Infrared and visible image fusion using fuzzy logic and population-based optimization

Jamal Saeedi, Karim Faez *

Electrical Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), 424, Hafez Ave., Tehran, Iran

ABSTRACT

This paper presents a new wavelet-based algorithm for the fusion of spatially registered infrared and visible images. Wavelet-based image fusion is the most common fusion method, which fuses the information from the source images in the wavelet transform domain according to some fusion rules. We specifically propose new fusion rules for fusion of low and high frequency wavelet coefficients of the source images in the second step of the wavelet-based image fusion algorithm. First, the source images are decomposed using dual-tree discrete wavelet transform (DT-DWT). Then, a fuzzy-based approach is used to fuse high frequency wavelet coefficients of the IR and visible images. Particularly, fuzzy logic is used to integrate the outputs of three different fusion rules (weighted averaging, selection using pixel-based decision map (PDM), and selection using region-based decision map (RDM)), based on a dissimilarity measure of the source images. The objective is to utilize the advantages of previous pixel- and region-based methods in a single scheme. The PDM is obtained based on local activity measurement in the DT-DWT domain of the source images. A new segmentation-based algorithm is also proposed to generate the RDM using the PDM. In addition, a new optimization-based approach using population-based optimization is proposed for the low frequency fusion rule instead of simple averaging. After fusing low and high frequency wavelet coefficients of the source images, the final fused image is obtained using the inverse DT-DWT. This new method provides improved subjective and objectives results as compared to previous image fusion methods.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The research presented in this paper is concerned with the problem of multi-sensor pixel-level image fusion. The objective is to produce reliable methods, which represent the visual information, obtained from a number of different imaging sensors, in a single fused image without causing distortion or loss of information. Multi-sensor image fusion, while decreases the amount of data, creates new images, which are more suitable for the purposes of human observation or for image processing tasks such as object detection, segmentation, or target recognition. Image fusion has many important applications such as medical imaging [1], microscopic imaging, remote sensing [2], computer vision, and robotics [3].

In this paper, we have concentrated on the fusion of visible and infrared images. The most important motivation for integrating visible and infrared (IR) images to produce a fused image, which is constructed by a combination of features, and allows improved detection and directly recognizable localization of a target in the IR image with respect to its background in the visible image [4]. Therefore, a human operator using a suitably fused representation of visible and IR images may be able to construct a more complete and accurate mental representation of the perceived scene, and results in a larger degree of situation awareness [5]. Fig. 1 illustrates how information from visible and IR images can improve situational awareness in a military campus.

Image fusion represents a specific case of multi-sensor information fusion in which all the information sources used represent imaging sensors. Analogous to other forms of information fusion, image fusion is usually performed at one of the three different processing levels: signal, feature, and decision [6]. Signal level image fusion, which is also known as pixel-level image fusion, represents fusion at the lowest level, where a number of raw input images are combined to produce a single fused image. Pixel-level image fusion methods consisting the simple pixel averaging to more complicated methods such as principal component analysis [7], and multi-resolution (MR) fusion methods [8].

Multi-scale decomposition-based (MSD) image fusion is a biologically motivated method, which fuses multiple images at

* Corresponding author. Tel.: +98 935 3592459; fax: +98 216 6406469.
E-mail addresses: jamal.saeedi@yahoo.com (J. Saeedi), kfaez@aut.ac.ir (K. Faez).

Fig. 1. Fusion of multi-sensor images, from left to right: infrared and visible "UN Camp" images, and fused image.

different spatial resolutions. MSD-based image fusion includes three main steps [9]:

1. The input images are decomposed into a resolution pyramid of numerous levels. Each level contains one or more bands representing low and high frequency information.
2. Following this decomposition, the coefficients or samples in each band of the source images are fused based on some fusion rules.
3. The fused pyramid is then reconstructed to generate the final fused image.

Up to now, several algorithms based on multiscale transforms have been proposed, which use multi-resolution transformations consisting of the Laplacian pyramid [10], gradient pyramid [11], morphological pyramid [12], and wavelet transform. Research results revealed that the wavelet transform schemes have more advantages over pyramid schemes such as increased directional information, no blocking artifacts, better signal-to-noise ratio, improved perception, and so forth [13].

The key step in the MSD-based image fusion is the coefficients combining step or fusion rule. The fusion rule is the process of merging the wavelet coefficients in an appropriate method in order to achieve the best quality in the final fused image. Some general alternatives to construct a fusion rule are illustrated in Fig. 2. As it can be seen in Fig. 2, the activity level measurement is used to determine the quality of each source image in the transform domain. Grouping and combining methods are also used to achieve the composite multi-scale representation of the fused image. The objective of the consistency verification process is generation of a composite MSD coefficient in a completely similar way from all its neighbors [14]. There are several alternatives in the processes distinguished by the dashed boxes in Fig. 2, and different combinations of these alternatives produce different fusion schemes.

Several pixel-based fusion methods exist in the literature [15–17], wherein each individual coefficient of the MR decomposition (or may be the coefficients in a small fixed neighborhood) is treated more or less independently. It is followed by many region-based fusion methods to combine objects instead of pixels [18–21]. For the fusion of IR and visible images, it could be more significant to combine objects rather than pixels, because there are some regions or objects in the IR image that are not visible in the visual image, and vice versa. Additional advantage of the region-based approaches is that the fusion process becomes more robust and this may help avoid some of the well-known disadvantages of pixel-based methods, such as blurring effects, high sensitivity to noise, and image misregistration [19].

Commonly, there are two ways for the region-based fusion methods: maximum selection (MS) and weighted averaging (WA). In the MS scheme, a region from one of the source images is selected based on the activity level measurement, and the other corresponding regions in the remaining images are neglected (e.g. in Ref. [18], normalized Shannon entropy is used for activity level measurement). In the WA scheme, first, the activity measurements of similar regions in different source images are obtained, and then the resultant coefficients for reconstruction are calculated from these measures via a weighted average of the coefficients in the similar regions of different sources (e.g. in Ref. [21], the weights are obtained based on Alpha-Stable modeling of the wavelet coefficients).

These fusion rules ignore some useful information. Selective operation makes the fused coefficients completely dependent on the coefficients with larger activity in the image regions and ignores other corresponding coefficients from other source images. Besides, in the weighted averaging scheme, choosing one weight as the activity level for all of the coefficients in a region cannot describe the uncertainty of each source image contributions, and results in a low-contrast fused image.
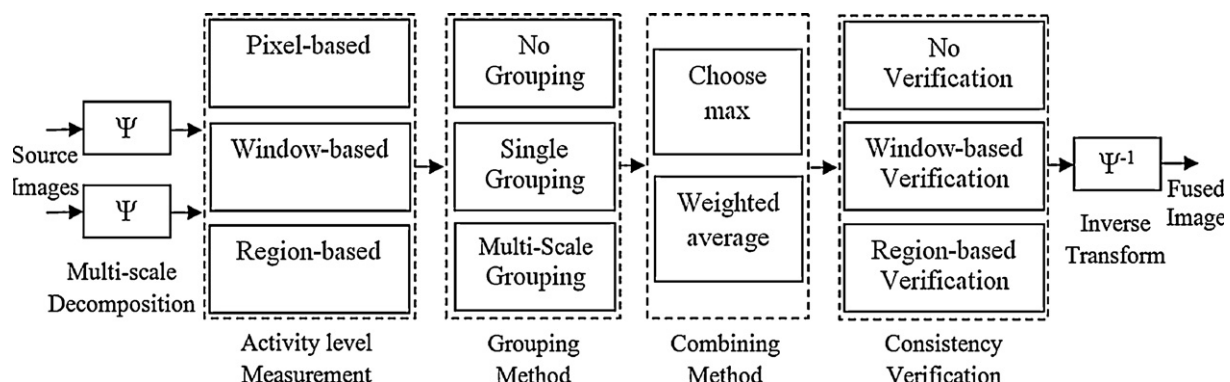

Fig. 2. A general framework for multi-resolution-based image fusion [14].

In this paper, we propose a new wavelet-based image fusion method, which combines pixel and region-based fusion algorithms using fuzzy logic. We specifically present new fusion rules to merge high and low frequency wavelet coefficients in an appropriate way in order to achieve the best quality in the fused image. The basic idea is to compute a dissimilarity measure of the source images to combine three different fusion rules to fuse high frequency wavelet coefficients, which are the weighted averaging, selection using pixel-based decision map, and selection using region-based decision map. In addition, we use an optimization-based method for low frequency fusion rule instead of simple averaging, because equal weighing of the input approximation sub-bands leads to the problem of contrast reduction. To evaluate the new fusion algorithm with other wavelet-based fusion methods, we use dual-tree discrete wavelet transform (DT-DWT), which introduces limited redundancy and allows the transform to provide approximate shift invariance and directionally selective filters, while preserving the usual properties of perfect reconstruction and computational efficiency [22,23].

The paper is structured as follows: in Section 2, an image segmentation algorithm using watershed transform is described. In Section 3, the proposed image fusion method is presented. Section 4 gives various results and comparisons. Finally, we conclude with a brief summary in Section 5.

2. Image segmentation algorithm

In this section, we describe an algorithm based on watershed segmentation for multi-valued images segmentation, especially for IR and visible images.

In the immersion-based watershed algorithm of [24], the gradient magnitude of an image is firstly calculated. This image is considered as a topographic relief in where the brightness value of each pixel corresponds to a physical elevation. The technique can simply be described by figuring that holes are pierced in each local minimum of the topographic relief. Eventually, the surface is slowly immersed into a 'lake', through that filling all the catchment basins, starting from the basin which is associated with the global minimum. As soon as two catchment basins tend to join, a dam is constructed. The procedure results in a partitioning of the image in many catchment basins of which the borders describe the watershed lines [25].

For multi-valued images, a single-valued method can be adopted by segmenting each band, separately, but jointly segmented images work better for the region-based image fusion algorithms. This is because the segmentation map will contain a smaller number of regions to represent all features in the both IR and visible images, and therefore reduces the computational cost.

Here we use a new method to obtain a joint gradient image from IR and visible images. First, a Gaussian derivative function is used to generate the gradient magnitude from the source images as:

$$G_S(x, y) = \sqrt{(S(x, y) * G'_x)^2 + (S(x, y) * G'_y)^2} \tag{1}$$

where $S(x, y)$ is the gray scale image, G'_x and G'_y are the Gaussian partial derivative filters in the x and y directions, and $*$ denotes convolution.

