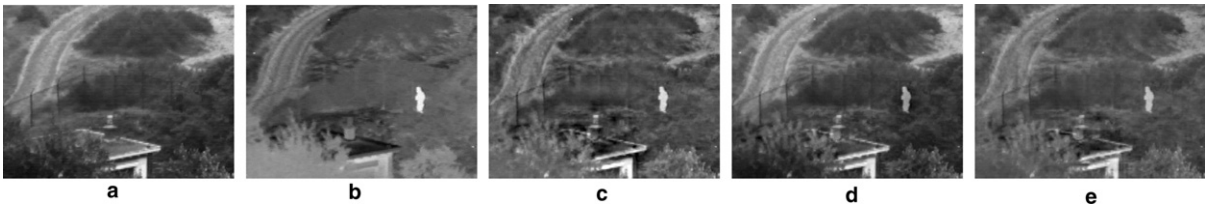

Fig. 11. Image fusion with night-vision image pair NV1812 (360×270 pixels). (a) Image A (II), (b) image B (IR), (c) fused by *a*DWT, (d) fused by Laplacian pyramid, and (e) fused by regular DWT.

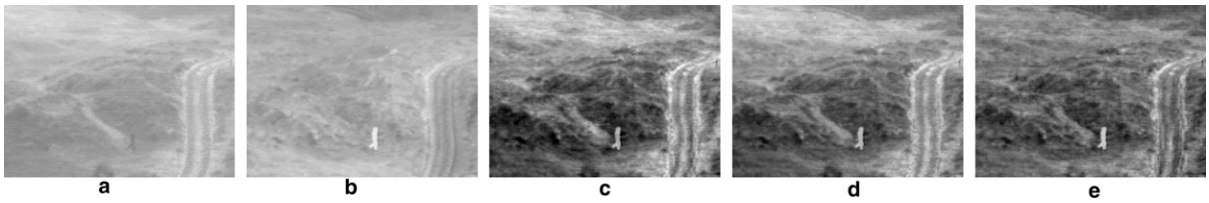

Fig. 12. Image fusion with night-vision image pair NV7415 (360 × 270 pixels). (a) Image A (II), (b) image B (IR), (c) fused by *a*DWT, (d) fused by Laplacian pyramid, and (e) fused by regular DWT.

Table 3
Spatial frequency (SF) values of all images

Image	Lena	Clock	Medical	Remote	NV0140	NV1812	NV7415
SF_A	16.3754	14.071	21.3343	53.7687	16.4374	15.0984	15.3057
SF_B	18.6206	11.1707	20.8579	46.8569	14.5631	11.8662	15.5508
SF_F	21.3353	16.5837	29.4372	61.7924	20.7935	18.3936	20.3924
SF_R	21.7955	17.1627	29.0218	64.1196	20.6961	18.2407	20.3503
rSFe	−0.0256	−0.0373	0.0112	−0.0401	0.0006	0.0035	−0.0023

In each column of the table, the SF of input image A and B (using denoised images for night-vision imagery), the fused image F (by the BP-*a*DWT algorithm with parameters indicated in Table 1), computed reference (not the ground truth) SF and the ratio of SF error were listed. For image Lena, the SF of ground truth image is $SF_{GT} = 20.9790$. The relative error is $(SF_R - SF_{GT})/SF_{GT} = (21.7955–20.9790)/20.9790 = 0.0389 \approx 4\%$.

For example, for image Lena, the SF of the ground truth image is $SF_{GT} = 20.9790$. The relative error between the computed reference SF and the ground truth SF is: $(SF_R - SF_{GT})/SF_{GT} = (21.7955 - 20.9790)/20.9790 = 0.0389 \approx 4\%$.

The definition of metric rSFe is self-integrated. Given two identical images, $I_A = I_B$, we have $Grad^D(I_R) = Grad^D(I_A)$ according to Eq. (10), thereby from Eqs. (7)–(9) $SF_R = SF_A$. The fused image (I_F) by *a*DWT is certainly identical to I_A, i.e., $SF_F = SF_A$. Clearly, in this case rSFe = 0 according to Eq. (12). In addition, IQI = 1 can be proved for this special fusion ($I_A = I_B$) with the definition of IQI in Eqs. (2)–(5).

