

Random forest-based scheme using feature and decision levels information for multi-focus image fusion

Nabeela Kausar · Abdul Majid

Received: 16 August 2014 / Accepted: 12 January 2015 / Published online: 6 February 2015
© Springer-Verlag London 2015

Abstract Often captured images are not focussed everywhere. Many applications of pattern recognition and computer vision require all parts of the image to be well-focussed. The all-in-focus image obtained, through the improved image fusion scheme, is useful for downstream tasks of image processing such as image enhancement, image segmentation, and edge detection. Mostly, fusion techniques have used feature-level information extracted from spatial or transform domain. In contrast, we have proposed a random forest (RF)-based novel scheme that has incorporated feature and decision levels information. In the proposed scheme, useful features are extracted from both spatial and transform domains. These features are used to train randomly generated trees of RF algorithm. The predicted information of trees is aggregated to construct more accurate decision map for fusion. Our proposed scheme has yielded better-fused image than the fused image produced by principal component analysis and Wavelet transform-based previous approaches that use simple feature-level information. Moreover, our approach has generated better-fused images than Support Vector Machine and Probabilistic Neural Network-based individual Machine Learning approaches. The performance of proposed scheme is evaluated using various qualitative and quantitative measures. The proposed scheme has reported 98.83, 97.29, 98.97, 97.78, and 98.14 % accuracy for standard images of Elaine, Barbara, Boat, Lena, and

Cameraman, respectively. Further, this scheme has yielded 97.94, 98.84, 97.55, and 98.09 % accuracy for the real blurred images of Calendar, Leaf, Tree, and Lab, respectively.

Keywords Image fusion · Multi-focus · Random forest · Ensemble · Machine learning

1 Introduction

Image fusion is a process of integrating useful information from two or more images to get an image which contains more information [1, 2]. Often captured images are not well-focussed everywhere due to limited field depth of optical lens of camera [3, 4]. When image is captured, only objects at a particular distance are focussed while other objects at larger distance than specified by the lens formula are not well-focussed. Thus, the captured image does not provide clear information, which is necessary for human visual perception as well as in the applications of pattern recognition and machine vision. To solve this problem image fusion approaches are developed. These approaches combine complementary information contained in the blurred images to provide more informative and clear images.

Previously, researchers have proposed several image fusion techniques in spatial and transform domains. In spatial domain, image fusion has been performed at pixel level, feature level, and decision level. The simplest technique is the pixel-level fusion that constructs the fused image by calculating pixel-by-pixel average of the blocks of input images. However, feature-level fusion techniques use the features information of the source images and construct the fused image using some fusion rule. Features

N. Kausar · A. Majid (✉)
Department of Computer and Information Sciences, Pakistan
Institute of Engineering and Applied Sciences,
Nilore, Islamabad 45650, Pakistan
e-mail: abdulmajid@pieas.edu.pk

N. Kausar
e-mail: nabeela.kausar@pieas.edu.pk

are extracted from principal component analysis (PCA) and Wavelet transform domains [5]. PCA-based fusion technique transformation performs conversion of the correlated variable to uncorrelated variables named as the principal components. It computes the compact dataset representation. The first principal component represents the most important data sample as it accounts for the maximum variance in the dataset and each succeeding component represents as much of the rest of the variance as possible.

Discrete Wavelet transform and its variant-based image fusion techniques are also proposed [6–9]. Multi-resolution transform-based fusion techniques are also developed using Laplacian pyramid [7], morphological pyramid [8], bi-orthogonal wavelet transform [6]. The performance of multi-resolution transform-based techniques is degraded due to down sampling. To minimize this effect dual-tree frame Wavelet-based technique has been developed [10]. The *multi-band vector Wavelet*-based fusion technique is proposed that computes the coefficients of high and low frequency bands. However, this technique produces some blocking effect in the fused image. Therefore, anisotropic diffusion arithmetic is also suggested as post-processing to obtain improved image [11]. The multi-band Wavelet is more generic compared to the two-band Wavelet transformation. Its performance is found superior than two-band Wavelet, due to its compact support, orthogonal aspects, and decomposition properties [11]. Recently, Goodman et al. [12] have developed *vector Wavelet* image fusion technique. This *Vector Wavelet*-based approach provides both synthesis and analysis operators. Therefore, this decomposition offers more design flexibility. It provides more benefits over scalar Wavelets with respect to short symmetry, support, and orthogonally [13].