It is followed by a morphological opening, which operates similar to the local median filtering to reduce step noise in the gradient image. The opening is based on a $\min(\cdot)$ followed by a $\max(\cdot)$ operation on a local neighborhood around the pixels. Then

a joint gradient image (JG) is obtained from the source images using the following formula:

$$JG(x, y) = \frac{1}{2} \left(\frac{\tilde{G}_A(x, y)}{\max\limits_{x,y} \tilde{G}_A(x, y)} + \frac{\tilde{G}_B(x, y)}{\max\limits_{x,y} \tilde{G}_B(x, y)} \right) \tag{2}$$

where $\tilde{G}_A$ and $\tilde{G}_B$ are the filtered gradient images from the IR and visible images.

Finally, the joint gradient image is used to obtain segmentation map via watershed algorithm. The watershed algorithm is known to suffer from the over segmentation problem due to noise and other local irregularities of the gradient image. A practical solution to this problem is to limit the number of allowable regions by incorporating a preprocessing step, which is designed to bring additional knowledge into the segmentation process. An approach used to control over segmentation is based on the concept of markers. A *marker* is a connected component belonging to an image. There are *internal* markers, associated with the object of interest, and *external* markers, associated with the background [26].

Here, we have used a binary version of the joint gradient (JG) image as the external marker (EM), and its edges as the internal marker (IM). Then we have used a morphological reconstruction based method to limit the number of regional minima in the JG image to control the over segmentation problem as follows:

$$\overline{JG} = \text{imimposemin}(JG, EM\&IM) \tag{3}$$

where the "imimposemin" function in the Matlab software modifies the joint gradient image using morphological reconstruction, so it only has regional minima wherever (EM & IM) is nonzero [26], and & is the logical AND operator.

Now, the watershed algorithm is applied to the modified JG image. Thus, at the end, we have a joint segmentation map for the both IR and visible images (see Fig. 3).

3. The proposed image fusion algorithm

In this section, we present the proposed multi-sensor image fusion algorithm. Fig. 4 shows the block diagram of the proposed method, which consists of a number of essential stages:

1. A segmentation map is obtained using the marker-based watershed segmentation algorithm based on all the source images.
2. The source images are decomposed into different directions and scales using the DT-DWT.
3. The pixel-based decision maps are obtained through activities measurement of high frequency wavelet coefficients in each scale and direction of the DT-DWT.
4. Having the segmentation map, the region-based decision maps are generated from pixel-based decision maps.
5. High frequency wavelet coefficients of the source images are integrated using fuzzy logic, which combines outputs of three different fusion rules based on a dissimilarity measure of the source images: selection using pixel-based decision map, selection using region-based decision map, and weighted averaging.
6. Low frequency wavelet coefficients of the final fused image are obtained via weighted averaging, in which the weights are optimally estimated using the Particle Swarm Optimization algorithm.
7. The inverse DT-DWT of the new low and high frequency wavelet coefficients generates the final fused image.

In the following subsections, we have provided more detailed explications of the image fusion process.

Fig. 3. Segmentation of IR and visible images, from left to right: IR image, visible image, the joint gradient image, and segmentation map.
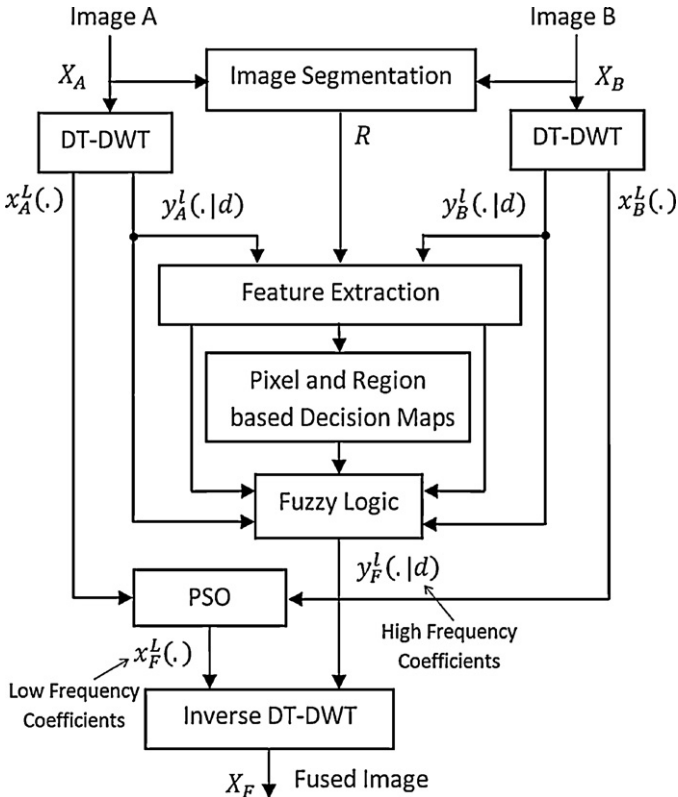

Fig. 4. Block diagram of the proposed image fusion algorithm.

3.1. Dual-tree discrete wavelet transform

The dual-tree discrete wavelet transform (DT-DWT) is a modified version of the discrete wavelet transform (DWT), and was proposed to conquer shift variance and directionality limitations of the DWT while maintaining the perfect reconstruction property with limited redundancy [22,23]. The DT-DWT is composed of two parallel DWT filter bank trees. The wavelet and scaling functions used in one tree can be defined as approximate Hilbert transforms of the functions in the other tree. The filters used in both trees are real, but the combined filters are referred to as analytic.

The DT-DWT employs two real DWTs, the first one (tree_a in Fig. 5) gives the real part of the transform while the second one (tree_b in Fig. 5) gives the imaginary part. The successful process of the DT-DWT is related to the differences between the filters in the two trees. To achieve optimal shift invariance, Kingsbury showed in Ref. [27] that the scaling basis functions for tree_b must be located halfway between those for tree_a at each level of the transform. In other words, the DT-DWT is achieved by a simple delay of one sample between the level 1 filters in each tree, and bye use of alternate odd-length and even-length linear-phase filters in the subsequent levels. These achieve the additional delay difference of half a sample

period, which is needed to obtain the correct total delay difference of one sample period at the sampling rate for that level. For example, at level 2, if filters H_{00a} and H_{01a} were even length, then H_{00b} and H_{01b} need to be odd length, and then vice versa for level 3, as shown in Fig. 5.

However, the odd/even filter method suffers from certain problems consisting of [27]: (a) the sub-sampling structure is not very symmetrical, (b) the two trees have slightly different frequency responses, and (c) the filter sets must be bi-orthogonal. Kingsbury proposed a Q-shift dual tree to overcome all problems above [27,28], in which all the filters beyond level 1 are even lengths, but they are no longer have strictly linear phase. They are designed to have a group delay of approximately 1/4 sample (q). Then the required delay difference of 1/2 sample ($2q$) is obtained using the time reverse of the tree_a filters in tree_b (i.e. the filters in the two trees are just the time reverse of each other, the same as are the analysis and reconstruction filters).

Moreover, because the filter coefficients are no longer symmetric, it is now possible to design the perfect reconstruction filter sets to be orthonormal, so that the reconstruction filters are just the time reverse of the equivalent analysis filters in both trees [27]. Therefore, all filters beyond level 1 are derived from the same orthonormal prototype set. This leads to a more symmetric sub-sampling structure, which preserves the key advantages of DT-DWT including approximate shift invariance and good directional selectivity. The symmetry of the sub-sampling process is important for hierarchical algorithms, which relate wavelet coefficients at one level to those at the same spatial location at the levels above or below. Therefore, we have used the Q-shift DT-DWT to decompose the input multi-sensor images. Here, the (9,7) filters were chosen for level 1 as a good compromise between smoothness and complexity (a much closer approximation to orthonormality at level 1 may be obtained from the (13,19)-tap filters) and Q-shift (14,14)-tap filters as the levels beyond [27].

This combination led to a complex extension of real signals. As the complex wavelets can distinguish between positive and negative frequencies, the diagonal sub-bands can be discriminated from horizontal and vertical sub-bands. Later on, horizontal and vertical sub-bands are divided giving six distinct sub-bands at each scale (at orientation $\pm 15°$, $\pm 45°$, and $\pm 75°$).

In this paper, the DT-DWT of an image x is denoted by y and is assumed in the different scales to be of the form:

$$y = \{y^1, y^2, \ldots, y^l, x^l\} \tag{4}$$

Here, x^l represents the approximation or low frequency sub-bands at the last decomposition level, while y^l represents the details or high frequency sub-bands at level l. In addition, y^l is composed of twelve directional sub-bands, six of which are real, and six are imaginary:

$$y^l = \begin{cases} real & y_{real}{}^l(i,j|1), y_{real}{}^l(i,j|2), \ldots, y_{real}{}^l(i,j|6) \\ imaginary & y_{imag}{}^l(i,j|1), y_{imag}{}^l(i,j|2), \ldots, y_{imag}{}^l(i,j|6) \end{cases} \tag{5}$$
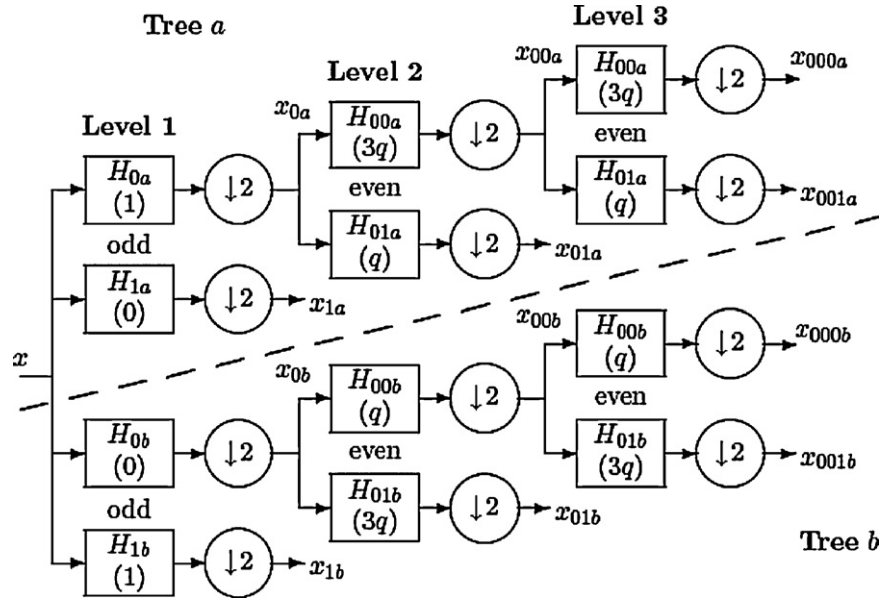
Fig. 5. The Q-shift version of the DT CWT, giving real and imaginary parts of complex coefficients from tree_a and tree_b, respectively. Figures in brackets indicate the delay for each filter, where $q = 1/4$ sample period [28].