The iterative BP-fusion algorithms (especially for *a*DWT) directed by the metric rSFe are practical for typical applications. Experiments (refer to Table 2) already showed that an optimal rSFe existed that can be located by separately varying the two parameters, the DWT decomposition level (L_d) and the wavelet length (L_w) with only several iterations. And the number of iterations may be further reduced by properly initializing parameters or optimizing the termination conditions. Values of the new metric rSFe are quite consistent with currently used metrics such as RMSE or IQI.

Overall, the iterative BP-*a*DWT was best at fusing inhomogeneous imagery. For night-vision imagery where brightness and contrast are very different between the two input images, the Laplacian pyramid also produced better-fused images judged on the IQI metric, however it caused an overshoot result at high contrast regions.

To obtain better-fused image quality, extra computations were required with iterative fusion. Keep in mind that the optimal fusion (described by the measurements, parameters, iterations, etc.) with the presented BP-fusion procedure is adaptive with respect to the input images. To illustrate the superiority of the iterative BP-fusion algorithms, the measurements of all fused images through three non-iterative methods by using a pair of default parameters ($L_d = 4$ and $L_w = 4$, as typically used in the literature) are shown in Table 4. By comparing Tables 4 and 1, the following points are

Table 4
The fused image measurements with two metrics (rSFe and IQI) over three non-iterative fusion methods by using default parameters of DWT level ($L_d = 4$) and wavelet length ($L_w = 4$)

Method	Metric	Lena	Clock	Medical	Remote	NV0140	NV1812	NV7415
*a*DWT	rSFe	−0.0276	−0.0480	−0.0417	−0.0512	−0.0145	−0.0239	−0.0083
	IQI	0.9747	0.9236	0.7495	0.8644	0.8575	0.6855	0.6950
Laplacian pyramid	rSFe	−0.0439	−0.2288	−0.3297	−0.3377	−0.2744	−0.1124	−0.1464
	IQI	0.9755	0.8887	0.6975	0.8308	0.8160	0.7389	0.7457
Regular DWT	rSFe	−0.0539	−0.2018	−0.5146	−0.4141	−0.3488	−0.2764	−0.0398
	IQI	0.9757	0.8978	0.5107	0.7728	0.6660	0.6599	0.7167

This table can be compared with Table 1 to have the impression that how good the iterative fusion algorithms are.

evident: (a) the optimized or converged results by iterative fusion algorithms are significantly better than those produced by non-iterative algorithms; (b) the location of the abs(rSFe) minimum is quite consistent with that of the IQI maximum; and (c) the change of function IQI with parameters (L_d, L_w) is less sensitive in fusing off-focus images (such as images Lena and Clock).

One may consider using the metric IQI as the feedback of an iterative BP-fusion algorithm. But the obvious drawback of using IQI is that the error IQI ($= \mathrm{IQI_F} - 1$) cannot provide the error direction (requirement (c) in Section 3.3) because the error IQI is always negative. We will report the further investigation of this aspect of IQI in a subsequent paper.

4.2. Qualitative analyses and discussion

In Section 4.1, a series of quantitative analyses were carried out in order to determine which of three image fusion methods (i.e., *a*DWT, Laplacian pyramid, and Regular DWT) produced the 'best' fused images. However, the evaluation results with quantitative metrics do not provide any information about the perceived visual quality of the fused imagery. Accordingly, a psychophysical experiment was designed and carried out below. The aim of this qualitative analysis was to determine which of the three fusion methods produce imagery with the highest perceived image quality by human observers viewing two different types of monitors.

4.2.1. Psychophysical experiment design

4.2.1.1. Participants. For the two monitor conditions, two groups of 15 observers participated (15 participants per monitor condition; 30 participants total). All observers had normal (20/20) or corrected to normal acuity, normal color vision, and no history of ocular pathologies. IRB-approved informed consent was obtained.

4.2.1.2. Apparatus. Stimuli were presented by two monitor conditions, which were by a Dell™ Dimension 8250 Desktop Computer on a Liquid Crystal Display (LCD) monitor, or by an eMachines™ T2040 Desktop Computer on a Cathode Ray Tube (CRT) monitor. The LCD was a Dell UltraSharp Monitor that had a resolution of 1280×1024 pixels, a frame-rate of 75 Hz, and a maximum luminance of 96 cd/m^2 with a gamma of 2.2. The CRT was a Proview PS709 Monitor that had a resolution of 1280×1024 pixels, a frame-rate of 60 Hz, and a maximum luminance of 62 cd/m^2 with a gamma of 2.2.