At decision level, image fusion techniques have been developed using individual Machine Learning (ML) approaches like Support Vector Machine (SVM) and Artificial Neural Network (ANN). SVM-based image fusion technique has been developed using dual tree complex Wavelet transform [10]. Zhao et al. have suggested SVMs for multi-source image fusion in [14]. SVM-based models perform binary classification by finding the decision surface that has maximum margin between the two closest points. SVM constructs an optimal hyper-plane that minimizes the error for unseen test data samples. ANN based fusion technique is developed using feature-level fusion approach in [15, 16]. Probabilistic Neural Network (PNN) [17] and Radial Basis Function Neural Network (RBFN) [18], two well-known variants of ANN, are proposed for multi-focus image fusion. PNN approach develops the decision of focussed/de-focussed image blocks using Spatial Frequency (SF), visibility, and edge features [17]. PNN approach based on Bayes theory utilizes the computational power efficiently. It has more flexible characteristic than back-propagation neural

networks. PNN organizes its functional structure into four layers; input, pattern, summation, and decision layer. However, the fusion decision of ML approaches is limited for diverse type of images.

To address the issue, ensemble models are developed for better decision making. In ensemble scheme, individual models and their diversities can yield better decision. The deficiency in one model can be replaced with the advantages of others. The ensemble decision is considered more accurate and generalized than single machine learning approach [19]. In this study, we proposed random forest (RF)-based ensemble approach for multi-focus image fusion. This approach has been used effectively in different classification and prediction problems [20].

The advantage of RF approach is the generation of diverse types of random trees with reduced variance, i.e. weak learners. Such type of weak learners is primarily required for ensemble [21]. RF-based ensemble approach could compensate for the deficiency of one tree with the advantages of other trees. Therefore, RF algorithm can effectively utilize the diversity of random trees as compared to individual approaches. RF algorithm randomly generates many tree classifiers and aggregates their predictions. The generated trees are trained on bootstrap samples of the training dataset and employs random sub-sampling of features. This strategy offers resistance to over training and over-fitting on input data [22]. RF approach has performed well in high-dimensional feature space, especially, when dataset is small and its structure is complex.

The combined decision obtained through RF is used for the construction of decision map for image fusion. That is why; the proposed RF-based scheme has performed better on diverse types of blurred images. In the proposed scheme, useful features information is extracted from both spatial and transforms domains. These features are then used to train randomly generated trees by RF algorithm. The predicted labels of random trees are aggregated to construct the decision map for better image fusion. The proposed scheme is evaluated in terms of accuracy, specificity, sensitivity, F -score, and Mathew Correlation Coefficient (MCC). However, the quality of fused images is assessed using Peak Signal to Noise Ratio (PSNR), Mutual Information (MI), Root Mean Square (RMSE), correlation, Structure Similarity Index Measure (SSIM), Spatial Frequency (SF), Mean Gradient (MG), and standard deviation (STD). Performance of the proposed scheme is compared to previous image fusion approaches for synthesized and real blurred images. It is found that our proposed scheme has better quality-fused images. The main novelty of our scheme is the effective development of RF-based ensemble scheme for multi-focus image problems. Secondly, this scheme has incorporated both the feature and decision levels information to fuse multi-focus images.

The remaining part of this paper is organized as: In Sect. 2, we have presented various stages of the proposed scheme, in detail. Experimental setup is explained in Sect. 3. Results are discussed in Sect. 4. Finally, in Sect. 5 conclusions are given.

