Pattern Anal Applic
DOI 10.1007/s10044-011-0235-9

THEORETICAL ADVANCES

A classification and fuzzy-based approach for digital multi-focus image fusion

Jamal Saeedi · Karim Faez

Received: 20 September 2010 / Accepted: 15 July 2011

Abstract This paper presents a new wavelet-based method for fusion of spatially registered multi-focus images. We have formulated the image fusion process as a two-class classification problem: in and out-of-focus classes. First, a 12–dimensional feature vector using dual-tree discrete wavelet transform (DT-DWT) sub-bands of the source images are extracted, and then a trained two-class fisher classifier projects it to the class labels. The classifier output is used as a decision map for fusing high-frequency wavelet coefficients of multi-focus source images in different directions and decomposition levels of the DT-DWT. In addition, there is an uncertainty for selecting high-frequency wavelet coefficients in smooth regions of source images, which causes some misclassified pixels in the classification output or the decision map. In order to solve this uncertainty and integrate as much information as possible from the source images into the fused image, we propose an algorithm based on fuzzy logic, which combines outputs of two different fusion rules based on a dissimilarity measure from the source images: Selection based on the decision map and weighted averaging. An estimation of the decision map is also used for fusing low-frequency wavelet coefficients of the source images 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 (more than 4.5 dB on average) as compared to previous fusion methods.

Keywords Multi-focus image fusion · Dual-tree discrete wavelet transform · Fisher classifier · Fuzzy logic

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 that represent the visual information obtained from a number of different imaging sensors, in a single fused image without causing distortion or loss of information. Image fusion gives a way to combine multiple images into a composite image, which is more suitable for the purposes of human visual perception and computer processing tasks such as segmentation, feature extraction, and target recognition.

Image fusion has many important applications such as digital auto-focusing [1], microscopic imaging, remote sensing [2], and medical imaging [3]. In this paper, we have concentrated on multi-focus image fusion. Due to the limited depth-of-focus of optical lenses in CCD devices, it is often impossible to take an image that contains all objects "in focus". The depth-of-focus is the range of distance from a camera that is reasonably sharp in the image obtained by that camera. One method to conquer this problem is to take several pictures with different focus points and combine them into a single composite, which contains the focused regions of all input images. This could be useful, for example, in digital camera design or in industrial inspection applications, where the need to visualize objects at very short distances makes difficult the preservation of the depth-of-focus [4, 5].

J. Saeedi · K. Faez (✉)
Electrical Engineering Department,
Amirkabir University of Technology
(Tehran Polytechnic), 424, Hafez Ave., Tehran, Iran
e-mail: kfaez@aut.ac.ir

J. Saeedi
e-mail: jamal.saeedi@yahoo.com

Published online: 19 August 2011

There are various methods for fusing multi-focus images. We have grouped them into three major categories; however, these categories may cover each other in various ways: (1) spatial domain methods, (2) optimization-based methods, and (3) multi-scale decomposition-based methods. In this study, we have focused on the wavelet-based approach, which is a subset of multi-scale decomposition-based methods. There are many reasons for selecting wavelet transform as a tool for image fusion, e.g., low computational complexity and the ability of the wavelet transform to capture important features in a picture. The most important reason is the time–frequency analysis of wavelet transform. Since in focus pixels of an image contain high-frequency information, we can simply use this ability of wavelet transform to determine in focus pixels.

In recent years, many researchers have recognized that multiscale transforms (MST) are very helpful for analyzing the information content of images for the purpose of image fusion [6–8]. Multi-scale decomposition-based (MSD) image fusion is a biologically motivated method, which fuses multiple images at different spatial resolutions. MSD-based image fusion includes three steps [9]:

- 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.
- Following this decomposition, the coefficients or samples in each band of the source images are fused based on some fusion rules.
- 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 are used multi-resolution transformations consisting of the Laplacian pyramid [7], gradient pyramid [10], morphological pyramid [11], and wavelet transform [12]. Research results disclose 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]. In recent years, several fusion schemes based on improved wavelet transforms have been developed [14–16]. Hill et al. [17] applied the shift invariant and directionally selective dual-tree complex wavelet transform (DT-CWT) to image fusion. The DT-CWT is an over-complete wavelet, which provides both good shift invariance and directional selectivity over the DWT.

In this paper, we propose a new multi-focus image fusion method, which is also based on multi-scale decomposition. Specifically, we present new fusion rules to merge high- and low-frequency wavelet coefficients in an appropriate way in order to obtain the best quality in the fused image. Since, the key step in the MSD-based image fusion is the coefficient combination step or the fusion rules, we have only focused on this issue. For this purpose, we have studied different fusion rules and their disadvantages. The three previously important fusion rules, are as follows [17]: The Maximum Selection (MS), which just picks the coefficients in each sub-band with the largest magnitude; The Weighted Average (WA), which is proposed by Burt and Kolczynski [18] and used a normalized correlation between the two image sub-bands over a small local area. The resulting coefficients for reconstruction are calculated from this measure via a WA of the two images coefficients; The Window-Based Verification (WBV), which is developed by Li et al. [12] and creates a binary decision map to choose between each pair of coefficients using a majority filter.

These fusion rules ignore some useful information and are sensitive to noise. Selective operation renders the fused coefficients completely dependent on the coefficients with larger average of local area energy and ignores the other corresponding coefficients. In the WA scheme, the weights were computed by a linear function, which cannot describe the uncertainty of each source image contributions. In addition, in coarser level of decomposition, because of passing through low-frequency filter banks, source images become smoother and, therefore, there are not enough differences between their wavelet coefficients for selecting one of them as in focus coefficients.

In this study, we aim to reach best performance that a wavelet-based method can produce. The best performance is related to manually picking in focus wavelet coefficients from the transformed source images using hand-made decision map to generate best possible fused image in the wavelet domain. Therefore, we present a new method for merging high- and low-frequency wavelet coefficients as a fusion rule using Fisher classifier and fuzzy logic. We use the inter-scale dependency between wavelet coefficients in the DT-DWT sub-bands to obtain a reliable decision map. First, a 12-dimensional feature vector is obtained using six directional sub-bands of DT-DWT in the first decomposition level of the source images. Then, this feature vector is projected to one-dimensional space using a trained Fisher classifier. We use the classifier output as a decision map for selecting high-frequency wavelet coefficients between two source images in the different directions and decomposition levels, equally. In addition, there is an uncertainty about selecting high-frequency wavelet coefficients in the smooth regions of two images, which causes some misclassified regions in the classifier output. In order to solve this uncertainty and integrate as much information of each source image as possible into the fused image, we propose an algorithm based on fuzzy logic, which combines

the outputs of two different fusion rules: Selection based on decision map, and weighted averaging. The classifier output is also used for selecting low-frequency wavelet coefficients between the source images instead of simple averaging.

For testing our new fusion rule, we use 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.

The paper is organized as follows: In Sect. 2, the proposed fusion algorithm using the Fisher classifier and fuzzy logic is presented. Section 3 gives various results and comparisons. Finally, we conclude with a brief summary in Sect. 4.

2 The proposed image fusion algorithm

In this Section, we present the proposed multi-focus image fusion algorithm. Figure 1 shows the block diagram of the proposed method, which consists of some essential stages:

1. The source images are decomposed into different directions and scales using DT-DWT.
2. A 12-dimensional feature vector is extracted using the directional sub-bands of DT-DWT in the first decomposition level.
3. A reliable decision map is obtained using a trained Fisher classifier.
4. The high-frequency wavelet coefficients of the source images are integrated using fuzzy logic, which combines outputs of two different fusion rules (selection based on decision map, and weighted averaging), based on a dissimilarity measure of the source images.
5. The low-frequency wavelet coefficients of the final fused image are selected from the source images using a down sampled version of the decision map, which is obtained in step 3.
6. The inverse DT-DWT of the new low- and high-frequency wavelet coefficients generates the final fused image.