4.2.1.3. Stimuli. The stimuli used in both the LCD and CRT monitor conditions consisted of six groups of images (5 images in each group, thus a total of 30 images as shown in Figs. 7–12). Note that no post-processing was imposed on the fused images. The simulated imagery from Lena was not used here because the three fused images (as shown in Fig. 6d–f) were visually identical (also refer to their quantitative evaluations in Table 1). The stimulus imagery consisted of two input images and the three fused images produced by one of the three fusion methods (*a*DWT, Laplacian pyramid, and regular DWT) for each pair of the input images.

4.2.1.4. Qualitative assessment methods. The methodology for the two monitor conditions was identical. In order to obtain an objective qualitative assessment of the three image fusion methods, a standard psychophysical *rank-ordering* paradigm [35,36] was employed. Essentially, this paradigm consisted of presenting the participants with the three fused images and asking each participant to rank order each of those three images based on their "image quality" by assigning each of the fused images a numerical value. Specifically, they were asked to assign the fused image perceived to possess 'high image quality' with a '1', the fused image perceived to possess 'medium image quality' with a '2', and the fused image perceived to possess 'low image quality' with a '3'. For each participant, the fused images to be ranked in each image set were arranged in a random order. Before beginning the experiment, each participant was asked to read standardized instructions which explained the task (refer to "Appendix A" for the set of instructions provided to each participant), and all participants were allowed ask questions regarding the task before beginning the experiment.

At the beginning of the experiment, participants were seated comfortably in a chair facing the monitor at a distance of 53 cm from the monitor and an unrestrained head position. For each trial, five images in each of six groups were simultaneously displayed, at their actual resolution except for the 'Clock' (Fig. 7) and 'Remote' (Fig. 9) imagery which were displayed at their half resolution (256×256, down sampled with the bicubic interpolation), with a MATLAB (Version 6.5) rendered Graphic User Interface (GUI). The two input images were presented above the three fused images (refer to Fig. 13 for an example of the spatial layout of the imagery for a given trial). Below each fused image was an input box where participants entered one of the three rank order values.

4.2.2. Experimental results and discussion

In the current experiment, each of the three fused images from the six groups of imagery could have been assigned a numerical value of 1, 2, or 3. Thus, within each condition, each of the three fused images was assigned 15 values (one value per participant). In order to carry out a meaningful quantitative analysis, the *rank frequency*, that is the number of times a fused image was assigned a rank value (i.e., the number of all the ones, twos, and threes for a given image), was taken as the operational definition of perceived image quality. For

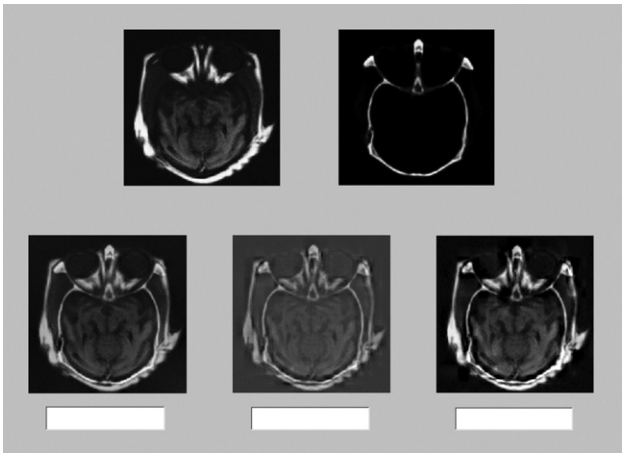

Fig. 13. The GUI for the qualitative analyses (with rank order)—the spatial layout of the imagery. For each participant, the three fused images (on the low row) in each set were displayed with a random order.

example, if a given image was assigned a high number of ones, relative to the other two values, this would indicate that most of the participants perceived that image as possessing the highest image quality relative to the other two fused images and vice versa if a given image was assigned a high number of threes.