2 Proposed scheme

The block diagram of RF-based proposed scheme is given in Fig. 1. The input dataset is formed by decomposing each input image into blocks of size $M \times N$ (Implementation detail about the blocks is given in Sect. 3). The nine useful features are extracted from each block of the blurred/un-blurred images. These useful features provide blur information contained in these images. The detail information about these features is given in Sect. 2.3. RF algorithm efficiently utilized the input dataset to generate decision/classification trees. The predicted labels of randomly generated trees of random forest are combined using majority voting to construct the decision map for the fused image. Finally, various quality measures are employed to report the performance of fused images. Now, we will describe different stages of the proposed scheme.

2.1 Dataset formulation

Input dataset is formed by decomposing two differently blurred images I_1 and I_2 into blocks of size $M \times N$ pixels.

For each image block, $B_{I_1}^{(i)} \in I_1$ and $B_{I_2}^{(i)} \in I_2$, nine-dimensional feature vector is constructed, i.e. $V_{I_1}^{(i)} = (v_{11}, v_{12}, v_{13}, \dots, v_{19})$ and $V_{I_2}^{(i)} = (v_{21}, v_{22}, v_{23}, \dots, v_{29})$. Next, we assigned target labels $t^{(i)}$ for i th feature vector of focussed/unfocussed blocks. For this purpose, a binary array is created as:

$$A^{(i)} = \begin{cases} 1 & \text{if } (v_{1j} \geq v_{2j}) \text{ for } j = 1, 2, \dots, 9 \\ 0 & \text{Otherwise} \end{cases} \tag{1}$$

These values in the array, $A^{(i)} = \{0, 1\}_{j=1}^9$, belong to one of two classes of focussed/unfocussed blocks, i.e. $A^{(i)} \in \{c_1, c_2\}$. Target labels $t^{(i)}$ are assigned using the majority class votes, i.e.

$$t^{(i)} = \arg \max_{c_j} \sum_{k=1}^9 \Delta(a_k^{(i)}, c_j), \text{ for } j = 0, 1 \tag{2}$$

where, $\Delta(d_k^{(i)}, c^j) = \begin{cases} 1 & \text{if } d_k^{(i)} = c_j \\ 0 & \text{otherwise} \end{cases}$. The image block is categorized to the class that receives the maximum votes.

2.2 Working of RF algorithm

Figure 2 demonstrates the working diagram of RF algorithm using bootstrap subspace data sampling technique. During forest growing process, bootstrap technique randomly selects samples without replacement from the input dataset of n_b blocks, $S = \{(V^{(i)}, t^{(i)})\}_{i=1}^{n_b}$. We investigated the performance of RF algorithm by selecting different number of random trees. However, we found the optimal