We use the coordinates (i, j) or the shorthand notation of $(\cdot)$, to index the spatial position of the coefficients. It should be mentioned that we have used the DT-DWT toolbox version 4.3 provided by Kingsbury[1] for implementation.

3.2. Generation of pixel and region-based decision maps

In the infrared images, usually the regions do not contain texture information, and the information within them is borders of different areas. This is because of temperature differences between adjacent areas, which are measured by the infrared sensors. On the other hand, the regions in the visual images contain textures, and in special conditions – such as poor light condition and foggy weather – the borders of different areas are not correctly detected. Wavelet-based image fusion methods are motivated by the fact that the human visual system is primarily sensitive to the contrast changes or edges. Based on this fact, the information of the infrared image (borders of areas) with the information of the visual image (texture of areas) can be combined in the wavelet transform domain. Generation of decision map (DM) for selecting the more important coefficients in the wavelet transform domain between the infrared and visual images is the main idea.

The activity level measurement is used to judge the quality of a given part of each source image in the transform domain. Here, we use the local standard deviation as the window-based activity, which is obtained using the magnitude of complex wavelet coefficients of the source images:

$$y_{abs}^l(\cdot|d) = \sqrt{y_{real}^l(\cdot|d)^2 + y_{imag}^l(\cdot|d)^2} \qquad (6)$$

Then,

$$F^l(i, j|d) = \sqrt{\frac{1}{W} \sum_{n=-N}^{N} \sum_{m=-M}^{M} \left(\left| y_{abs}^l(i+n, j+m|d) \right| - \left| \bar{y}_{abs}^l(\cdot|d) \right| \right)^2} \qquad (7)$$

where $y_{imag}^l(\cdot|d)$ and $y_{real}^l(\cdot|d)$ are the imaginary and real high frequency wavelet coefficients, l is the decomposition level, and $d =$ 1, 2, . . . , 6 is the direction of high frequency wavelet coefficients (± 15, ± 45, ± 75). W is the number of pixels in the local window $[2N + 1, 2M + 1]$, and $(\cdot)$ is the shorthand notation for spatial position.

After extracting the local feature using the magnitude of complex wavelet coefficients, it is followed by nonlinear averaging in the local window for taking into account the neighbor dependency and smoothing on the feature space:

$$NF^l(i, j|d) = \sum_{n=-N}^{N} \sum_{m=-M}^{M} G(n, m) \times F^l(i+n, j+m|d) \qquad (8)$$

where F is the local feature, $G(n,m)$ is the local window's weight, which is obtained by the Gaussian filter.

After extracting local features, pixel-based decision map (PDM) in each decomposition level and direction is obtained using the following formula:

$$PDM^l(\cdot|d) = \begin{cases} 1 & \text{if } NF_A^l(\cdot|d) > NF_B^l(\cdot|d) \\ 0 & \text{otherwise} \end{cases} \qquad (9)$$

where $NF_A^l(\cdot|d)$ and $NF_B^l(\cdot|d)$ are the extracted features using (8) for the input source images, l is the decomposition level, $d = 1, 2, \ldots, 6$ is the direction of high frequency wavelet coefficients (± 15, ± 45, ± 75), and $(\cdot)$ is the shorthand notation for spatial position.

Region-based activity measurements are also used to obtain the decision map for selecting more important high frequency wavelet coefficients between source images, but it is very sensitive to the large coefficients in the region and is not accurate [18,19]. Therefore, we use a new method to generate the region-based decision map (RDM). First, a joint segmentation map based on all source images is obtained, which is explained in Section 2. Then, by looking into the labeled region, if numbers of one in the corresponding region within the PDM are more than numbers of zero, then the same region in the RDM is set to one, and otherwise the corresponding region in the RDM is set to zero. This is expressed in the following equation:

$$RDM^l(R|d) = \begin{cases} 1 & \text{if } \left(\sum_{n \in R} PDM^l(n|d) \right) > \frac{|R|}{2} \\ 0 & \text{otherwise} \end{cases} \qquad (10)$$

_____
[1] http://www.eng.cam.ac.uk/~ngk/.
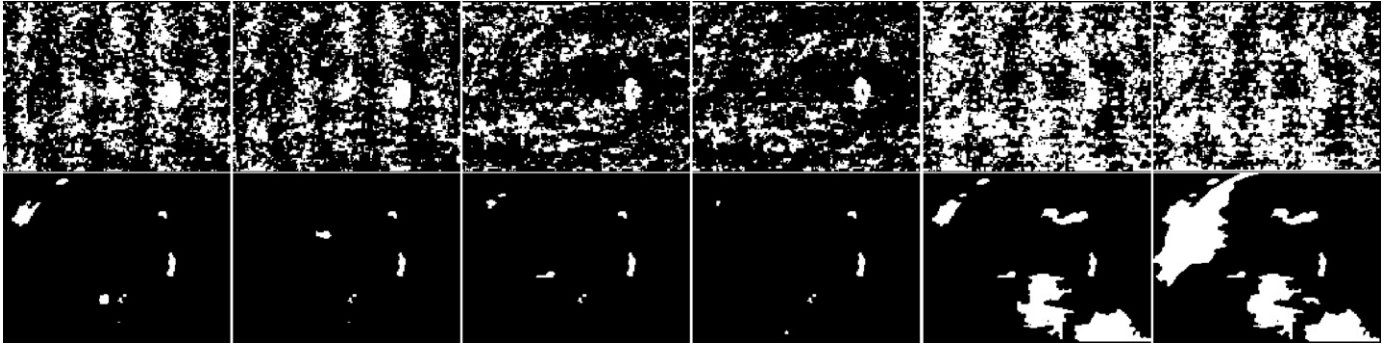
Fig. 6. Pixel- and region-based decision maps (first and second row) obtained from "UN Camp" images in the first decomposition level and six directional sub-bands of the DT-DWT, from left to right: −15, +15, −45, +45, −75 and +75 directions, respectively. Black regions indicate that high frequency wavelet coefficients in the DT-DWT sub-bands of the final fused image should be selected from the visible image, and white regions indicate that high frequency wavelet coefficients in the DT-DWT sub-bands of the final fused image should be selected from the IR image.

where R is the region with size $|R|$, PDM is the pixel-based decision map using (9), l is the level of decomposition, and $d = 1, 2, \ldots, 6$ is the direction of high frequency wavelet coefficients ($\pm 15, \pm 45, \pm 75$).

Fig. 6 shows pixel- and region-based decision maps obtained from IR and visible "UN Camp" images in the first decomposition level and six directional sub-bands of DT-DWT.

3.3. Fusion rule for high frequency wavelet coefficients

Having the pixel and region-based decision maps, there are two strategies for fusion of high frequency wavelet coefficient. Indeed, we can use either the PDM or the RDM for selecting more important high frequency wavelet coefficients between the source images using the following:

$$y_1^l(\cdot|d) = \text{PDM}^l(\cdot|d) \times y_A^l(\cdot|d) + (1 - \text{PDM}^l(\cdot|d)) \times y_B^l(\cdot|d) \quad (11)$$

and

$$y_2^l(\cdot|d) = \text{RDM}^l(\cdot|d) \times y_A^l(\cdot|d) + (1 - \text{RDM}^l(\cdot|d)) \times y_B^l(\cdot|d) \quad (12)$$

where $y_A^l(\cdot|d)$ and $y_B^l(\cdot|d)$ are the high frequency wavelet coefficients of source images, l is the level of decomposition, $d = 1, 2, \ldots, 6$ is the direction of high frequency wavelet coefficients ($\pm 15, \pm 45, \pm 75$), PDM and RDM are the pixel and region-based decision maps using (9) and (10), and ($\cdot$) is the shorthand notation for spatial position.

As it can be seen in Fig. 6, using region-based decision map (RDM) for fusing high frequency wavelet coefficients using (12) will eliminate a lot of information from the infrared image in the final fused image. On the other hand, using pixel-based decision map (PDM) for fusion of high frequency wavelet coefficients using (10) affected by problems such as noise, image misregistration, and usually results in artifacts in the fused image. In addition, in some regions, we cannot make a good decision for selecting more important wavelet coefficients between the source images, because there is not enough difference between the extracted features from these regions. Therefore, a weighted averaging rule is proposed for these pixels:

$$y_3^l(\cdot|d) = \frac{NF_A^l(\cdot|d) \times y_A^l(\cdot|d) + NF_B^l(\cdot|d) \times y_B^l(\cdot|d)}{NF_A^l(\cdot|d) + NF_B^l(\cdot|d)} \quad (13)$$

where $y_A^l(\cdot|d)$ and $y_B^l(\cdot|d)$ are the high frequency wavelet coefficients of source images, l is the level of decomposition, $d = 1, 2, \ldots, 6$ is the direction of high frequency wavelet coefficients ($\pm 15, \pm 45, \pm 75$), $NF_A^l(\cdot|d)$ and $NF_B^l(\cdot|d)$ are the extracted features using (8) for the input source images A and B, and ($\cdot$) is the shorthand notation for spatial position.

None of the three fusion rules expressed in (11)–(13) results in a good fused image, individually. Indeed, we use fuzzy logic to design a good fusion method by combining the three fusion rules to transfer as much information as possible from the source images into the fused image. To this end, a dissimilarity measure (DIS) of the source images is defined. The DIS is intended to quantify the degree of 'dissimilarity' between the source images. In the following expression, this measure is defined as:

$$DNF^l(i, j|d) = |NF_A^l(i, j|d) - NF_B^l(i, j|d)| \quad (14)$$

Then,

$$\text{DIS}^l(i, j|d) = \sqrt{\sin\left(\frac{\pi \times DNF^l(i, j|d)}{2 \times T}\right)} \quad (15)$$

where $NF_A^l(\cdot|d)$ and $NF_B^l(\cdot|d)$ are the extracted features using (8) for the input source images A and B, l is the level of decomposition, $d = 1, 2, \ldots, 6$ is the direction of high frequency wavelet coefficients ($\pm 15, \pm 45, \pm 75$), and $T = \max_{i,j} DNF^l(i, j|d)$. Indeed, we have used sinusoid function in (15), in order to normalize the dissimilarity measure between zero and one.

By analyzing the DIS measure, we can determine where and to what extent the source images differ, and use this information to combine the three different fusion rules expressed in (11)–(13). First, we define the following linguistic rules for fusion process, which are obtained based on our observation:

IF the DIS measure at a given position is high (i.e. the sources are distinctly different at that position) **THEN** we use the first fusion rule (selection using the pixel-based decision map).
IF the DIS measure at a given position is medium (i.e. the sources are different at that position) **THEN** we use the second fusion rule (selection using the region-based decision map).
IF the DIS measure at a given position is low (i.e. the sources are similar at that position) **THEN** we use the third fusion rule (weighted averaging).