For each condition the rank frequency were summed across the six image groups, which resulted in the *summed rank frequency* (refer to Column 'Sum' in Tables 5–7). The data obtained from both monitor conditions are shown in Figs. 14 and 15. From both figures, it is clear that the *a*DWT image fusion method was assigned the highest number of ones (i.e., this fusion method produced imagery that was perceived to have the highest image quality compared to the imagery produced by the other two fusion methods); the Laplacian pyramid fusion method produced imagery that was assigned the highest number of twos (i.e., essentially perceived as "second best"); and the DWT fusion method produced imagery that was assigned the highest number of threes (i.e., essentially perceived as "third best"). Thus it appears that, regardless the type of stimulus display (LCD or CRT), the participants in both conditions perceived the *a*DWT fused imagery as having the highest image quality relative to the imagery produced by of the other two image fusion methods. Rank order frequencies for each image type are shown in Tables 5 and 6.

In order to determine if the results from the two monitor conditions were significantly different, a $2 \times 3 \times 3$ (i.e., *group* by *fusion-method* by *rank-frequency*) three-way repeated measures Analysis of Variance (ANOVA) was carried out on the "summed rank frequency" for each of the three different fusion methods using SPSS (Version 12). The result of this analysis was a non-significant interaction ($F_{(4,40)} = 0.721$, $p = 0.583$), indicating

Table 5
The number of rank orders in the LCD condition with 15 participants

Method	Rank	Clock	Medical	Remote	NV0140	NV1812	NV7415	Sum
*a*DWT	1	**10**	**9**	**14**	**10**	6	5	**54**
	2	2	6	1	4	6	5	24
	3	3	0	0	1	3	5	12
Laplacian pyramid	1	**7**	5	1	3	**9**	3	28
	2	4	**9**	**13**	**9**	5	5	**45**
	3	4	1	1	3	1	**7**	17
Regular DWT	1	1	1	0	2	0	**7**	11
	2	**9**	0	1	2	4	5	21
	3	5	**14**	**14**	**11**	**11**	3	**58**

The value in column 'Sum' means the sum of ranks along that row (i.e., across six types of images).

Table 6
The number of rank orders in the CRT condition with 15 participants

Method	Rank	Clock	Medical	Remote	NV0140	NV1812	NV7415	Sum
*a*DWT	1	**14**	**12**	**11**	**9**	**8**	**9**	**63**
	2	0	2	1	5	4	3	15
	3	1	1	3	1	3	3	12
Laplacian pyramid	1	1	2	4	5	**7**	2	21
	2	2	**11**	**10**	**7**	6	3	**39**
	3	**12**	2	1	3	2	**10**	30
Regular DWT	1	0	1	0	1	0	4	6
	2	**13**	2	4	3	5	**9**	36
	3	2	**12**	**11**	**11**	**10**	2	**48**

Table 7
The number of rank orders in the LCD and CRT combined condition with a total of 30 participants

Method	Rank	Clock	Medical	Remote	NV0140	NV1812	NV7415	Sum
*a*DWT	1	**24**	**21**	**25**	**19**	**14**	**14**	**117**
	2	2	8	2	9	10	8	39
	3	4	1	3	2	6	8	24
Laplacian pyramid	1	8	7	5	8	**16**	5	49
	2	6	**20**	**23**	**16**	11	8	**84**
	3	**16**	3	2	6	3	**17**	47
Regular DWT	1	1	2	0	3	0	11	17
	2	**22**	2	5	5	9	**14**	57
	3	7	**26**	**25**	**22**	**21**	5	**106**

that the results from the two monitor conditions were not significantly different. Thus, regardless the monitor type (LCD or CRT), participants ranked imagery produced by the *a*DWT method highest relative to the other two fusion methods. Since the interaction was not significant, the data from the two monitor conditions was combined (i.e., summed across groups). In order to determine if the differences in fused imagery preference (i.e., aDWT being best; Laplacian being second best; and Regular DWT being third best) were significant, a 3×3 (i.e., *fusion-method* by *rank-frequency*) two-way repeated measures ANOVA was carried out. The interaction between the fusion method and the rank frequency was found to be significant ($F_{(4,20)} = 5.309$, $p = 0.004$), indicating a significant difference for the rank frequency. The combined data (i.e., $n = 30$) from the two monitor conditions, where the frequencies of rank order values were summed across the six image groups (e.g., the "summed rank frequency") are depicted in Fig. 16, the combined data for each image group is shown in Table 7.

Thus, the results from the current qualitative assessment of fused image quality are in agreement with the quantitative analyses (i.e., the rSFe and IQI metrics).