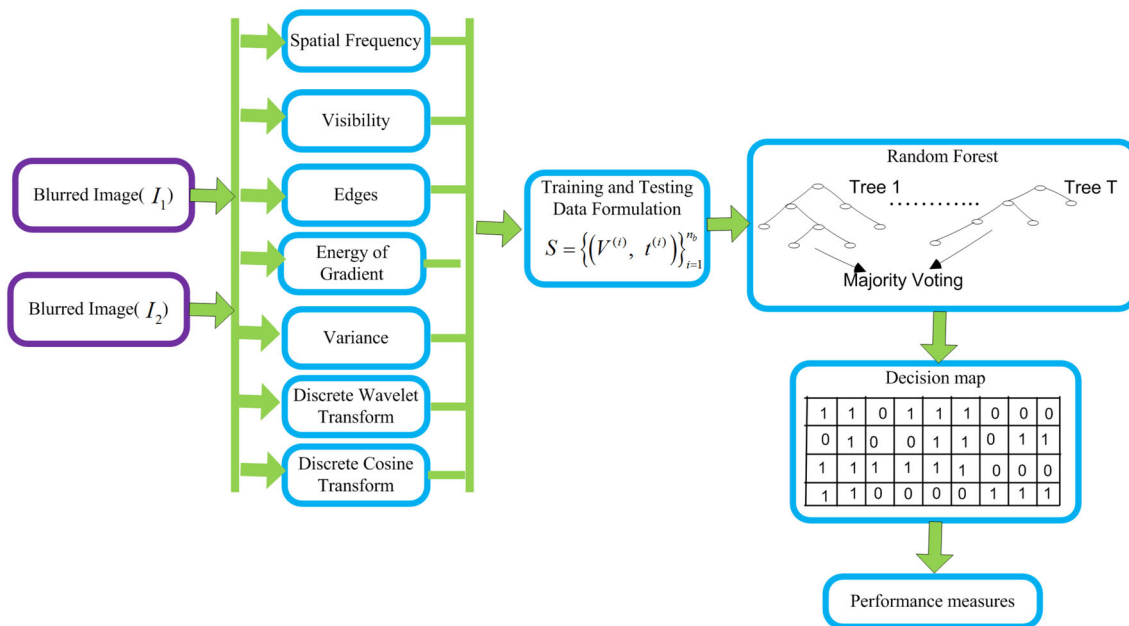
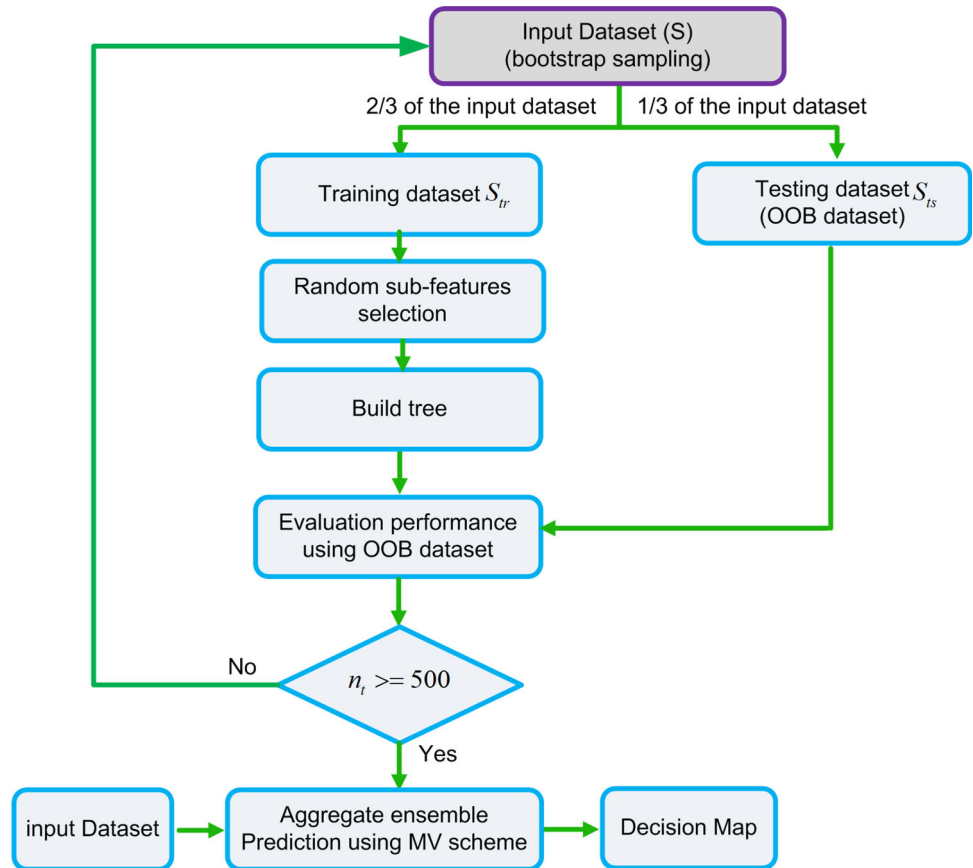


Fig. 1 Block diagram of the proposed RF based fusion scheme

Fig. 2 Working of RF algorithm to construct decision map



performance for $n_t = 500$ random trees. Each tree is built by randomly selecting 2/3rds of the input dataset with replacement called training dataset $S_{tr} = \{(V^{(i)}, t^{(i)})\}_{i=1}^{i=n_{tr}}$. The remaining 1/3rds testing dataset of n_s blocks, $S_{ts} = \{(V^{(i)}, t^{(i)})\}_{i=1}^{i=n_s}$, is kept for the evaluation.