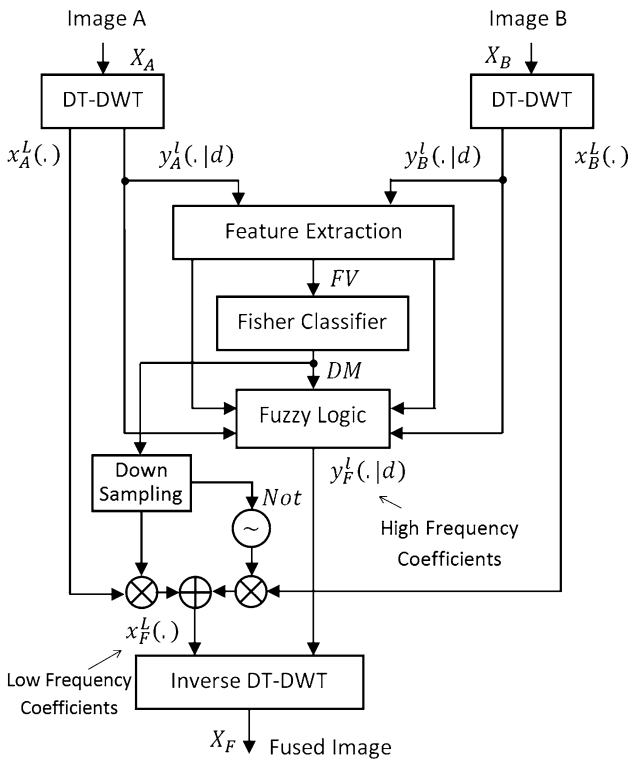


Fig. 1 Block diagram of proposed image fusion algorithm

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

2.1 Dual-tree discrete wavelet transform

The DT-DWT is a modified version of the DWT and was proposed to overcome shift variance and directionality limitations of the DWT while maintaining the perfect reconstruction property with limited redundancy [19–21]. 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. This combination led to a complex extension of real signals. As complex wavelets can distinguish between positive and negative frequencies, the diagonal sub-bands can be discriminated from horizontal and vertical sub-bands. Subsequently, horizontal and vertical sub-bands are decomposed 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:

$$y = \{y^1, y^2 \ldots y^L, x^L\}. \tag{1}$$

Here, x^L represents the approximations 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 12 directional sub-bands, six of which are real, and six imaginary:

$$y^l = \begin{cases} \text{real} & y^l_{\text{real}}(i,j|1), y^l_{\text{real}}(i,j|2), \ldots, y^l_{\text{real}}(i,j|6) \\ \text{imaginary} & y^l_{\text{imag}}(i,j|1), y^l_{\text{imag}}(i,j|2), \ldots, y^l_{\text{imag}}(i,j|6) \end{cases}. \tag{2}$$

We use the coordinates (i, j) or the shorthand notation of $(.)$, to index the spatial position of the coefficients.

2.2 Feature extraction

The high-frequency wavelet coefficients reflect the image edges and detailed information. Based on the mechanism of imaging in the optical system, the bandwidth of system function for in focus images is wider for out-of-focus images [22]. Therefore, the absolute values of high-frequency wavelet coefficients of clear images are larger than blurred images.

Using the information of high-frequency wavelet coefficients is a straightforward approach to distinguish between in and out-of-focus regions; however, we can use some local features (such as energy, mean, standard deviation, and normalized Shannon entropy), to enhance the information of wavelet coefficients. Commonly, absolute values of wavelet coefficients are used for calculating local features as follows:

$$\text{energy}(i,j) = \frac{1}{W}\sum_{n=-N}^{N}\sum_{m=-M}^{M}|y(i+n,j+m)|^2 \tag{3}$$

$$\text{mean}(i,j) = \frac{1}{W}\sum_{n=-N}^{N}\sum_{m=-M}^{M}|y(i+n,j+m)| \tag{4}$$

$$\text{std}(i,j) = \left[\frac{1}{W}\sum_{n=-N}^{N}\sum_{m=-M}^{M}\left(|y(i+n,j+m)| - |\bar{y}|\right)^2\right]^{1/2} \tag{5}$$

$$\text{entropy}(i,j) = -\sum_{l=0}^{L} h(l) \times \text{Log}_2(h(l)) \tag{6}$$

where $y(i, j)$'s are the wavelet coefficients at the position of (i, j), $(2N + 1, 2M + 1)$ is the size of local window, W is the number of coefficients in the local window, h is normalized histogram of absolute values of wavelet coefficients in a local window at the position of (i, j), and L is the number of levels.

Here, we use the magnitude of complex wavelet coefficients at the first decomposition level for calculating the local features:

$$y^1_{\text{abs}}(.|d) = \sqrt{y^1_{\text{real}}(.|d)^2 + y^1_{\text{imag}}(.|d)^2} \tag{7}$$

where $y^1_{\text{real}}(.|d)$ and $y^1_{\text{imag}}(.|d)$ are the real and imaginary wavelet coefficients of the DT-DWT at the first level and orientation d, which can be $\pm 15°$, $\pm 45°$, and $\pm 75°$.

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

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

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

We use two different local windows (3×3, and 7×7) to calculate the local features in the six sub-bands of DT-DWT of source images. Calculating the features using two different local windows can reduce the existent uncertainty in the border of in and out-of-focus regions. Therefore, 12 feature images are obtained for each source image using two local windows in the six directional sub-bands of DT-DWT. Then, the differences between extracted features of the source images are used as a feature vector for the classification problem:

$$FV^d_{\text{LW}} = NF^d_{\text{LW}}(A) - NF^d_{\text{LW}}(B) \tag{9}$$

where A and B are the source images, NF is obtained using (8), d's are the six high-frequency directional sub-bands of DT-DWT, and LW are the two local windows.

Finally, the 12-dimensional features vector is defined as follows:

$$FV = (FV^1_{3\times3}, FV^2_{3\times3}\ldots FV^6_{3\times3}, FV^1_{7\times7}, FV^2_{7\times7}\ldots FV^6_{7\times7}). \tag{10}$$

Thus, at the end of feature extraction step we have a stack of 12 transformed images, which is obtained using six high-frequency directional sub-bands of DT-DWT of two source images with different focus points.

2.3 Fisher classifier

Having the feature vector, for classification of wavelet coefficients as either in focus or out-of-focus, we use the Fisher classifier. Compared with the Neural Network (NN) and the Support Vector Machine (SVM), the Fisher classifier is easier to train, faster to classify, needs fewer training samples, and does not suffer from overtraining problems [23, 24].

For a feature vector X, the Fisher classifier projects X into one dimension Y in direction W using:

$$Y = W^T X. \tag{11}$$

The Fisher criterion finds the optimal projection direction W_0 by maximizing the ratio of the between-class scatter to the within-class scatter. Let S_w and S_b be the within- and between-class scatter matrices, respectively,

$$S_w = \sum_{k=1}^{K}\sum_{x\in\text{class}K}(x-u_k)(x-u_k)^T \tag{12}$$

$$S_b = \sum_{k=1}^{K}(u_k-u_0)(u_k-u_0)^T \tag{13}$$

$$u_0 = \sum_{k=1}^{K} u_k \tag{14}$$

where u_k is the mean vector of the kth class, u_0 is the global mean vector, and K is the number of classes.