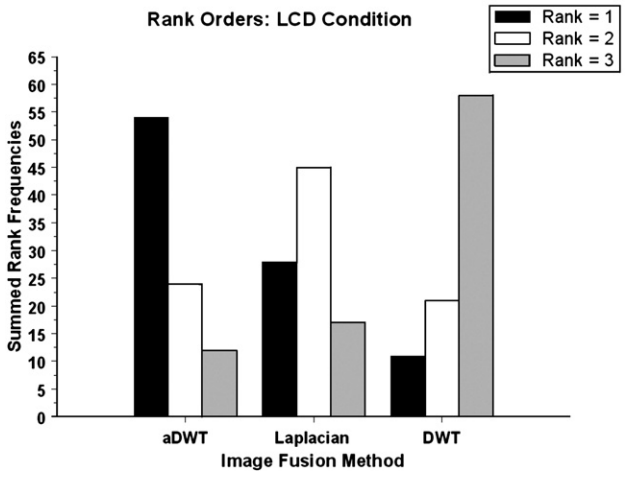

Fig. 14. The rank frequencies in the LCD condition (with 15 participants)—the rank frequencies summed across six types of images over three fusion methods.

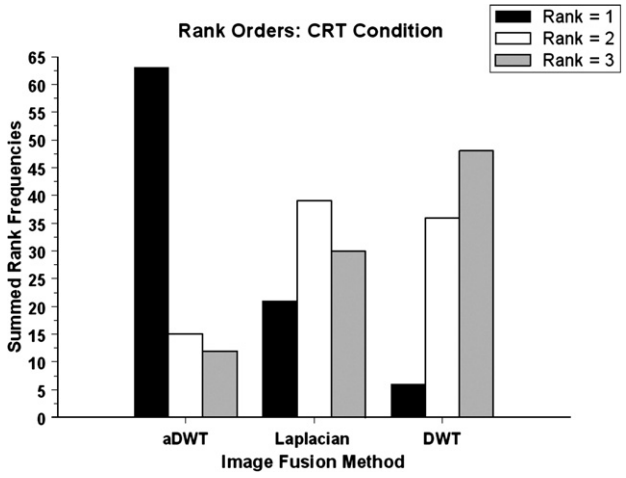

Fig. 15. The rank frequencies in the CRT condition (with 15 participants)—the rank frequencies summed across six types of images over three fusion methods.

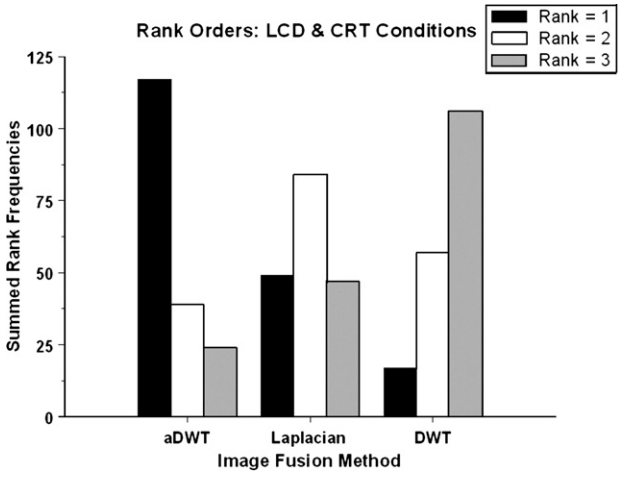

Fig. 16. The rank frequencies in the combined condition (LCD plus CRT, with 30 participants)—the rank frequencies summed across six types of images in both monitor conditions over three fusion methods.

Specifically, the fused imagery produced by the *a*DWT method was most frequently perceived as having the highest image quality relative to the other two methods (Laplacian pyramid and the regular DWT).

4.3. Future work

We plan to extend the definition of metric rSFe to multi-band (more than two) sensor fusion. We also plan to further investigate the effects of noise on rSFe.