To determine the best node for split, we selected a random subset (subspace) of m variables from training data. In general, the value of m variables is determined from D -dimensional feature vector, i.e. $m \approx \sqrt{D}$. However, in the proposed scheme, for nine-dimensional feature vector, the value of $m \approx 3$. These selected variables subset is used to compute the best node split according to Gini criteria [22]. This fitness criterion estimates the importance of variables. This measures the degree of association between variables and classification results. The lowest value of this measure is taken as the best split of each node. Gini measure is computed using the following equation as:

$$\text{Gini}(d, v_i) = \sum_{i=1}^p \frac{a_i}{n_s} I(d_{u_i}) \tag{3}$$

where, p represents the number of children at node d . Here, n_s shows the number of input feature vectors used for training. The Gini impurity $I(d_{v_i})$ gives the distribution of class label in the node. For a feature variable $v_i \in V$ with

p values at node d , $v_i = \{u_1, u_2, \dots, u_p\}$ the value of $I(d_{u_i})$ is computed as:

$$I(d_{u_i}) = 1 - \sum_{i=0}^c \left(\frac{n_{c_i}}{a_i} \right)^2 \tag{4}$$

where, n_{c_i} is the number of samples with value u_i belong to class c_i and a_i indicates the number of samples with value u_i at that node d .

After the construction of classification trees on training dataset, out-of-bag (OOB) testing data $S_{ts} = \{(V^{(i)}, t^{(i)})\}_{i=1}^{i=n_s}$ is used to evaluate the predictions performance. To build diverse type of weak classifiers, for good generalization, we have constructed 500 random classification trees with average accuracy is kept in the range of 60–70 %. In the next stage, the whole input dataset of n_b blocks, $S = \{(V^{(i)}, t^{(i)})\}_{i=1}^{i=n_b}$, is provided to 500 classification trees for the construction of depth map.

Figure 3 demonstrates an exemplary ensemble of random trees with associated features and predicted labels. In this figure, for each vector of input dataset, $S = \{(V^{(i)}, t^{(i)})\}_{i=1}^{i=n_b}$, the predicted labels are extracted using 500 RF trees as: $\hat{t}_1 = f_1(V^i), \hat{t}_2 = f_2(V^i), \dots, \hat{t}_{500} = f_{500}(V^i)$, where $\hat{t}_j = f_j(V^i)$ indicates the predicted value of f_j tree for

i th feature vector V^i . The predicted labels are aggregated using well-known majority voting (MV) technique i.e., $\hat{t}^i = MV(\hat{t}_1^i, \hat{t}_2^i, \dots, \hat{t}_{500}^i)$. The predicted values of \hat{t}^i are used to construct the decision map.

2.3 Image fusion using decision map

After construction of decision map, each block of fused image $B_F^{(i)}$ is selected from input images I_1 and I_2 using the corresponding value of the decision map as:

$$B_F^{(i)} = \begin{cases} B_{I_2}^{(i)} & \text{for } \hat{t}^i = 1 \\ B_{I_1}^{(i)} & \text{for } \hat{t}^i = 0 \end{cases} \quad (5)$$

White and black regions in the binary images of decision maps, in Figs. 4 and 5, show the complimentary information to be picked from two source images. In Fig. 4c, the black regions indicate the focussed parts of the image in Fig. 4a and the white regions of the image in the Fig. 4b. Similar is the case for the Calendar image in Fig. 5. Both these figures demonstrate that our scheme has accurately selected the focussed image blocks from the input images.

2.4 Feature extraction stage

To form diverse type of training data for RF model, we have selected the most informative/discriminative features both from spatial and frequency domain. These features are used effectively in the literature to measure the image blur

[18, 23–27]. The dataset contains nine features in which five features are taken from the spatial domain and remaining four features are extracted from the frequency domain. Spatial domain features consist of visibility, spatial frequency, energy of gradient, variance, and edges. The remaining four frequency domain features include one discrete cosine transform (DCT) high frequency and three detail bands of DWT. The details of these features are given as:

Visibility (VIS): This feature is derived from human visual system [23]. It measures the difference among intensities of the block pixels and mean intensity value of the image block and shows the clarity of image block. It is given by:

$$VIS = \sum_{m=1}^M \sum_{n=1}^N \left| \frac{B(m, n) - \mu_B}{\mu_B^{\alpha+1}} \right| \quad (6)$$

where α is a constant and μ_B is mean intensity value of image block.