The optimal projection direction is the eigenvector of $S_w^{-1}S_b$ corresponding to its largest eigenvalue [25]. For a two-class classification problem, we do not need to calculate the eigenvectors of $S_w^{-1}S_b$. It is shown that the optimal projection direction is [25]:

$$W_0 = S_w^{-1}(u_1 - u_2). \tag{15}$$

Figure 2 shows train images and their corresponding class labels for training Fisher classifier (black regions indicate out-of-focus area and white regions indicate in focus area for the first source image and vice versa). Let Y_1 and Y_2 be the projections of training images for in and out-of-focus classes to the optimal projection direction W_0, and let $E[Y_1]$ and $E[Y_2]$ be the means of Y_1 and Y_2, respectively. Suppose $E[Y_1] > E[Y_2]$, then for a pair of test images with feature vector X, the decision can be made as:

$$Y = W_0^T X. \tag{16}$$

Then,

$$C(X) = \begin{cases} \text{class1} & \text{if } Y > \frac{E[Y_1]+E[Y_2]}{2} \\ \text{class2} & \text{otherwise} \end{cases}. \tag{17}$$

It is known that, if the feature vector X is jointly Gaussian distributed and the two classes have the same covariance matrices, then the Fisher classifier is optimal in a minimum classification error sense [25].

Table 1 shows classification accuracy using different local features (Mean, Energy, Normalized Shannon Entropy, and Standard Deviation) for test images. It should be mentioned that for calculating the classification accuracy, the first handmade classification maps are extracted from different multi-focus images using cut and paste method. Then, these handmade classification maps are compared with the classification outputs for different test images. As it can be seen in Table 1, the best result is related to the standard deviation as the local feature for the classification problem. Also using two local features, while increasing the complexity, is not effective for the classification accuracy. The handmade classification maps and the classification results using the standard deviation as the local feature, for a number of multi-focus images are shown in Fig. 3.

After completion of classification, there may existed some misclassified pixels. In order to solve this problem, we have used post-processing. Li [12] applied a majority filter for this purpose. In the Li method, if the number of 1's pixels are more than 0's pixels in a local window (e.g. 5×5) around a central pixel in the binary decision map, then the value of central pixel is set to 1, and if the number of 0's pixels are more than 1's pixels, then the value of central pixel is set to 0.

In this paper, we have proposed a new method based on area of separate regions in the binary classification map. First, each separate region in the binary classification map is labeled (using the "bwlabel" function in the Matlab software). Then, the area of each labeled region is calculated (using the "regionprops" function). Finally, the binary value of each labeled region in the decision map, for which its area is smaller than a specific threshold, is complemented (i.e., from one to zero and vice versa). In order to determine the threshold value, the calculated areas

Fig. 2 **a** Training images used in the experiment, **b** class labels (*black regions* indicate out-of-focus areas and *white regions* indicate in focus areas for the first training image and for the second training image, *black regions* indicate in of focus areas and *white regions* indicate out-of-focus areas)

Table 1 Classification accuracy using different local features

Feature	Mean	Energy	Std.	Entropy	Mean + Std.	Energy + Std.
Classification Accuracy	88.11%	81.83%	**88.53%**	85.67%	87.50%	86.49%

Bold value indicate better performance

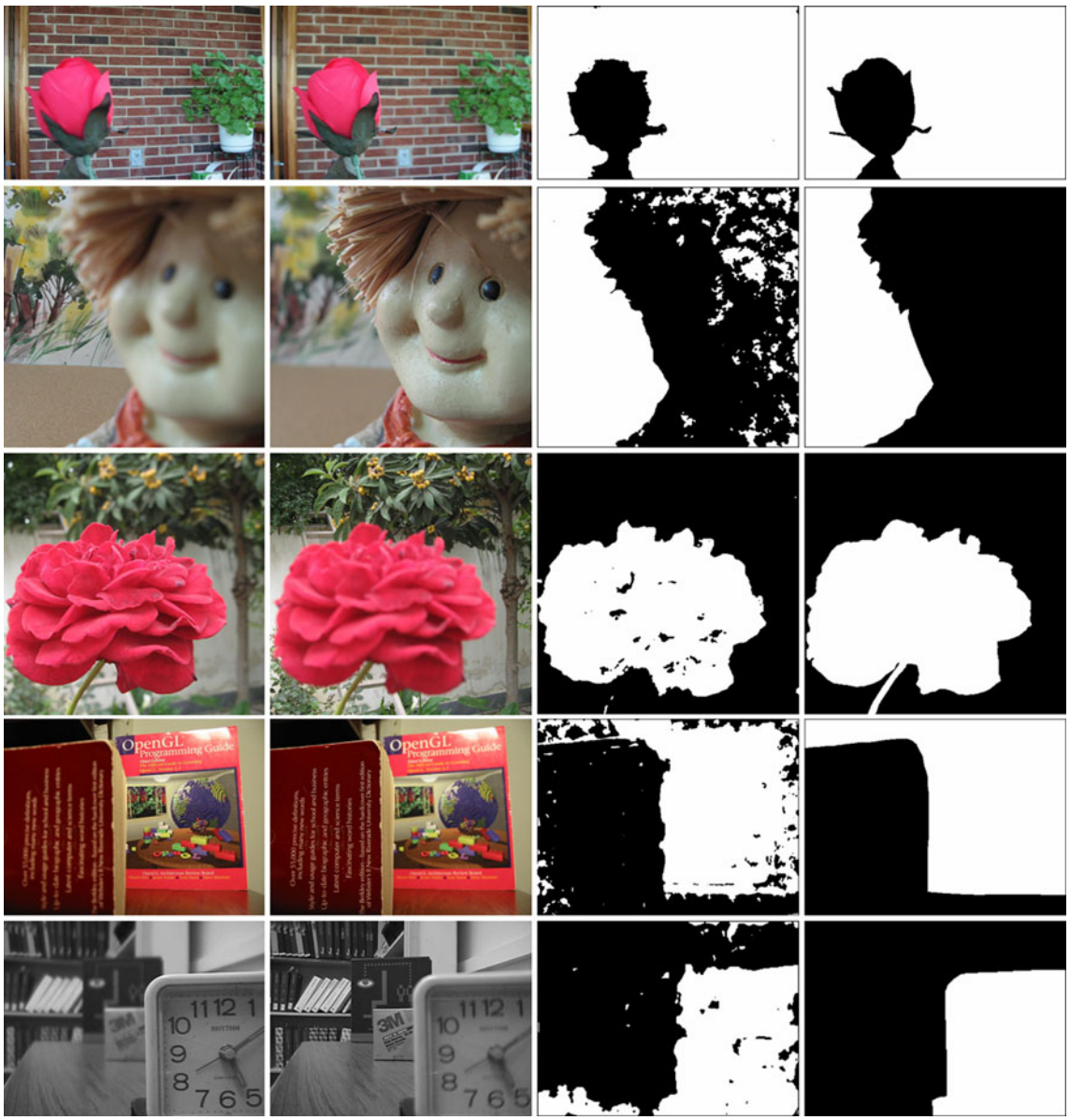


Fig. 3 Classification results for the different multi-focus test images. *First and second columns from top to bottom* pairs of multi-focus "Flower", "Doll", "Rose", "Book", and "Disk" images, *third and fourth columns* the classification result, and the handmade classification maps, respectively

of the isolated regions in the decision map are sorted. Then, the second value from the top is selected as the threshold. It is assumed that there are two dominant regions in the classification map (i.e. in and out-of-focus regions), and the isolated regions with smaller areas are misclassified. Figure 4 shows the results of post-processing using majority filter and the proposed method in this paper. In addition, Table 2 shows the classification accuracy after post-processing using these methods. As it can be observed in Table 2, the improvement using our post-processing method is promising (more than 6.5%).