Then, for constructing standard rules from linguistic ones, we define small, medium, and large membership functions. Fuzzy membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value or degree of membership between zero and one. MF is often given the designation of μ. Here, we use the triangular membership function, which is shown in Fig. 7. In order to construct the three fuzzy membership functions, one threshold (T in Fig. 7) must be defined. The threshold should be set to partition the feature space (DIS) into the three regions of linguistic variables (i.e. small, medium, and large) in an appropriate way. In order to find the best value for the threshold,
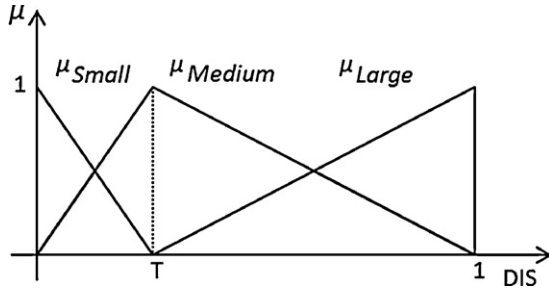
Fig. 7. Triangular fuzzy membership functions.

we have performed an experiment using different datasets. Fig. 8 shows Petrovic index of fused image versus different thresholds obtained from different multi-sensor datasets [29]. The Petrovic index considers the amount of edge information transferred from the input images to the fused image using the Sobel edge detector to calculate the strength and orientation information at each pixel in both source and the fused images. Based on the experiment, the threshold is selected which maximizes the Petrovic index of final fused image for different datasets ($T \approx 0.2$). It should be mentioned that the averaging fusion rule is used to fuse low frequency wavelet coefficients of the source images in this experiment.

Finally, the fusion rule for integrating high frequency wavelet coefficients from the source images is defined as follows:

$$y_F^l(\cdot|d) = \sum_{i=1}^{3} y_i^l(\cdot|d) \times \mu_i(\mathrm{DIS}^l(\cdot|d)) \qquad (16)$$

where $y_1^l(\cdot|d)$, $y_2^l(\cdot|d)$ and $y_3^l(\cdot|d)$ are obtained using (11)–(13), μ_1, μ_2 and μ_3 are the large, medium and small membership functions, respectively, l is the level of decomposition, $d = 1, 2, \ldots, 6$ is the direction of high frequency wavelet coefficients (± 15, ± 45, ± 75), and $(\cdot)$ is shorthand notation for spatial position.

3.4. Fusion rule for low frequency wavelet coefficients

The information contained in the low frequency (LF) coefficients is vital for the natural appearance of an image, because LF coefficients contain the bulk of the energy of an image. Commonly, averaging is a popular method to fuse LF sub-bands of the source images in most wavelet-based methods:

$$x_F^L(i,j) = \frac{x_A^L(i,j) + x_B^L(i,j)}{2} \qquad (17)$$

where $x_F^L(i,j)$, $x_A^L(i,j)$ and $x_B^L(i,j)$ are the fused and input LF wavelet coefficients, and L represents the coarsest resolution level.

However, equal weighing of the input approximation sub-bands leads to the problem of contrast reduction, because IR and visible images have different contrast conditions. Petrovic and Xydeas proposed a LF fusion rule, which is performed using an arithmetic combination of the input LF coefficients [30]. Specifically, the offset zero-mean addition is used, which is defined as:

$$x_F^L(i,j) = x_A^L(i,j) + x_B^L(i,j) - \frac{\mu_A + \mu_B}{2} \qquad (18)$$

where μ_A and μ_B are the mean values of the two input LF sub-bands, and L represents the coarsest resolution level. Like other arithmetic fusion methods, the fusion defined in (18) is weak to the destructive superposition, especially if the input LF sub-bands have opposing illumination levels [30].

An alternative approach is a weighted averaging proposed in Ref. [21], in which the weights are obtained using a region-based activity measurement (the normalized Shannon entropy) of the low frequency wavelet coefficients:

$$W_S(R) = \frac{1}{|R|} \sum_{n \in R} x_S^{L2}(n) \times \log_2 x_S^{L2}(n) \qquad (19)$$

where R is the region with size $|R|$, and x_S^L are the input LF wavelet coefficients, and L represents the coarsest resolution level. Hence, the composite LF sub-band is generated using:

$$x_F^L(i,j) = \frac{W_A(i,j) \times x_A^L(i,j) + W_B(i,j) \times x_B^L(i,j)}{W_A(i,j) + W_B(i,j)} \qquad (20)$$

where $x_F^L(i,j)$, $x_A^L(i,j)$ and $x_B^L(i,j)$ are the fused and input LF wavelet coefficients, $W_A(i,j)$ and $W_B(i,j)$ are obtained using (19), and L represents the coarsest resolution level.

Averaging, arithmetic combination [30], and weighted averaging [21] fusion rules cannot lead to an optimal result. In this paper, we have proposed a new weighted averaging rule for the fusion of the LF wavelet coefficients, in which the weights are estimated using an optimization algorithm:

$$x_F^L(i,j) = W_1(i,j) \times x_A^L(i,j) + W_2(i,j) \times x^L(i,j) \qquad (21)$$

where $x_F^L(i,j)$, $x_A^L(i,j)$ and $x_B^L(i,j)$ are the fused and input LF wavelet coefficients, W_1 and W_2 are the values in the range [0,1], and $W_1 + W_2 = 1$.

An optimization algorithm finds the optimal weights with maximizing the entropy index from the final fused image. The entropy metric measures the information content in an image. An image
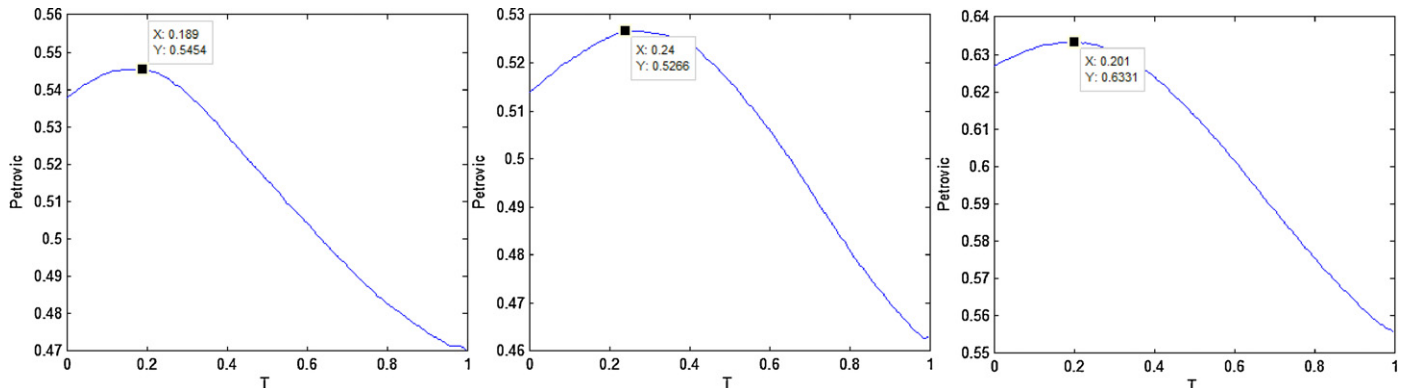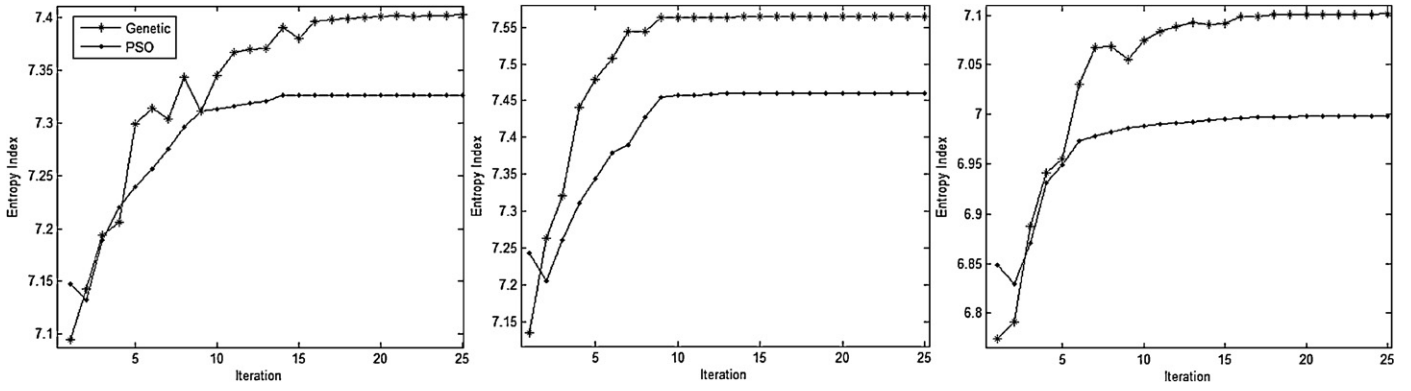

Fig. 8. The Petrovic index of fused images versus different thresholds (T for building fuzzy membership functions) obtained from different multi-sensor datasets, from left to right: "UN Camp", "Dune", and "Trees".

Fig. 9. The entropy index of fused images versus iteration of the PSO and genetic algorithms obtained from different multi-sensor datasets, from left to right: "UN Camp", "Dune", and "Trees".

Table 1
The entropy index of fused images obtained from different datasets using different LF fusion rules in the proposed fusion scheme.

Low frequency fusion rule	UN Camp	Dune	Trees
Averaging	6.63	6.85	6.43
Weighted averaging [21]	6.85	7.11	6.61
arithmetic combination [30]	6.97	7.21	6.71
PSO	7.30	7.45	7.01
Genetic	7.41	7.56	7.11

with high information content will have high entropy. The entropy index is defined as:

$$H = -\sum_{i=0}^{K} h(i) \times \log_2(h(i)) \tag{22}$$

where h is the normalized histogram of reconstructed or fused image, and K is the number of gray levels in the image. In addition, the final fused image is obtained using inverse DT-DWT:

$$X_F = \psi^{-1}(x_F^L, y_F) \tag{23}$$

in which ψ is the inverse DT-DWT, y_F and x_F^L are the composite high and low frequency wavelet coefficients, which are obtained using (16) and (21).

3.5. Optimization algorithm

Over the years, many real-parameter optimization algorithms have been developed by using point-by-point [31,32] as well as multi-point approaches [33,34]. While a point-by-point approach begins with one guessed solution and updates the solution iteratively for reaching near the optimum solution, a multi-point method deals with a number of solutions in each iteration. Starting with a number of guessed solutions, the multi-point algorithm updates one or more solutions in a synergistic way for steering the population toward the optimum. In this paper, we focus our attention on the multi-point optimization algorithms, which we refer here as the population-based optimization algorithms.