5. Conclusions

We presented a new metric for image fusion based on an extended definition of spatial frequency, which is the ratio of spatial frequency error (rSFe). Experiments showed that evaluation with rSFe is very consistent with the RMSE (root mean square error) and IQI (image quality index) metrics, but the rSFe metric is sensitive to small changes in image quality and can also provide more information of the fusion process—under-fused ($rSFe < 0$) or over-fused ($rSFe > 0$)—that makes it useful for iterative fusion processes such as the one proposed here. With the absolute value and the sign of rSFe back propagated (BP) to the fusion algorithm, an iterative BP-fusion procedure can be directed, thus an optimized fused image can be provided. The iterative BP-*a*DWT (advanced DWT), which incorporated PCA (principal component analysis) and morphological processing into a regular DWT fusion procedure, has been shown to significantly improve upon the regular DWT or Laplacian pyramid on the basis of the quantitative IQI metric and the qualitative perceptual evaluation with rank ordering on the fused imagery.

Acknowledgements

This work was supported by grant #N00014-03-1-0224 from the Office of Naval Research. We wish to thank Lex Toet and the TNO Human Factors Research Institute who provided the night-vision imagery. Finally, thanks are given to the anonymous reviewers for many helpful and constructive suggestions.

Appendix A. Standard set of instructions utilized in the rank ordering experiment

"Please rank order the three fused images on the lower row with 1 (best), 2 (medium), or 3 (worst) based on their *image quality* by considering the two input images (e.g., from different sensors such as visible band and infrared) in the upper row. The best image quality (with rank order = 1) means that it contains the most useful information (combined from the two input images) but the least artifacts or distortion."

References

[1] L.D. Keys, N.J. Schmidt, B.E. Phillips, A prototype example of sensor fusion used for a siting analysis, in: Technical Papers 1990, ACSM-ASPRS Annual Conf. Image Processing and Remote Sensing, 1990, vol. 4, pp. 238–249.
[2] R.H. Rogers, L. Wood, The history and status of merging multiple sensor data: an overview, in: Technical Papers 1990, ACSMASPRS Annual Conf. Image Processing and Remote Sensing, 1990, vol. 4, pp. 352–360.
[3] E.A. Essock, M.J. Sinai, J.S. McCarley, W.K. Krebs, J.K. DeFord, Perceptual ability with real-world nighttime scenes: imageintensified, infrared, and fused-color imagery, Hum. Factors 41 (3) (1999) 438–452.
[4] E.A. Essock, J.S. McCarley, M.J. Sinai, J.K. DeFord, Human perception of sensor-fused imagery, in: R.R. Hoffman, A.B. Markman (Eds.), Interpreting Remote Sensing Imagery: Human Factors, Lewis Publishers, Boca Raton, Florida, 2001.
[5] C. Pohl, J.L. Van Genderen, Review article: multisensor image fusion in remote sensing: concepts, methods and applications, Int. J. Remote Sens. 19 (5) (1998) 823–854.
[6] O. Rockinger, Pixel level fusion of image sequences using wavelet frames, in: Proc. 16th Leeds Applied Shape Research Workshop, Leeds University Press, 1996, pp. 149–154.
[7] H. Li, B.S. Manjunath, S.K. Mitra, Multisensor image fusion using the wavelet transform, Graph. Models Image Process. 57 (3) (1995) 235–245.
[8] T. Pu, G. Ni, Contrast-based image fusion using the discrete wavelet transform, Opt. Eng. 39 (8) (2000) 2075–2082.
[9] D.A. Yocky, Image merging and data fusion by means of the discrete two-dimensional wavelet transform, J. Opt. Soc. Am. A 12 (9) (1995) 1834–1841.
[10] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, R. Arbiol, Image fusion with additive multiresolution wavelet decomposition: applications to spot1 landsat images, J. Opt. Soc. Am. A 16 (1999) 467–474.
[11] F. Jahard, D.A. Fish, A.A. Rio, C.P. Thompson, Far/near infrared adapted pyramid-based fusion for automotive night vision, in: IEEE Proc. 6th Int. Conf. on Image Processing and its Applications (IPA97), 1997, pp. 886–890.
[12] B. Ajazzi, L. Alparone, S. Baronti, R. Carla, Assessment of pyramid-based multisensor image data fusion, Proc. SPIE 3500 (1998) 237–248.
[13] M. Beauchemin, K.B. Fung, X. Geng, A method based on local variance for quality assessment of multiresolution image fusion, in: Proc. IAPRS, 2002, vol. XXXIV, p. B-32 ff.
[14] L.W. Leung, B. King, V. Vohora, Comparison of image data fusion techniques using entropy and INI, in: Proc. ACRS, vol. 1, 2001, pp. 152–157.
[15] G. Piella, H. Heijmans, A new quality metric for image fusion, in: Proc. 2003 Int. Conf. on Image Processing, Barcelona, Spain, 2003.
[16] Y. Zheng, E.A. Essock, B.C. Hansen, An advanced image fusion algorithm based on wavelet transform—incorporation with PCA and morphological processing, Proc. SPIE 5298 (2004) 177–187.
[17] S. Li, J.T. Kwok, Y. Wang, Combination of images with diverse focuses using the spatial frequency, Infusionstherapie 2 (3) (2001) 169–176.
[18] P.J. Burt, E.H. Adelson, Merging images through pattern decomposition, Proc. SPIE 575 (1985) 173–182.
[19] Z. Wang, A.C. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9 (3) (2002) 81–84.