2.4 Fusion rule for high-frequency coefficients

The classifier output obtained in the previous subsection is used as a decision map (DM) for selecting high-frequency wavelet coefficients between the multi-focus source images in different directions and decomposition levels of the DT-DWT. In fact, the fusion rule is defined as follows:

$$y_1^l(.|d) = y_A^l(.|d) \times \mathrm{DM}(.) + y_B^l(.|d) \times (1 - \mathrm{DM}(.)) \quad (18)$$

where $y_A^l(.|d)$ and $y_B^l(.|d)$ are the high-frequency wavelet coefficients of the source images A and B, at level l and

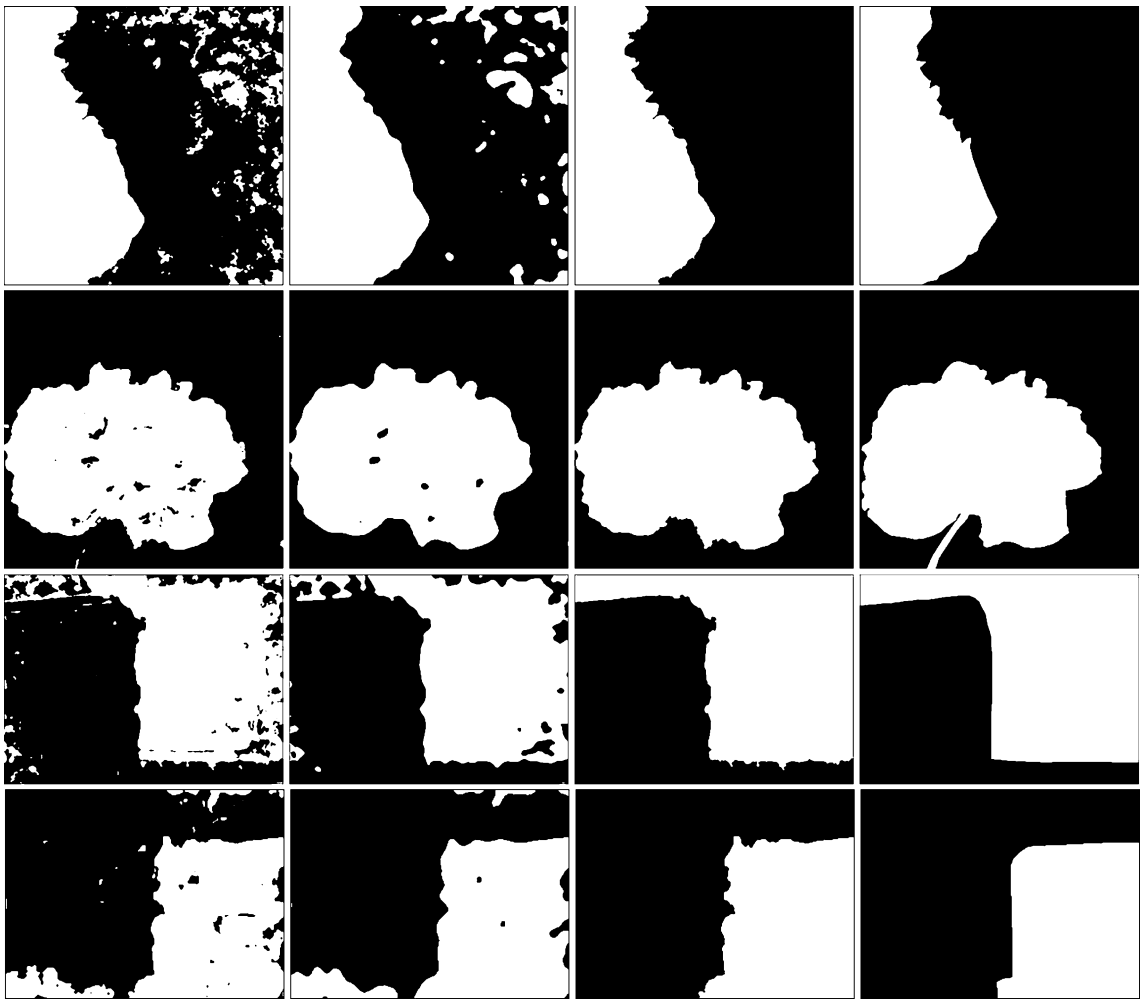


Fig. 4 *First column* the classification results for multi-focus "Doll", "Rose", "Book", and "Disk" images, *second column* post-processing results using majority filter, *third column* the post-processing results using proposed method and *fourth column* handmade classification map

Table 2 Classification accuracy after post-processing

Method	Before post-processing	Majority filter [12]	Proposed
Classification accuracy	88.53%	90.15%	**95.07%**

Bold value indicate better performance

orientation d, $l = 1, 2\ldots\mathrm{L}$, and $d = 1, 2\ldots6$ which can be $\pm15°$, $\pm45°$, and $\pm75°$. Also in the coarser decomposition level, an estimation of DM using down sampling is used.

Although the decision maps obtained using the Fisher classifier are very accurate, it does not result in an optimal fused image. Indeed, there are many smooth regions in the source images, and the extracted features could not distinguish them enough to classify them as in or out-of-focus pixels. Therefore, some pixels are misclassified in the decision map. In order to solve this uncertainty problem, a weighted averaging rule is used for these regions:

$$y_2^l(.|d) = \frac{y_A^l(.|d) \times W_A(.) + y_B^l(.|d) \times W_B(.)}{W_A(.) + W_B(.)} \tag{19}$$

where $W_A(.)$ and $W_B(.)$ are obtained using the following formula:

$$W_S(i,j) = \frac{1}{12} \sum_{\mathrm{LW}=3\times3,7\times7} \sum_{d=1}^{6} NF_{\mathrm{LWS}}^d(i,j) \tag{20}$$

where NF is the extracted feature for each source image using (8).

We plan to design a good fusion method for combining two fusion rules using (18) and (19) to integrate as much information as possible into the fused image. We define a dissimilarity measure (DIS) for this purpose. The DIS is intended to measure the degree of 'dissimilarity' from the

source images. In the following expression, this measure is defined as:

$$AV(i,j) = \frac{1}{12} \sum_{\mathrm{LW}=3\times3,7\times7} \sum_{d=1}^{6} \left| NF^{d}_{\mathrm{LW}}(A) - NF^{d}_{\mathrm{LW}}(B) \right|. \tag{21}$$

Then,

$$\mathrm{DIS}(i,j) = \left[\sin\left(\frac{\pi \times AV(i,j)}{2 \times T} \right) \right]^{1/2} \tag{22}$$

where $T = \max_{i,j}(AV(i,j))$.

By analyzing the DIS measure, we can determine where and to what extent the source images differ, and use this information to combine the fusion rules. First, we define the following linguistic rules for the proposed fusion rule:

IF the DIS measure at a given position is high (i.e., the sources are different at that position) **THEN** we use the first fusion rule (selection using DM)
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 and large membership functions. Fuzzy membership function is a curve, which 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 trapezoidal membership function, which is shown in Fig. 5. Finally, the fusion rule for high-frequency coefficients is defined as follows:

$$y^{l}_{F}(.|d) = \sum_{i=1}^{2} y^{l}_{i}(.|d) \times \mu_i(\mathrm{DIS}(.)) \tag{23}$$

where y^{l}_{1} and y^{l}_{2} are obtained using (18) and (19), and μ_1 and μ_2 are the large and small membership functions, respectively.