Here, we have considered two population-based search algorithms (Particle Swarm Optimization and genetic algorithm (GA)) for the optimization problem. Fig. 9 shows the entropy index of fused images obtained from different IR and visible datasets at different iterations of PSO and different generation of genetic algorithms for the optimization problem. In addition, Table 1 shows the entropy index for comparison between different alternative LF fusion rules in the proposed fusion scheme. The high amount of the entropy index for the proposed method indicates that the fused image contains higher information compared to other methods. In

addition to objective evaluation, a sample image has been selected for comparing different LF fusion rules, subjectively (see Fig. 10). As it can be seen in Fig. 10, the image contrast in the proposed method is far better compared to other methods.

Generally, the results indicate that both GAs and PSO can be used in the optimization of parameters during image fusion. In terms of computational effort, the GA approach is faster, although it should be noted that neither algorithms take what can be considered an unacceptably long time to determine their results. With respect to accuracy of model parameters, the GA determines values, which are better than what does the PSO. Thus, it must be concluded that for the process of optimization, the GA approach is superior to the PSO approach. Techniques such as PSO and GA are inspired by the nature, and have proved themselves to be effective solutions to optimization problems. Appropriate setting of control parameters of PSO and GA algorithms is a key point for their success. In general, some form of trial-and-error tuning is necessary for each particular instance of optimization problem. In the following sub-sections, we briefly describe PSO and genetic algorithms.

3.5.1. Particle Swarm Optimization

Kennedy and Eberhart [33] firstly proposed the PSO algorithm for optimization. PSO is a population-based search algorithm based on the simulation of the social behavior of birds within a flock. In the PSO, particles are flown through hyper-dimensional search space. Changes to the position of the particles within the search space are based on the social-psychological tendency of individuals to emulate the success of other individuals [33].

The position of each particle is changed according to its own experience and that of its neighbors. Let $x_i(t)$ denotes the position of particle p_i, at iteration t. The position of p_i is then changed by adding a velocity $v_i(t)$ to the current position, i.e.:

$$x_i(t) = x_i(t-1) + v_i(t) \tag{24}$$

The velocity vector reflects the socially exchanged information, and is generally defined in the following way:

$$\begin{aligned} v_i(t) = &\, W \times v_i(t-1) \\ &+ C_1 r_1 \times (x_{pbest_i}(t) - x_i(t)) + C_2 r_2 \times (x_{leader}(t) - x_i(t)) \end{aligned} \tag{25}$$

where W is the inertia weight employed to control the impact of the previous history of velocities on the current velocity of a given particle, and x_{pbest} is the personal best position of a given particle, so far. That is, the position of the particle that has provided the greatest success. x_{leader} is position of the best particle of the entire swarm. C_1 and C_2 are learning factors and represent the attraction that a particle has toward either its own success or that of its neighbors, and $r_1, r_2 \in [0,1]$ are random values.
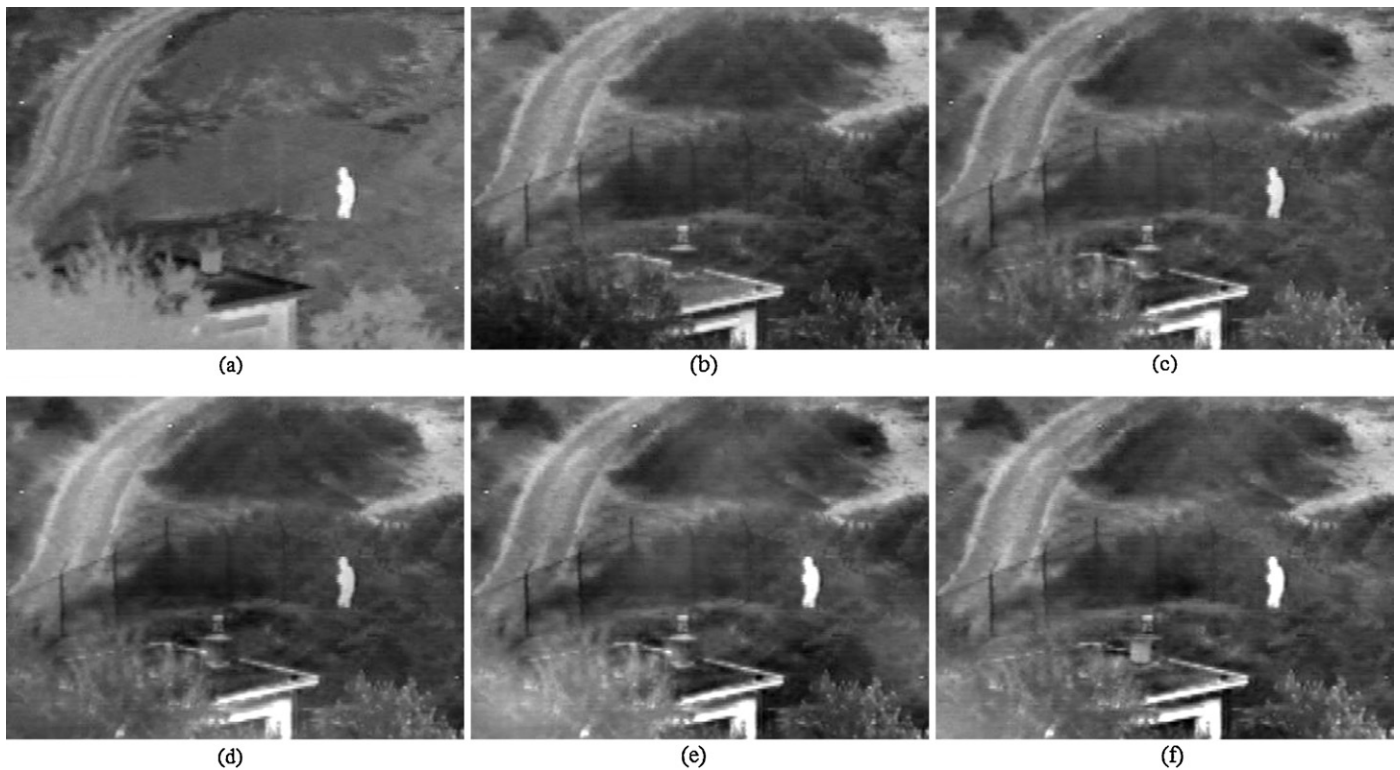
Fig. 10. Subjective image fusion results using different low frequency fusion rule in the proposed fusion scheme. (a) "UN Camp" IR image, (b) visible image, and fused images using (c) averaging, (d) weighted averaging [21], (e) arithmetic combination [30] and (f) the proposed method based on PSO algorithm.

The role of the inertia weight W, in (25), is critical for the PSO's convergence behavior. The inertia weight is employed to control the impact of the previous history of velocities on the current one. Accordingly, the parameter W adjusts the trade-off between the global and local exploration abilities of the swarm. A large inertia weight facilitates global exploration (searching new areas), while a small one tends to facilitate local exploration, i.e. fine-tuning the current search area. A suitable value for the inertia weight W usually provides balance between global and local exploration abilities and consequently results in a reduction of the number of iterations required to locate the optimum solution. Initially, the inertia weight was constant. However, experimental results indicated that it is better to initially set the inertia to a large value, in order to promote global exploration of the search space, and gradually decrease it to get more refined solutions. Thus, an initial value around 1.2 and a gradual decline toward 0 can be considered as a good choice for W. In Refs. [35,36] a decaying inertia weight is proposed and tested, with the aim of favoring global search at the start of the algorithm and local search later.

The parameters C_1 and C_2, in Eq. (25), are not critical for PSO's convergence. However, proper fine-tuning may result in faster convergence and alleviation of local minima. An extended study of the acceleration parameter in the first version of PSO is given in Ref. [37]. As default values, $C_1 = C_2 = 2$ were proposed, but experimental results indicate that $C_1 = C_2 = 0.5$ might provide even better results. Recent work reports that it might be even better to choose a larger cognitive parameter, C_1, than a social parameter, C_2, but with $C_1 + C_2 = 4$ [38]. The parameters r_1 and r_2 are used to maintain the diversity of the population, and they are uniformly distributed in the range [0,1].

Here, default values for the parameters C_1 and C_2 were used: $C_1 = 2.4$, $C_2 = 1.6$. Although the choice of the parameter values seems not to be critical for the success of the methods, it appears that

Table 2
The PSO parameters selected for the optimization.

Operation	Parameters
Particle number	32
Inertia weight	$W_{init} = 1.2$, $W_{final} = 0.4$
Learning factors	$C_1 = 2.4$, $C_2 = 1.6$
Maximum iteration	30

faster convergence can be obtained by proper fine-tuning. The balance between the global and local exploration abilities of the PSO is mainly controlled by the inertia weight, since the positions of the particles are updated according to the classical PSO strategy. A time decreasing inertia weight value, i.e. start from 1.2 and gradually decrease toward 0.4, proved to be superior than a constant value. This is because the larger inertia weights at the beginning help to find good seeds and the later small inertia weights facilitate a fine search.