Spatial Frequency (SF): This is a measure of activity level in the image [24]. It computes the difference of frequency among the rows and columns of image blocks. The larger value of SF shows greater focus in the image block. This frequency is given as:

$$SF = \sqrt{(RF)^2 + (CF)^2} \quad (7)$$

Here, RF and CF are the row and column frequencies, respectively. The RF and CF are computed as:

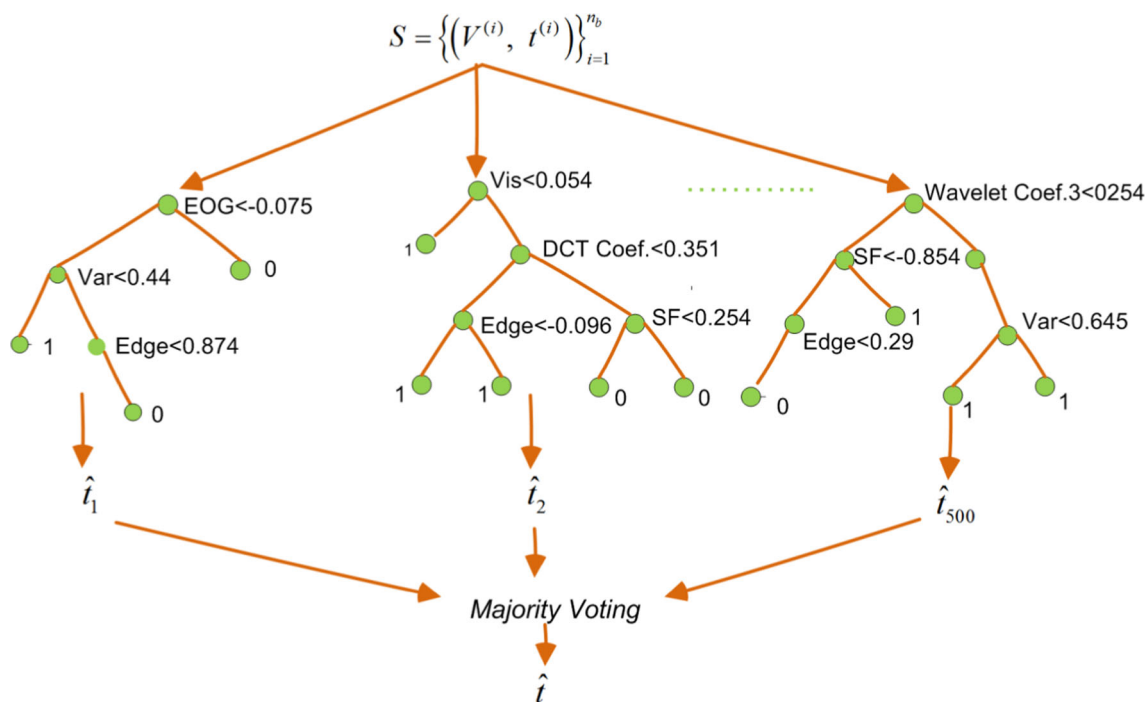


Fig. 3 Ensemble tree along with associated features and final predicted label

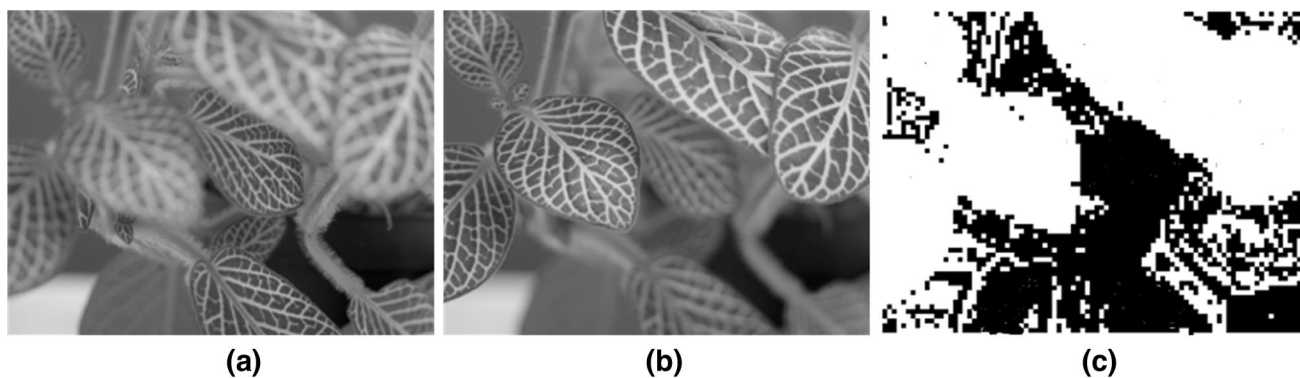


Fig. 4 Real blurred leaf image (a) image 1 (b) image 2 (c) corresponding binary image of the decision map

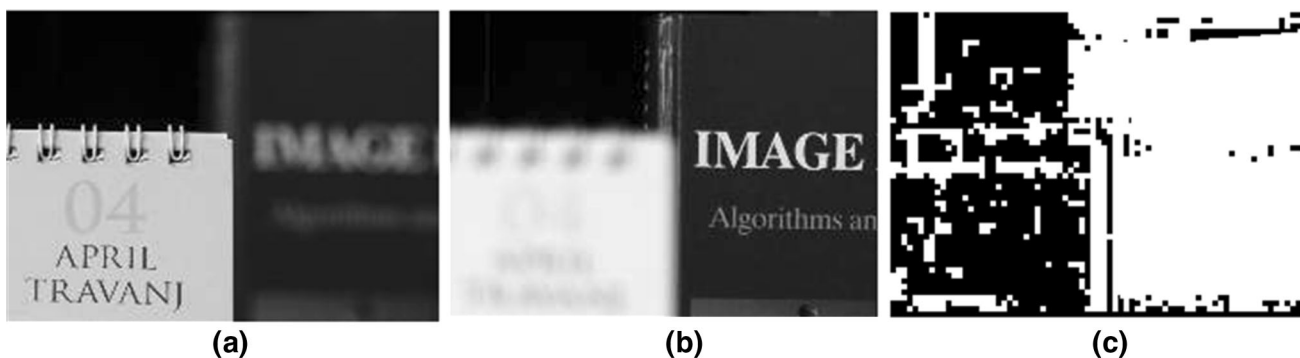


Fig. 5 Real blurred Calendar image (a) image 1 (b) image 2 (c) corresponding binary image of the decision map

$$RF = \sqrt{\frac{1}{MN} \sum_{m=1}^M \sum_{n=2}^N (B(m, n) - B(m, n-1))^2} \quad (8)$$

$$CF = \sqrt{\frac{1}{MN} \sum_{m=1}^M \sum_{n=2}^N (B(m, n) - B(m-1, n))^2} \quad (9)$$

Edges: Edge feature indicates focus of an image. The focussed image blocks consist of more edges than the corresponding de-focussed image blocks. In our current work we have used Canny edge detection algorithm, as proposed in [25], to detect the edges in the image blocks.

Energy of Gradient (EOG): EOG is used to measure focus of an image. The smaller value of EOG shows the block is de-focussed and focussed image block has the greater value of EOG [26]. It is calculated as:

$$EOG = \sum_{i=1}^M \sum_{j=1}^N (B_i^2 + B_j^2) \quad (10)$$

where, $B_i = B(i+1, j) - B(i, j)$ and $B_j = B(i, j+1) - B(i, j)$

Variance: Variance measures focus of the image by computing the gray level contrast between the pixels of

image block and mean value of block. The value of variance is given as:

$$VAR = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N (B(m, n) - \mu_B)^2 \quad (11)$$

where, μ_B is mean value of the image block.