The proposed fusion rule for the high-frequency coefficients causes blurring effect around the border area between the in and out-of-focus regions in the fused image. Although the decision maps obtained using the Fisher classifier are very accurate, but the border points between the in and out-of-focus regions are not extracted, appropriately. In order to solve this problem, we change the first fusion rule expressed in (18). First, the border of in and out-of-focus regions is obtained using the decision map. This border is then widened (using the "imdilate" function of Matlab software). Figure 6 shows the extracted border maps (BM) using the decision maps of different multi-focus images. Finally, the BM is used for solving the blurring effect around the border regions as:

$$y^{l}_{3}(.|d) = \begin{cases} y^{l}_{A}(.|d) & \text{if } F_{3\times3}\left(\left|y^{l}_{A}(.|d)\right|\right) > F_{3\times3}\left(\left|y^{l}_{B}(.|d)\right|\right) \\ y^{l}_{B}(.|d) & \text{otherwise} \end{cases} \tag{24}$$

where $F_{3\times3}$ is the standard deviation using (5) in a 3×3 local window. Then, the first fusion rule is changed as follows:

$$\tilde{y}^{l}_{1}(i,j|d) = \begin{cases} y^{l}_{3}(i,j|d) & \text{if } \mathrm{BM}(i,j) = 0 \\ y^{l}_{1}(i,j|d) & \text{otherwise} \end{cases}. \tag{25}$$

Figure 7 shows the subjective results using this correction for the different multi-focus images. As it can be seen in Fig. 7, the border regions are appropriately retrieved.

It should be mentioned that we use an estimation of the DM, DIS and BM using down sampling for coarser decomposition levels for the fusion rule. In fact, the proposed method is based on multi-scale grouping, which uses a similar approach for fusion of high-frequency wavelet coefficients in different directions and scales of dual-tree discrete wavelet transform.

Usually, using more decomposition levels for wavelet-based image fusion results in a better fused image. In the previous wavelet-based methods, the decision map is obtained by comparing the activity level measurement of wavelet coefficients between source images in different

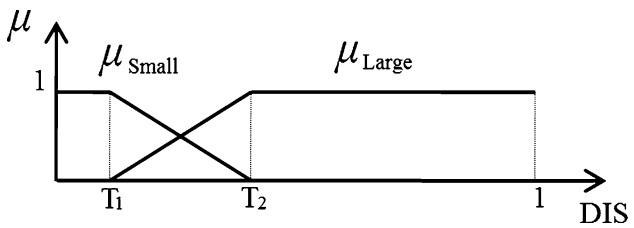


Fig. 5 Fuzzy membership function, $T_1 = 0.1$ and $T_2 = 0.25$, which are obtained using test images by trial and error

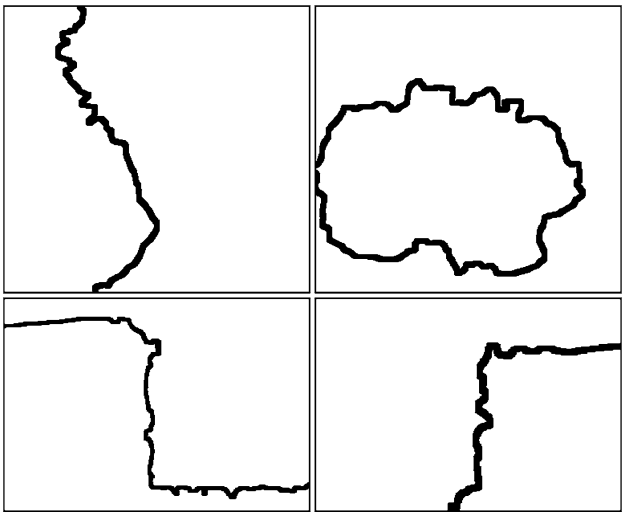


Fig. 6 Border map for multi-focus "Doll", "Rose", "Disk", and "Book" images (*from top left, clockwise*)

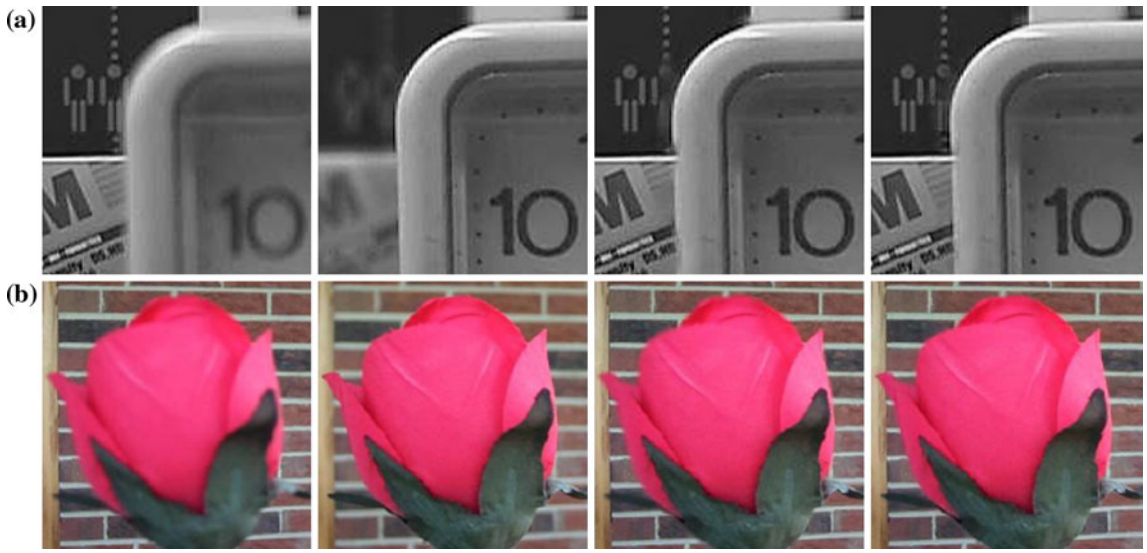


Fig. 7 Subjective fusion results, **a** *from left to right* a part of Multi-focus "Disk" images, results without and with border correction, and **b** *from left to right* a part of Multi-focus "Flower" images, results without and with border correction

directions and scales. In addition to lower accuracy of the decision maps in finer scales, also in coarser scales due to loss of high-frequency information of images (because of passing through low-frequency filter banks), the decision maps were wrongly extracted for selecting more important coefficients. Therefore, a certain levels numbers as the best level in the wavelet analysis was selected. Even in some literature, algorithms for determining the optimal decomposition level have been suggested [26].

In the proposed method, because of using an accurate and reliable decision map in different wavelet decomposition levels, the problem of selecting appropriate levels is solved. In fact, the experimental results show that for N-2 decomposition levels, the best results are obtained, where N is the possible decomposition levels for an image (e.g., for a 512 × 512 image, $N = \log_2 512 = 8$).

Table 3 shows a comparison between previous fusion rules for high-frequency wavelet coefficients consisting of MS, WA [18], and WBV [12], and the proposed method in this paper. In this experiment, which has been performed on different multi-focus images, the best value of the PSNR index with respect to the decomposition levels for the fused images (compared to the handmade or reference image) are obtained for different methods. Handmade or reference image is obtained by transferring in focus regions of source images into one image. As it can be observed in Table 3, the best results for the proposed method are obtained using N-2 decomposition levels, while for other fusion rules the best results are obtained in different decomposition levels. The purpose of this experiment is determining optimal decomposition levels for image fusion process, and comparisons between different methods will be fully discussed in Sect. 3.