In addition, the number of iterations and particles is set based on convergence of the PSO algorithm. More particles could speed up the convergence of the PSO algorithm; however, it increases its computation time. In addition, using too much particles and iterations is ineffective on the convergence of PSO algorithm. Table 2 shows entropy index obtained for different runs of PSO algorithm using different particles number. The number of iterations is set to 25 for this experiment. Fig. 11(a) shows the entropy index of fused image obtained from "UN Camp" IR and visible dataset at different iterations of the PSO algorithm using different particles number. Based on this experiment, we have chosen 32 particles in the PSO algorithm, as a good compromise between accuracy and complexity. In summary, the PSO parameters selected for the optimization problem are listed in Table 2.
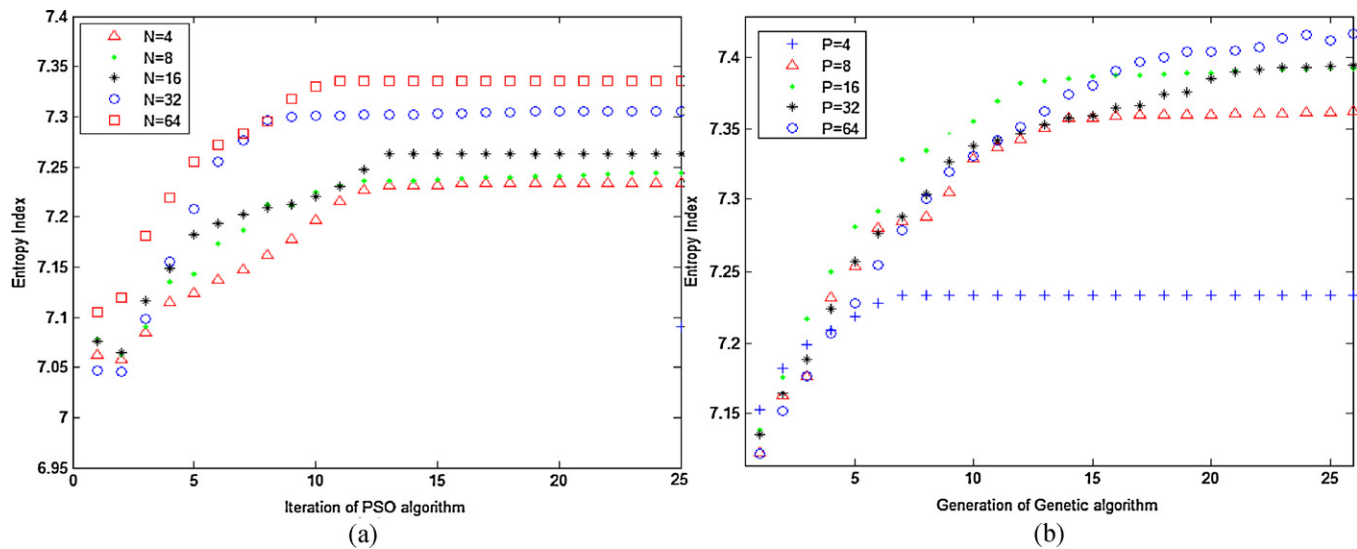
Fig. 11. The entropy index of fused image obtained from "UN Camp" dataset at (a) different iterations of the PSO algorithm using different particles number and (b) different generation of genetic algorithm using different population sizes.

3.5.2. Genetic algorithm

In genetic algorithm (GA) [34], a candidate solution for a specific problem is called an individual or a chromosome and consists of a linear list of genes. Each individual stands for a point in the search space, and therefore a possible solution to the problem. A population consists of a finite number of individuals. Each individual is decided by an evaluating system to obtain its fitness value. Based on this fitness value and undergoing genetic operators, a new population is iteratively generated with each successive population referred to as a generation.

Three basic genetic operators are sequentially applied to each individual with certain probabilities during each generation, i.e. selection, crossover (recombination), and mutation. First, a number of best individuals are picked based on a user defined fitness function. The remaining individuals are discarded. Next, a number of individuals are selected and paired with each other. Each individual pair generates one offspring by partially exchanging their genes around one or more randomly selected crossing points. At the end, a certain number of individuals are selected and the mutation operations are applied (i.e. a randomly selected gene of an individual abruptly changes its value). The GA is called a population-based technique because instead of operating on a single potential solution, it uses a population of potential solutions. The larger the population, the greater the diversity of the members of the population, and the larger the area searched by the population. Here, we have used Matlab optimization toolbox for implementation of genetic algorithm. Based on the experimental results (see Fig. 11(b)) the population size is set to 16 and the generation size is set to 20 for genetic optimization as a good compromise between accuracy and complexity.

Table 3
Relative computation time of different image fusion techniques.

| Method	Unit of time
Averaging	0.332
LT [10]	0.771
MS	0.739
WA [15]	0.821
WBV [16]	0.984
Lewis [18]	1.703
Proposed	2.253

3.6. Computation time

It is interesting to evaluate the proposed method from a practical point of view that is the computation time. The proposed method consists of two steps with different computation times. In the first step, high frequency wavelet coefficients of the fused image are obtained using a simple fuzzy rule, and therefore it has low computation time. In the second step, a population-based optimization algorithm is used to generate low frequency wavelet coefficients of the fused image, and consequently the optimization algorithm increases the computation time.

Table 3 summarizes the relative computation time of the various image fusion methods considered in this paper. All computations have been performed on a Pentium IV personal computer, using a 2.00-GHz processor, running Windows Vista. The low computation times of different image fusion methods listed in Table 3 are because of their simplicity. In addition, the computation time of the proposed method can be better in some ways.

Table 4
The objective results of final fused images using different decomposition levels in the proposed algorithm.

Metric	Dataset	N							
		1	2	3	4	5	6	7	8
Petrovic	UN Camp	0.162	0.200	0.297	0.421	0.495	0.532	0.552	0.481
	Dune	0.224	0.267	0.364	0.452	0.504	0.534	0.537	0.443
	Trees	0.248	0.316	0.436	0.539	0.592	0.614	0.630	0.520
Entropy	UN Camp	6.869	6.940	6.980	7.031	7.164	7.246	7.305	7.223
	Dune	7.065	7.122	7.169	7.253	7.363	7.433	7.453	7.671
	Trees	6.577	6.667	6.706	6.786	6.905	7.011	7.051	7.145
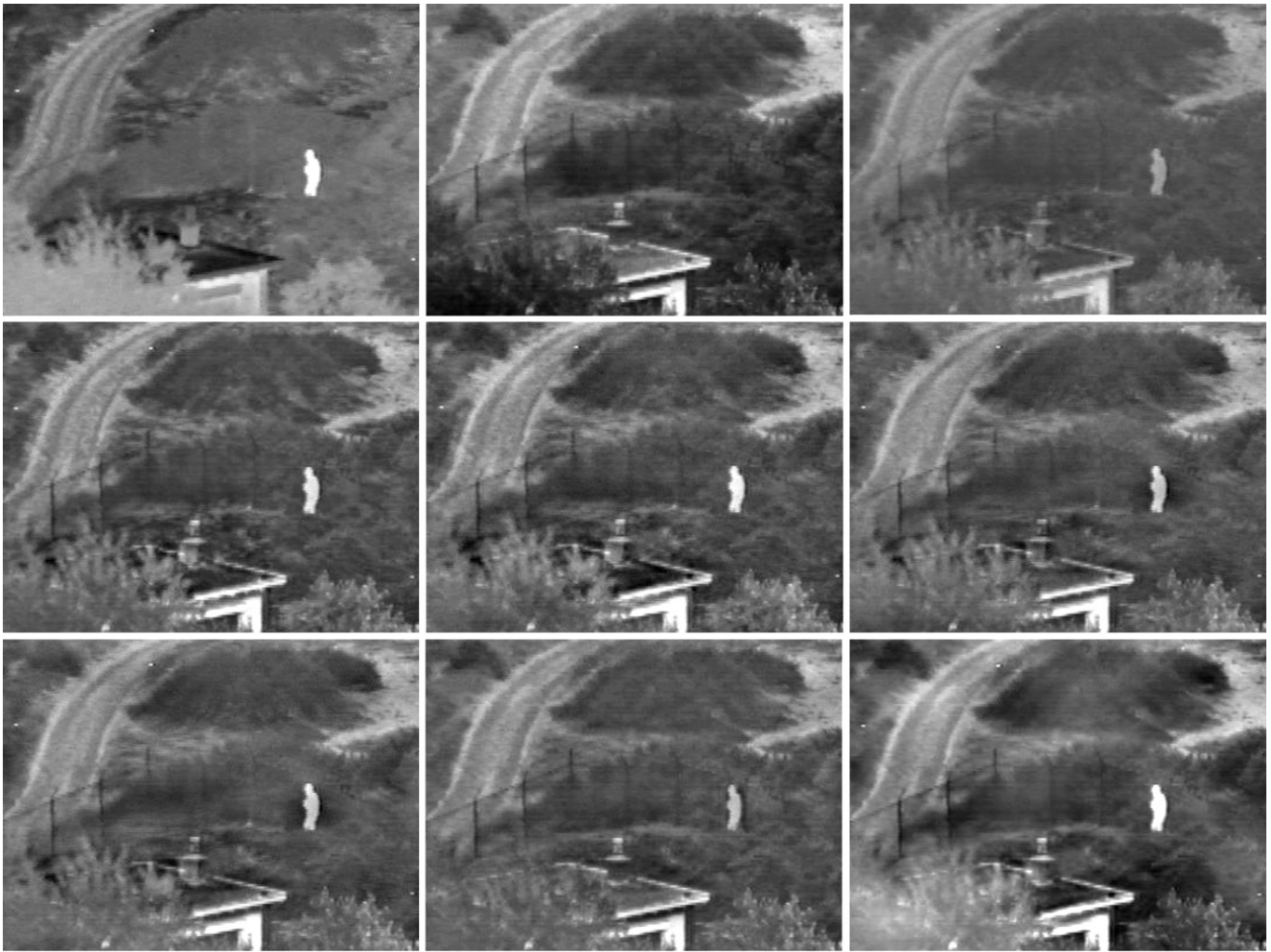
Fig. 12. Subjective fusion results of "UN Camp" images: original IR image, original visible image, fused image using averaging method (top-row: left to right), fused images using LT [10], MS, WA [15] methods (middle-row), and fused images using WBV [16], Lewis [18], the proposed methods (bottom-row).

Since the proposed method is fully implemented using Matlab software, it is possible to use an optimum programming language such as C++ to reduce the computation time. In addition, a population-based algorithm can be used in parallel optimization form or synchronous algorithm, which speeds up the algorithm. A parallel optimization algorithm allows higher throughput (solving more complex problems in the same time span), more sophisticated finite element formulations, and higher accuracy.

4. Experimental results

The proposed image fusion method was tested against several state-of-the-art image fusion methods including the simple averaging, the Laplacian Transform[2] (LT) [10], the dual-tree discrete wavelet transform with three different fusion rules including maximum selection (MS), weighted average (WA) [15], window based verification (WBV) [16], and Lewis's region-based algorithm [18]. In the LT method five decomposition levels is used for image decomposition and maximum absolute selection rule is used as the high frequency fusion rule. For the DT-DWT based methods, the

available "AntonB" mother wavelet is used for the filter banks and five decomposition levels is used for image decomposition. In addition, the simple averaging is used for the low frequency fusion rule in the LT and DT-DWT based methods.

The images used in experiments are surveillance images from TNO Human Factors, publicly available at the Image Fusion web site [40]. Image sequence "UN Camp" consists of 32 pair images and the image sequences "Trees" and "Dune" contain 19 pair images. In addition, we have applied the contrast stretching to enhance the representation of the source image. Other image processing techniques like histogram equalization can also be used.

Two metrics are considered in this paper, which do not require ground truth images for evaluation. The first metric is the Xydeas and Petrovic index, which is proposed in Ref. [29], which considers the amount of edge information transferred from the input images to the fused image using a Sobel edge detector to calculate the strength and orientation information at each pixel in both source and the fused images. The second metric is the entropy index, which measures the information content in an image. An image with high information content will have high entropy. The entropy index is defined using (21).