[20] A.M. Eskicioglu, P.S. Fisher, Image quality measure and their performance, IEEE Trans. Commun. 43 (12) (1995) 2959–2965.
[21] P.J. Burt, The pyramid as structure for efficient computation, in: A. Rosenfeld (Ed.), Multiresolution Image Processing and Analysis, Springer-Verlag, New York/Berlin, 1984, pp. 6–35.
[22] P.J. Burt, E. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. Com-31 (4) (1983) 532–540.
[23] A. Toet, Image fusion by a ratio of low pass pyramid, Pattern Recogn. Lett. 9 (4) (1989) 245–253.
[24] A. Toet, L.J. Van Ruyven, J.M. Valeton, Merging thermal and visual images by contrast pyramid, Opt. Eng. 28 (7) (1989) 789–792.
[25] A. Toet, Hierarchical image fusion, Mach. Vision Appl. 3 (1) (1990) 1–11.
[26] A. Toet, Multiscale contrast enhancement with application to image fusion, Opt. Eng. 31 (5) (1992) 1026–1031.
[27] P.J. Burt, A gradient pyramid basis for pattern-selective image fusion, SID Int. Symp. Digest Tech. Papers 16 (1985) 467–470.
[28] P.J. Burt, R.J. Lolczynski, Enhanced image capture through fusion, in: IEEE Proc. 4th Int. Conf. Computer Vision, 1993, pp. 173–182.
[29] S. Richard, F. Sims, M.A. Phillips, Target signature consistency of image data fusion alternatives, Opt. Eng. 36 (3) (1997) 743–754.
[30] A. Toet, A morphological pyramid image decomposition, Pattern Recogn. Lett. 9 (4) (1989) 255–261.
[31] G.K. Matsopoulos, S. Marshall, J. Brunt, Multiresolution morphological fusion of MR and CT images of the human brain, IEE Proc. Vision Image Signal Process. 141 (3) (1994) 137–142.
[32] G.K. Matsopoulos, S. Marshall, Application of morphological pyramids: fusion of MR and CT phantoms, J. Visual Commun. Image Represent. 6 (2) (1995) 196–207.
[33] R.C. Gonzalez, R.E. Woods, Digital Image Processing, second ed., Prentice Hall, Upper Saddle River, NJ, 2002.
[34] D.W. Ruck, S.K. Rogers, M. Kabrisky, P.S. Maybeck, M.E. Oxley, Comparative analysis of backpropagation and the extended Kalman filter for training multilayer perceptrons, IEEE Trans. Pattern Anal. Mach. Intell. 14 (6) (1992) 686–691.
[35] T. Engen, Psychophysics: Scaling Methods, in: J.W. Kling, L.A. Riggs (Eds.), Experimental Psychology, Sensation and Perception, vol. 1, Holt, Rinehart and Winston Inc., New York, 1972, pp. 47–86.
[36] J.C. Falmagne, Psychophysical measurement and theory, in: K.R. Boff, L. Kaufman, J.P. Thomas (Eds.), Handbook of Perception and Human Performance, Sensory Processes and Perception, vol. 1, John Wiley & Sons, New York, 1986, pp. 1-1–1-64.
[37] Z. Zhang, S. Sun, F. Zheng, Image fusion based on median filters and SOFM neural networks: a three-step scheme, Signal Process. 81 (6) (2001) 1325–1330.
[38] V.S. Petrovic, C.S. Xydeas, Sensor noise effects on signal-level image fusion performance, Inform. Fusion 4 (3) (2003) 167–183.