Table 3 The best value of the $PSNR_{dB}$ index with respect to decomposition levels obtained for different multi-focus images and different fusion rules

Test Image	MS	WA [18]	WBV [12]	Proposed	N
"Disk"	35.29	37.41	37.76	42.56	5
	$L = 4$	$L = 5$	$L = 5$	$L = 3$	
"Lab"	35.83	37.64	37.85	41.62	5
	$L = 5$	$L = 5$	$L = 5$	$L = 3$	
"Pepsi"	39.37	40.77	41.57	44.34	8
	$L = 5$	$L = 5$	$L = 5$	$L = 6$	
"Flower"	33.42	34.92	34.80	41.36	8
	$L = 4$	$L = 7$	$L = 7$	$L = 6$	
"Rose"	32.54	33.95	33.77	37.54	8
	$L = 4$	$L = 5$	$L = 5$	$L = 6$	
"Doll"	38.22	38.56	39.14	42.75	8
	$L = 6$	$L = 7$	$L = 7$	$L = 6$	
"Book"	37.97	39.56	39.98	44.87	8
	$L = 3$	$L = 4$	$L = 4$	$L = 6$	

2.5 Fusion rule for low-frequency coefficients

The information contained in the low-frequency (LF) coefficients is very important for the natural appearance of an image, because LF coefficients contain the bulk of the energy of an image. Commonly, the averaging is an accepted method to fuse approximation or low-frequency coefficients of the source images in the wavelet-based fusion methods:

$$x_F^L(.) = \frac{x_A^L(.) + x_B^L(.)}{2} \tag{26}$$

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

However, equally weighing the input approximation or LF sub-bands in the final fused image causes the problem of contrast reduction, because the source images have different contrast conditions. In [27], LF fusion rule is performed using arithmetic combinations of input LF coefficients. Specifically, the offset zero-mean addition is used, which is defined as:

$$x_F^L(.) = x_A^L(.) + x_B^L(.) - \frac{\mu_A + \mu_B}{2} \tag{27}$$

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 (26) is weak to the destructive superposition, especially if the input LF sub-bands have opposing illumination levels.

In this paper, we have proposed a new method for low-frequency fusion rule. Since, there are not enough differences between LF coefficients of the source images to distinguish them as in or out-of-focus coefficients, we have used an estimation of the decision map obtained by the Fisher classifier for LF fusion rule, which is defined as:

$$x_F^L(.) = x_A^L(.) \times \widehat{\mathrm{DM}}(.) + x_B^L(.) \times (1 - \widehat{\mathrm{DM}}(.)) \tag{28}$$

where $\widehat{\mathrm{DM}}(.)$ is obtained via down sampling of the DM, which is the classification output.

Finally, after combining low- and high-frequency dual-tree discrete wavelet coefficients of the source images, the final fused image is obtained using inverse DT-DWT:

$$X_F = \psi^{-1}(y_F, x_F^L) \tag{29}$$

where ψ^{-1} is the inverse DT-DWT, y_F are the high-frequency coefficients using (23), and x_F^L are the low-frequency coefficients using (28).

Table 4 shows the PSNR index for three different alternative low-frequency fusion rules in our fusion scheme. This experiment was performed on the multi-focus images dataset, and average PSNRs for all test images are shown in Table 4. Unlike many authors who believe that the LF fusion rules have little influence on the overall fusion performance [27], it can be seen in Table 4 that the proposed LF fusion rule is very effective (more than +1.46 dB on average).

Table 4 $\mathrm{PSNR}_{\mathrm{dB}}$ index for different low frequency fusion rules

Low-frequency fusion rule	Averaging	Arithmetic combinations [27]	Proposed
$\mathrm{PSNR}_{\mathrm{dB}}$	40.46	40.87	**42.33**

Bold value indicate better performance

3 Experimental results

In this Section, we compare our proposed multi-focus image fusion algorithm with some of the best state-of-the-art fusion methods. The proposed image fusion method was tested against several methods consisting of the simple averaging, principle component analysis (PCA)[1] [28], spatial frequency (SF)[2] [29], and the Dual-Tree Discrete Wavelet Transform[3] with three different fusion rules: MS, WA [18], and WBV [12]. For the DT-DWT based methods, the available "AntonB" mother wavelet is used. In addition, the result of using the handmade decision map in the DT-DWT based image fusion is used as the best result (Best). In other words, the handmade decision map is used for selecting wavelet coefficients between the source images, instead of using the classification output as the decision map in the proposed algorithm.

The evaluation indices of image fusion include subjective and objective indices. Objective indices can be divided into three categories [30]. One category reflects the image features, such as entropy and gradient. The second reflects the relation of the fused image to the source images, such as mutual information and Petrovic index [31]. Finally, the third reflects the relation of the fused image to the standard image, such as correlation coefficient and peak signal to noise ratio (PSNR).

Two metrics are considered in this paper. The first metric is the Xydeas and Petrovic metric, proposed in [31], which considers the amount of edge information transferred from the input images to the fused images 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 PSNR index, which is defined as:

$$\mathrm{PSNR} = 10 \log \frac{255^2}{\mathrm{RMSE}^2} \tag{30}$$

in which the root mean square error (RMSE) is defined as the following formula:

$$\mathrm{RMSE}^2 = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} [R(x,y) - F(x,y)]^2. \tag{31}$$

An image with higher PSNR index will have better fused image. It should be mentioned that for image fusion experiments, a ground-truth image was prepared by cutting and pasting method [12], which is performed by transferring in focus regions of source images into one

[1] Image Fusion Toolbox for MATLAB developed by Oliver Rockinger: http://www.metapix.de/toolbox.htm.

[2] The Image Fusion Toolkit for Matlab developed by Eduardo Canga: http://www.imagefusion.org/.

[3] Available at: http://taco.poly.edu/WaveletSoftware/dt2D.html.

Fig. 8 Multi-focus source images, from top *left, clockwise*: "Pepsi", "Clock", "Jug", and "Lab" images

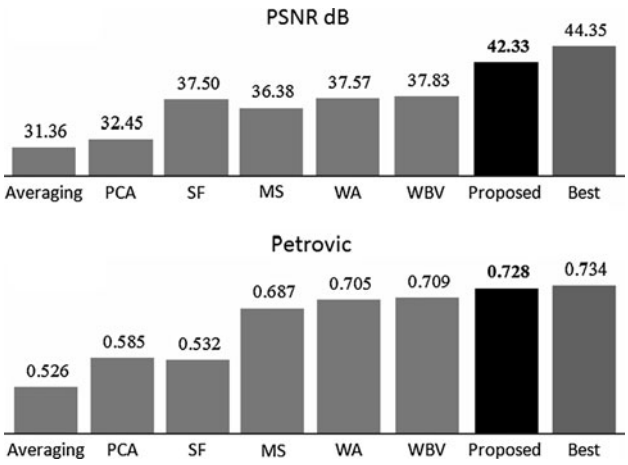


Fig. 9 PSNR and Petrovic indexes for different image fusion methods

image. Image fusion results will be discussed in the two following subsections.

3.1 Fusion results of natural images

In the first experiment, nine pairs of natural multi-focus source images, as shown in Figs. 3 and 8 are used to evaluate the proposed method. The multi-focus images consisting of "Disk", "Lab", "Pepsi" and "Clock" are publicly available at the Image fusion web site [32], "Flower" and "Book" are publicly available at [33], and additional multi-focus images consisting of "Doll", "Rose" and "Jug" are provided by us.