It should be mentioned that the original size of the source images is 270×360. Because of using the dual-tree discrete wavelet transform in the proposed algorithm, the images size should be a factor

---

[2] Image Fusion Toolbox for MATLAB developed by Oliver Rockinger [39]: http://www.metapix.de/toolbox.htm.
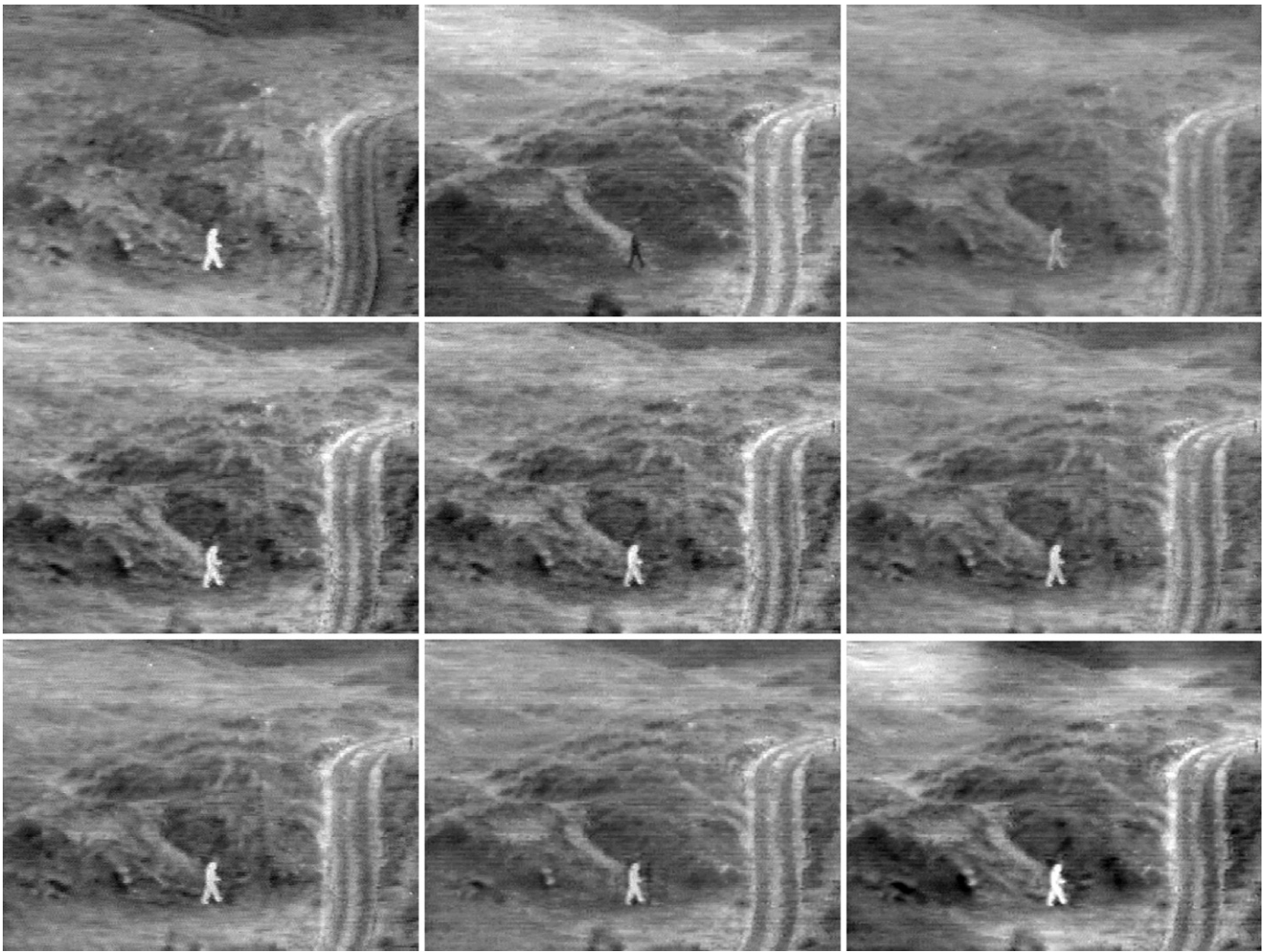
Fig. 13. Subjective fusion results of "Dune" images: original IR image, original visible image, fused image using averaging method (top-row: left to right), fused images using LT [10], MS, WA [15] methods (middle-row), and fused images using WBV [16], Lewis [18], the proposed methods (bottom-row).

of two. Therefore, we have generated 512×512 images by repeating border elements of original images symmetrically. The possible decomposition levels for a 512×512 image is $N = \log_2 512 = 8$. Different decomposition levels result in different performance of the proposed algorithm. In order to obtained best decomposition level, we have studied the effect of different decomposition levels on the performance of fused images. Table 4 shows the objective results of final fused images using different decomposition levels in the proposed algorithm. It can be seen in Table 4 that the effect of decomposition levels on the proposed algorithm is more noticeable for the Petrovic index compared to entropy index. The Petrovic index considers the amount of edge information transferred from the input images to the fused image [29]. The experimental results demonstrate that for $N - 1$ decomposition levels, the best results are obtained for different datasets.

Figs. 12–14 show the results of different image fusion algorithms for the "UN Camp", "Dune" and "Trees" multi-sensor images, respectively. Subjective or visual comparison between different methods indicates superiority of the proposed method against other image fusion algorithms. In Fig. 12, it is clear that the fence details from the visible image and the person details from the IR image are far better transferred into the fused image in the proposed method than in the other algorithms. The fused image obtained by the averaging method is dark and image contrast is

low. The different contrast conditions of the IR and visible images cause this problem. While the fused images obtained by the previous pixel-based methods in the multi-resolution domain, i.e. LT, MS, WA, and WBV algorithms have low contrast conditions, the fused images also have distortions (darkness) around the person and the roof. In addition, these methods are sensitive to the noise and image misregistration.

In the fused image obtained by Lewis's region-based method, which is also implemented in the dual-tree discrete wavelet transform domain, some details such as the contours of trees and the bright points are not transferred into the fused image. In Fig. 13, the roads' details from the visible image and the person's details from the IR image, and also in Fig. 14 the trees' details from the visible image and the person's details from the IR image, are better transferred into the fused image in the proposed method. Generally, it can be seen in Figs. 12–14 that the person is brighter in the proposed method, and the contrast of the fused images is far better compared to other methods.

Table 5 shows the average performance results from different image fusion methods and different datasets. It should be mentioned that the obtained results are different from the results listed in previous publications, because here we have applied the contrast stretching to enhance the representation of the source images.
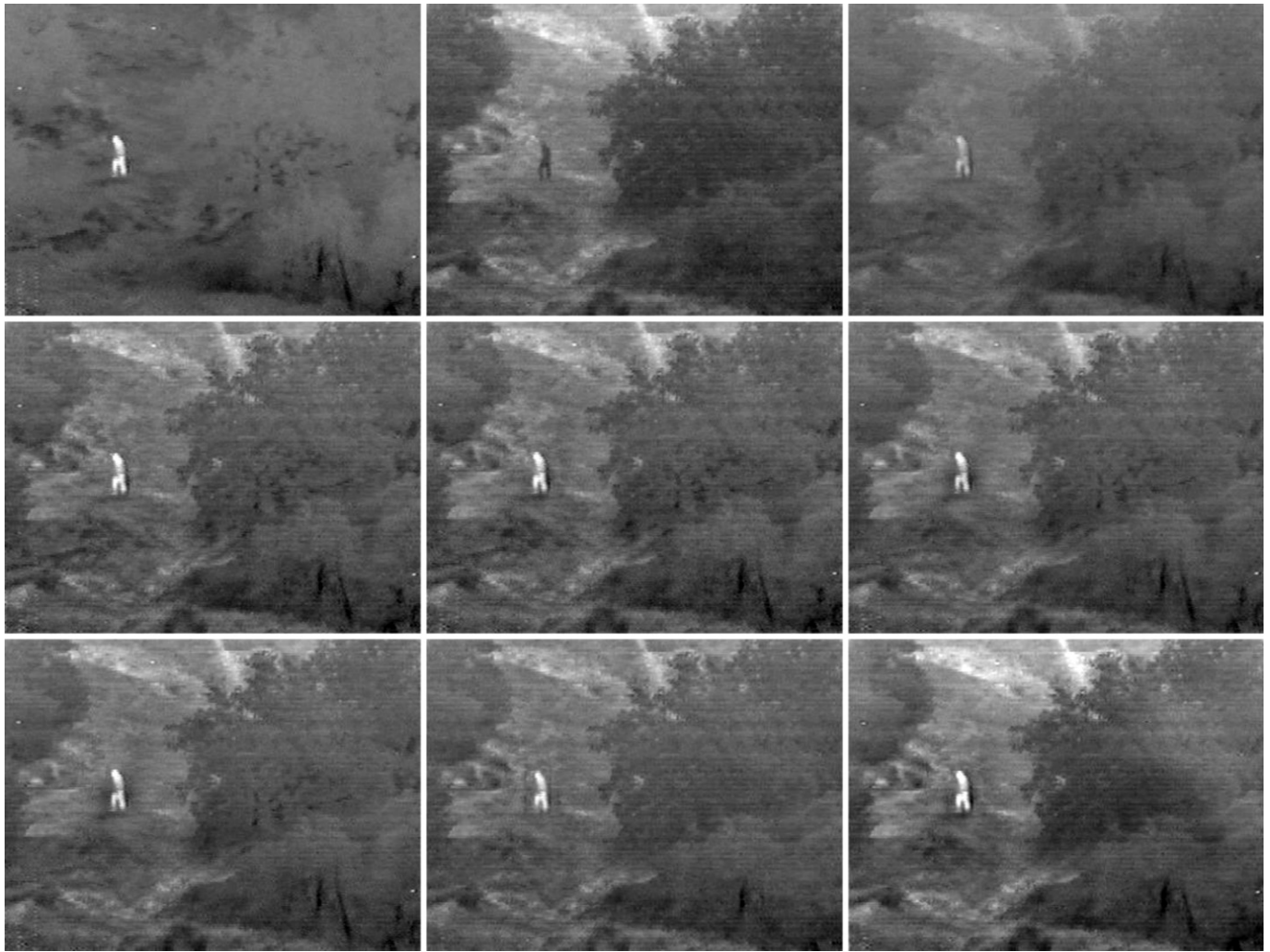
Fig. 14. Subjective fusion results of "Trees" images: original IR image, original visible image, fused image using averaging method (top-row: left to right), fused images using LT [10], MS, WA [15] methods (middle-row), and fused images using WBV [16], Lewis [18], the proposed methods (bottom-row).