Figure 9 shows objective results for different multi-focus image fusion algorithms. It should be mentioned that the average results between nine pair test images are presented. It can be seen in Fig. 9 that our proposed image fusion algorithm has better PSNR index compared to the other methods (more than +4.5 dB), and the obtained results for the proposed method are very close to the best results (less than 2.2 dB). The best result is related to manually picking in focus wavelet coefficients from the transformed source images using hand-made decision map to generate best possible fused image in the wavelet domain. In addition, evaluation of different methods using Petrovic metric has shown that our proposed algorithm has better performance for transferring edges information from source images to the corresponding fused image. When looking more closely at the results, we observe the following:

- Synthesizing the composite image by averaging the corresponding pixels of the source images is a simple method for image fusion. In addition to simplicity, this method usually causes many undesirable effects on the fused images, such as low contrast. The averaging method produces worst results in terms of PSNR and Petrovic indexes, compared with other methods.
- The PCA method produces fused image using weighted averaging, in which the weights for each input image are obtained from the eigenvector corresponding to the largest eigenvalue [28]. The PCA approach also generates weak objective results compared with other methods.
- Shutao et al. [29] developed a method based on spatial frequency, which is computationally simple and can be used in real-time applications. The spatial frequency measures the overall activity level in an image. The spatial frequency method suffers from the blocking effect, and this problem leads to weak objective results. The consequence of blocking effect on the Petrovic index is more than SNR metric. Because the Petrovic

index measures the similarity between the edges information from source images and the fused image, the blocking effect changes the edges information of fused image compared to the source images.

- As we have discussed in Sect. 2, the previous wavelet-based methods (MS, WA, and WBV) cannot produce reliable decision maps, and, therefore, they result in lower performances compared to the proposed method. Another important reason is that these methods use the averaging method for low-frequency fusion rule. As we have shown in Sect. 2.5, the proposed method for low-frequency fusion rule outperforms the averaging method. The WBV method produces the best results among the pervious wavelet-based methods. Because it uses a WBV to modify the decision maps, which is obtained using the MS method for high-frequency fusion rule.
- The proposed method, which is based on the fuzzy logic and the classification, generates best results compared to other methods. Generation of very accurate decision maps for high- and low-frequency fusion rules is the main reason for outshining the proposed method.

In addition to objective assessment, two samples have been selected here to subjectively evaluate different results. Figure 10 shows the fused results using different methods for multi-focus "Disk" images. As it can be observed in Fig. 10, the PCA method produces blur images with low contrast condition. The result of spatial frequency (SF) method has the blocking effects, which can be seen around the clock. The previous wavelet-based methods including MS, WA, and WBV have caused blurring in the fused images, which can be seen on the left of the clock and top of the book. The result of our proposed method avoids the mentioned problems (blurring, blocking effect), and also produces a clearer fused image with high contrast condition. Figure 11 shows another subjective result for multi-focus "Lab" images. As it can be seen in Fig. 11, specifically in the bottom row (magnified head regions of the fused results), the PCA method produces blur images with low contrast condition. The result of SF method has the blocking effects (around the head). The previous wavelet-based methods including MS, WA, and WBV have caused artifacts in the fused images (around the head). These artifacts are because of using inaccurate decision maps (produced by these methods) for high-frequency fusion rule when multi-focus images have misregistration problem. The result of our proposed method avoids the mentioned problems (blurring, blocking effect, artifacts), and also produces a clearer fused image with high contrast condition.

3.2 Fusion results of artificial images

To confirm the effectiveness of the proposed algorithm, ten standard images shown in Fig. 12 are used as ground truth. From each image, Gaussian blurring creates two out-of-focus images. Then, the blurred images with different focus points are taken as the source images, and the original image is taken as the reference image for image fusion process.

Figure 13 shows objective results for different multi-focus image fusion algorithms. It can be seen in Fig. 13 that our proposed image fusion algorithm has better PSNR

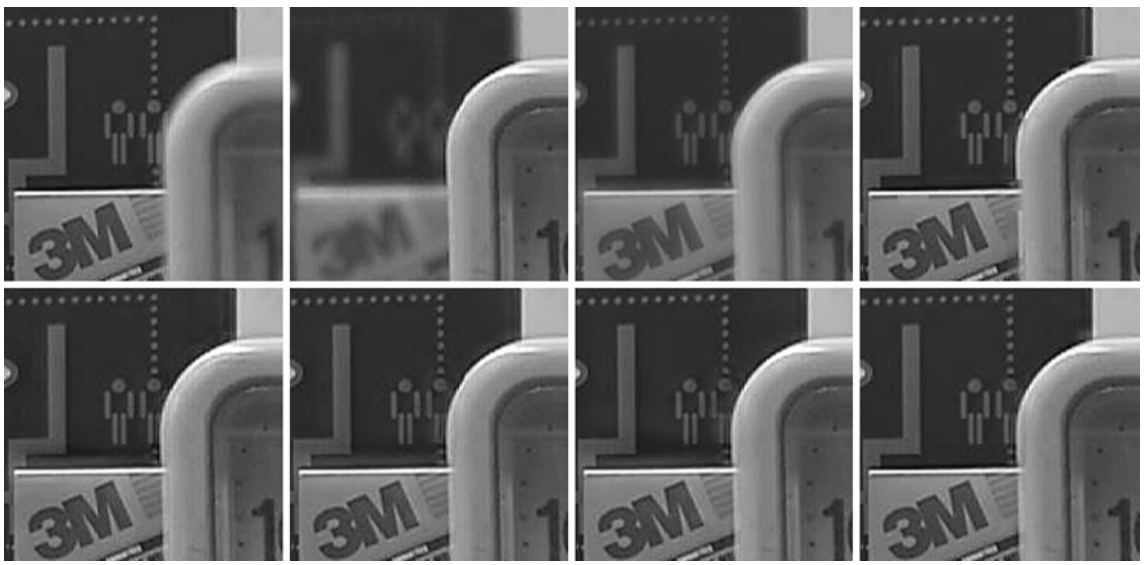


Fig. 10 Subjective fusion results. *Top row, left to right* a part of multi-focus "Disk" images, the fused results using PCA, SF methods. *Bottom row, left to right* the fused results using MS, WA, WBV, and the proposed methods

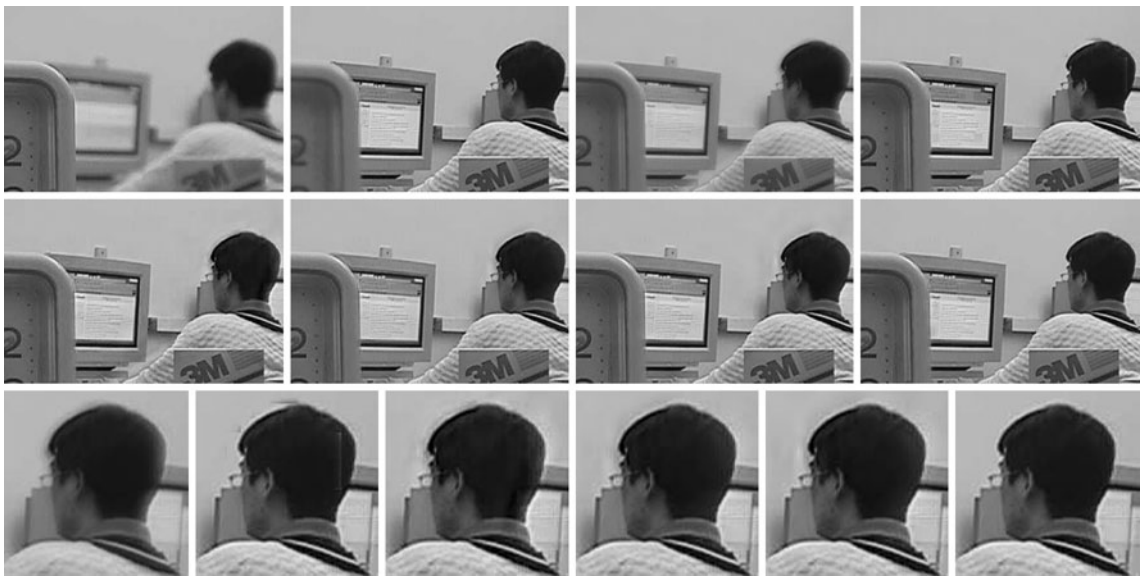


Fig. 11 Subjective fusion results. *Top row, left to right* a part of multi-focus "Lab" images, the fused results using PCA, SF methods. *Middle row, left to right* the fused results using MS, WA, WBV, and the proposed methods. *Bottom row, left to right* magnified head regions of the fused results produced with PCA, SF, MS, WA, WBV, and the proposed method

Fig. 12 Ten ground truth images

index compared to other methods (more than +3.17 dB), and the results obtained for the proposed method are very close to the best results (less than 2.05 dB). In addition, evaluation of different methods using Petrovic metric has shown that our proposed algorithm has better performance for transferring information of edges from source images to the corresponding fused image.

4 Conclusions

In this paper, we have presented a new wavelet-based multi-focus image fusion method using Fisher classifier and Fuzzy logic. 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. Generating a reliable decision map using the information of high-frequency wavelet coefficients in the six directional sub-bands of the DT-DWT is the main idea. In addition, we have used fuzzy logic based on our observation to solve existent uncertainty in the smooth regions of source images and in the border of in- and out-of-focus regions. This new method used DT-DWT for finer frequency decomposition and shift invariant property as compared to discrete wavelet transform. The experimental results demonstrated that the proposed method outperforms the standard fusion methods in the fusion of multi-focus images.

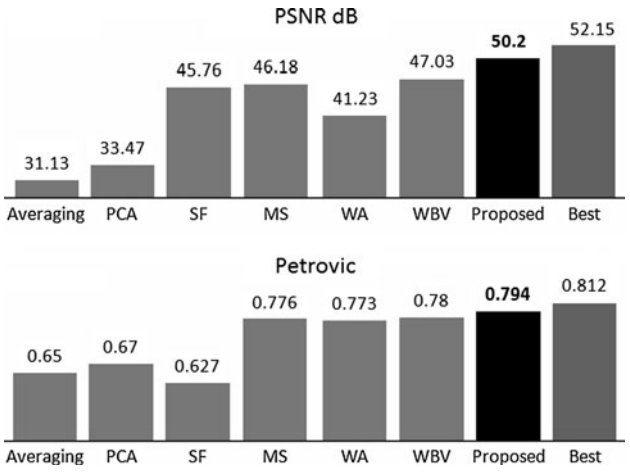


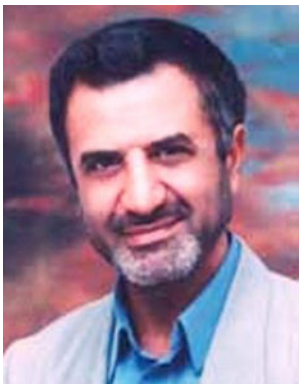
Fig. 13 PSNR and Petrovic indexes for different image fusion methods

References

1. Saeedi J, Faez K (2009) Fisher classifier and Fuzzy logic based multi-focus image fusion. In: IEEE International Conference on Intelligent computing and intelligent systems, pp 420–425
2. Yang XH, Zh L, Jing G, Liu L, Hua ZH (2007) Fusion of multi-spectral and panchromatic images using fuzzy rule. Communications in nonlinear science and numerical simulation, vol 12. pp 1334–1350
3. Garg S, Kiran U, Mohan K, Tiwary R (2006) 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, vol 17–18. pp 7680–7683
4. Zhang Zh, Blum RS (1999) A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. Proc IEEE 87(8): 1315–1326
5. Song Y, Li M, Li Q, Sun L (2006) A new wavelet based multi-focus image fusion scheme and its application on optical microscopy. In: Proceedings of the 2006 IEEE international conference on robotics and biomimetics, pp 401–405
6. Rosenfeld A, Thurston M (1971) Edge and curve detection for visual scene analysis. IEEE Trans Comput C-20:562–569
7. Burt PJ, Adelson EH (1983) The Laplacian pyramid as a compact image code. IEEE Trans Commun 31(4):532–540
8. Lindeberg T (1994) Scale–Space theory in computer vision. Kluwer Academic Publisher, Dordrecht
9. Mueller D, Maeder A, O'Shea1 P (2006) The Generalised image fusion toolkit (GIFT), pp 1–16
10. Burt PJ (1992) A gradient pyramid basis for pattern selective image fusion. In: Proceedings of the society for information display conference, pp 467–470
11. Toet A (1989) A morphological pyramidal image decomposition. Pattern Recognit Lett 9(3):255–261
12. Li H, Manjunath BS, Mitra SK (1995) Multi-sensor image fusion using the wavelet transform. Graph Models Image Process 57(3):235–245
13. Pajares G, Cruz JM (2004) A wavelet-based image fusion tutorial. Pattern Recognit 37(9):1855–1872
14. Rockinger O (1997) Image sequence fusion using a shift invariant wavelet transform. In: IEEE international conference on image process, pp 288–291
15. Nu′fez J, Otazu X, Fors O, Pala V, Arbiol R (1999) Multiresolution based image fusion with additive wavelet decomposition. IEEE Trans Geosci Remote Sens 37(3):1204–1211
16. Chibani Y, Houacine A (2003) Redundant versus orthogonal wavelet decomposition for multisensor image fusion. Pattern Recogn 36(4):1785–1794
17. Hill P, Canagarajah N, Bull D (2002) Image fusion using complex wavelets. In: Proceedings of the British machine vision conference, pp 487–496
18. Burt PJ, Kolczynski RJ (1993) Enhanced image capture through fusion. In: Proceedings of the 4th international conference on computer vision, pp. 173–182
19. Kingsbury N (2000) A Dual-tree complex wavelet transform with improved orthogonality and symmetry properties. ICIP 2: 375–378
20. Kingsbury N (1998) The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters. In: Proceedings of 8th IEEE DSP Workshop, Utah, pp 9–12
21. Selesnick IW, Baraniuk RG, Kingsbury NG (2005) The dual-tree complex wavelet transform. In: IEEE signal processing magazine, pp 124–152
22. Wei S, Ke W (2007) A multi-focus image fusion algorithm with DT-CWT. In: International conference on computational intelligence and security, pp 147–151
23. Zheng Y, Li H, Doermann D (2004) Machine printed text and handwriting identification in noisy document images. IEEE Trans Pattern Anal Mach Intell 26(3):337–353
24. Kumar S, Gupta R, Khanna N, Chaudhury S, Joshi SD (2007) Text extraction and document image segmentation using matched wavelets and MRF model. IEEE Trans Image Process 16(8): 2117–2128
25. Fukunaga K (1990) Introduction to statistical pattern recognition, 2nd edn. Academic Press, New York
26. Kannan K, Perumal SA (2007) Optimal decomposition level of discrete wavelet transform for pixel based fusion of multi-focused images. In: International conference on computational intelligence and multimedia applications, pp 314–318
27. Petrovic VS, Xydeas CS (2004) Gradient-based multiresolution image fusion. IEEE Trans Image Process 13(2):228–237
28. Rockinger O, Fechner T (1998) Pixel-level image fusion: the case of image sequences. Proc SPIE 3374:378–388
29. Li S, James, Kwok T, Wang Y (2001) Combination of Images with diverse focuses using the spatial frequency. Inf Fusion 2(3): 169–176
30. Niu Y, Shen L (2009) A novel approach to image fusion based on multi-objective optimization. In: Proceedings of the 6th World congress on intelligent control, pp 9911–9915
31. Petrovic V, Xydeas C (2000) On the effects of sensor noise in pixel-level image fusion performance. In: Proceedings of the third international conference on image fusion, vol 2. pp 14–19
32. http://imagefusion.org
33. http://www.imgfsr.com

Author Biographies

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 multiresolution 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.