The metric's values confirm the subjective assessment, that our proposed image fusion algorithm generally integrates more details from the visible and IR images into the fused images. It can be seen in Table 5 that the entropy index for the proposed method is better compared to other fusion methods, which indicates that the fused image obtained by the proposed method has more visual information and better contrast compared to the fused images using other fusion algorithms. In addition, evaluation of different methods using Petrovic metric has shown that the proposed algorithm has better performance for transferring edges information from the source images to the corresponding fused image.

5. Conclusions

In this paper, we have presented a new wavelet based multi-sensor image fusion method using fuzzy logic and population-based optimization. Proposing new fusion rules for merging high and low frequency wavelet coefficients, which is the second step in the wavelet-based image fusion, is the main novelty of this paper.

We specifically have presented a fuzzy-based approach for fusing high frequency wavelet coefficients of the IR and visible images, which combines the outputs of three different fusion rules based on a dissimilarity measure of the source images. This new method applies the advantages of the previous pixel and region-based methods using fuzzy logic. Proposing an optimization method for fusing low frequency wavelet coefficients of the source image is another novelty of this paper, which applies the population-based optimization in the low frequency fusion scheme to maximize the entropy of the final fused image.

In addition, the proposed method used the DT-DWT for finer frequency decomposition and shift invariant property compared

Table 5
Average performance results of different image fusion methods.

Metric	Method	UN Camp	Dune	Trees
Petrovic	Averaging	0.332	0.338	0.351
	LT [10]	0.471	0.451	0.541
	MS	0.439	0.409	0.472
	WA [15]	0.421	0.396	0.494
	WBV [16]	0.424	0.390	0.492
	Lewis [18]	0.503	0.485	0.553
	Proposed	0.552	0.530	0.630
Entropy	Averaging	6.29	6.75	6.47
	LT [10]	6.71	7.03	6.61
	MS	6.79	7.01	6.63
	WA [15]	6.72	6.92	6.58
	WBV [16]	6.76	6.97	6.57
	Lewis [18]	6.66	6.88	6.54
	Proposed	7.41	7.56	7.11

to discrete wavelet transform. The experimental results demonstrated that the proposed method outperforms the standard fusion methods in the fusion of IR and visible images.

In this paper, we did not investigate the robustness of the proposed method to the noise. In the proposed algorithm, we have used an image segmentation algorithm, which is sensitive to the noise and for noisy images results in oversegmentation with incorrect regions. In addition, we have not considered the noise parameters for the high frequency fusion rule, and therefore the results will be sensitive to the noise. As a result, we have to use an image denoising algorithm as a pre-processing step to work with the proposed algorithm in noisy situation. One aspect, which we would like to explore in the future, is to design a robust image fusion algorithm for fusing noisy multi-sensor images, in which the two processes of image fusion and denoising are considered all together.

References

[1] S. Garg, U. Kiran, K. Mohan, R. Tiwary, Multilevel medical image fusion using segmented image by level set evolution with region competition, in: 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2006, pp. 7680–7683.
[2] J. Saeedi, K. Faez, A new pan-sharpening method using multiobjective particle swarm optimization and the shiftable contourlet transform, Photogrammetry & Remote Sensing 66 (2011) 365–381.
[3] M. Kam, X. Zhu, P. Kalata, Sensor fusion for mobile robot navigation, Proceedings of the IEEE 85 (1997) 108–119.
[4] A. Toet, J.K. IJspeert, A.M. Waxman, M. Aguilar, Perceptual evaluation of different image fusion schemes, Displays 24 (2003) 25–37.
[5] A. Toet, E.M. Franken, Fusion of visible and thermal imagery improves situational awareness, Displays 18 (1997) 85–95.
[6] V.S. Petrovic, Multisensor Pixel-level Image Fusion, Ph.D. Thesis, Manchester School of Engineering, UK, 2001.
[7] S.S. Kumar, S. Muttan, PCA-based image fusion, Proceedings of the SPIE 6233 (2006) 62331T.
[8] J. Saeedi, K. Faez, S. Mozaffari, Multi-focus image fusion based on fuzzy and wavelet transform, in: Iberoamerican Congress on Pattern Recognition CIARP, 2009, pp. 970–977.
[9] D. Mueller, A. Maeder, P. O'Shea, The Generalised Image Fusion Toolkit (GIFT), 2006.
[10] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, IEEE Transactions on Communications 31 (4) (1983) 532–540.
[11] P.J. Burt, A gradient pyramid basis for pattern selective image fusion, in: Proc. of the Society for Information Display Conf., 1992, pp. 467–470.
[12] A. Toet, A morphological pyramidal image decomposition, Pattern Recognition Letter 9 (3) (1989) 255–261.
[13] G. Pajares, J.M. Cruz, A wavelet-based image fusion tutorial, Pattern Recognition 37 (9) (2004) 1855–1872.
[14] R.S. Blum, Z. Liu, Multi-sensor Image Fusion and Its Applications, Taylor & Francis Group, New York, 2006.
[15] P.J. Burt, R.J. Kolczynski, Enhanced image capture through fusion, in: Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany, 1993, pp. 173–182.
[16] H. Li, B.S. Manjunath, S.K. Mitra, Multisensor image fusion using the wavelet transform, Graphical Models and Image Processing 57 (3) (1995) 235–245.
[17] I. Koren, A. Laine, F. Taylor, Image fusion using steerable dyadic wavelet transforms, in: Proceedings of the IEEE International Conference on Image Processing, Washington, DC, 1995, pp. 232–235.
[18] J. Lewis, R. O'Callaghan, S. Nikolov, D. Bull, N. Canagarajah, Pixel- and region-based image fusion with complex wavelets, Information Fusion 8 (2) (2007) 119–130.
[19] G. Piella, H. Heijmans, Multiresolution image fusion guided by a multimodal segmentation, in: Proceedings of Advanced Concepts of Intelligent Systems, Ghent, Belgium, 2002, pp. 1–8.
[20] N. Cvejic, D. Bull, N. Canagarajah, Region-based multimodal image fusion using ICA bases, IEEE Sensors Journal 7 (5) (2007) 743–775.
[21] T. Wan, N. Canagarajah, A. Achim, Segmentation-driven image fusion based on alpha-stable modeling of wavelet coefficients, IEEE Transactions on Multimedia 11 (4) (2009) 624–633.
[22] N.G. Kingsbury, The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters, in: Proc. 8th IEEE DSP Workshop, Utah, August, 1998, pp. 9–12.
[23] I.W. Selesnick, R.G. Baraniuk, N.G. Kingsbury, The dual-tree complex wavelet transform, IEEE Signal Processing Magazine (2005) 124–152.
[24] L. Vincent, P. Soille, Watersheds in digital spaces: an efficient algorithm based on immersion simulations, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (1991) 583–593.
[25] P. Scheunders, J. Sijbers, Multiscale watershed segmentation of multivalued images, IEEE Proceedings (2002).
[26] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Tom Robbins, Upper Saddle River, NJ, 2002.
[27] N. Kingsbury, A dual-tree complex wavelet transform with improved orthogonality and symmetry properties, in: International Conference on Image Processing, 2000, pp. 375–378.
[28] N. Kingsbury, Design of q-shift complex wavelets for image processing using frequency domain energy minimization, in: International Conference on Image Processing, 2003, pp. 1013–1016.
[29] V. Petrovic, C. Xydeas, On the effects of sensor noise in pixel-level image fusion performance, in: Proceedings of the Third International Conference on Image Fusion, vol. 2, 2000, pp. 14–19.
[30] V.S. Petrovic, C.S. Xydeas, Gradient-based multiresolution image fusion, IEEE Transactions on Image Processing 13 (2) (2004) 228–237.
[31] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.
[32] S. Rao, Optimization: Theory and Applications, Wiley, New York, 1984.
[33] J. Kennedy, R.C. Eberhart, A discrete binary version of the particle swarm algorithm, in: Proceedings of the 1997 IEEE Conference on Systems, Man, and Cybernetics, Piscataway, NJ, 1997, pp. 4104–4109.
[34] E. Goldberg, Genetic Algorithms in Search, Optimization & Machine Learning, Addison-Wesley, 1989.
[35] Y. Shi, R.C. Eberhart, Parameter selection in particle swarm optimization, in: Proceedings of the Seventh Annual Conference on Evolutionary Programming, 1998, pp. 591–600.
[36] Y. Shi, R.C. Eberhart, Empirical study of particle swarm optimization, in: Proceedings of the IEEE Congress on Evolutionary Computation (CEC), 1999, pp. 1945–1950.
[37] J. Kennedy, The behavior of particles, in: V.W. Porto, N. Saravanan, D. Waagen, A.E. Eiben (Eds.), Evolutionary Programming VII, Springer, 1998, pp. 581–590.
[38] A. Carlisle, G. Dozier, An off-the-shelf PSO, in: Proceedings of the Particle Swarm Optimization Workshop, 2001, pp. 1–6.
[39] O. Rockinger, T. Fechner, Pixel-level image fusion: the case of image sequences, Proceedings of the SPIE 3374 (1998) 378–388.
[40] Available at: http://imagefusion.org.

Jamal Saeedi was born in Amol, Iran, in 1983. He received the B.Sc. degree in Biomedical Engineering from Sahand University of Tabriz, Iran in 2006 and M.Sc. degree in Electronic Engineering from Amirkabir University of Tehran, Iran in 2009. He is currently a Ph.D. student in the Electrical Engineering department of Amirkabir University of Technology, Tehran, Iran. He works in the field of signal and image processing, specializing particularly in multi-resolution analysis, remote sensing, image fusion, speech enhancement, biomedical signal compression, and fuzzy image processing.

Karim Faez was born in Semnan, Iran. He received his B.S. degree in Electrical Engineering from Tehran Polytechnic University as the first rank in June 1973, and his M.S. and Ph.D. degrees in Computer Science from University of California at Los Angeles (UCLA) in 1977 and 1980, respectively. Prof. Faez was with Iran Telecommunication Research Center (1981–1983) before joining Amirkabir University of Technology (Tehran Polytechnic) in Iran, where he is now a professor of Electrical Engineering. He was the founder of the Computer Engineering Department of Amirkabir University in 1989 and he has served as the first chairman during April 1989–September 1992. Professor Faez was the chairman of planning committee for Computer Engineering and Computer Science of Ministry of Science, Research and Technology (during 1988–1996). His research interests are in Pattern Recognition, Image Processing, Neural Networks, Signal Processing, Farsi Handwritten Processing, Earthquake Signal Processing, Fault Tolerant System Design, Computer Networks, and Hardware Design. He is a member of IEEE, IEICE, and ACM.