Discrete Wavelet Transform (DWT): DWT coefficients show the focus of image. The activity level of this transform can be measured at coefficients level by treating each coefficient separately [27]. Another approach computes window-based activity (WBA) by averaging the coefficients over blocks of the image [18, 27]. DWT decomposition of image blocks gives three detail sub-bands and one approximation sub-band. The approximation sub-band does not contain much of the edge information. Therefore, it does not contain useful information about the focus of the image. Hence, the approximation band is not included in the feature set for RF training. Only three detail sub-bands HL, LH, and HH are used in the construction of feature space. The activity level is computed by averaging the Wavelet coefficients over image blocks for each detail sub-band.

High Frequency Energy: Discrete Cosine Transform (DCT) coefficients highlight the variation in frequency information of the image. The high frequency coefficients

of this transform give the focus measure of the image blocks, which can be obtained by removing the direct components. The input image $B(u, v)$ in DCT domain is defined as:

$$B(u, v) = \frac{2}{L} c(u)c(v) \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} B(i, j) \cos \frac{(2i+1)\pi u}{2L} \cos \frac{(2j+1)\pi v}{2L} \tag{12}$$

here, $c(0)=1/\sqrt{2}$, $c(u) = 1$, and $u, v = 0, 1, \dots, L-1$.

DCT high frequency energy (E) is computed as:

$$E = \left[\sum_{u=0}^{L-1} \sum_{v=0}^{L-1} B^2(u, v) \right] - B^2(0, 0) \tag{13}$$

2.5 Performance measures

To demonstrate the comparative performance of multi-resolution transforms, individual learners, and combine decision model, we used several performance measures. They are explained as follows:

Accuracy (Acc): The prediction Acc of a classifier is calculated as:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \times 100 \tag{14}$$

where, TP (True Positive) is number of correctively predicted labels of positive class and FP is number of incorrectly predicted labels of positive class. TN is number of correctively predicted labels of negative class and FN is number of incorrectly predicted labels of negative class.

Sensitivity (Sn) and Specificity (Sp): These measures show the capability of classifier to predict the positive (negative) class. These measures are calculated as:

$$Sn = \frac{TP}{TP + FN} \times 100 \tag{15}$$

$$Sp = \frac{TN}{FP + TN} \times 100 \tag{16}$$

Mathew Correlation Coefficients (MCC): This measure is discrete form of Pearson’s correlation coefficient. It accounts for both overprediction and underprediction. Its value varies between -1 and 1. If MCC value is 1, it indicates the classifier makes no mistakes in prediction and if -1 then it shows the classifier prediction is wrong. However 0 MCC value means random prediction. MCC is calculated as:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{[TP + FP][TP + FN][TN + FP][TN + FN]}} \tag{17}$$

F-Score: F-score evaluates the statistical tests. It uses Precision and Recall to compute the prediction accuracy and can be considered as weighted average of precision and recall.

Precision is number of correct prediction divided by total number of returned predictions and recall is correct predictions divided by number of predictions. Its value is computed as:

$$F \text{ - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{18}$$

where, Precision = $\frac{TP}{TP+FP}$ and Recall = $\frac{TP}{TP+FN}$

Standard Deviation (STD): STD measures the contrast in the fused image. An image with high contrast would have a greater STD value. Its value is computed as:

$$STD = \sqrt{\sum_{i=1}^L (i - i')^2 h_F(i)} \tag{19}$$

where, L is the number of gray levels and $i' = \sum_{i=1}^L i h_F$ and h_F is the normalized histogram of fused image.

Mean Gradient (MG): MG is the mean value of the gradient of the final fused image. Gradient usually represents the image details and clear image regions. Then, a larger value of mean gradient indicates that the image contains more image details and clear image regions. The mean gradient is calculated as:

$$MG = \frac{1}{(M-1)(N-1)} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \sqrt{\left(\frac{\partial B_F(i, j)}{\partial i}\right)^2 + \left(\frac{\partial B_F(i, j)}{\partial j}\right)^2} \tag{20}$$

Spatial Frequency (SF): It is an important measure of the image details. The higher spatial frequency value indicates the more image details and further details are given in previous Sect. 2.3.

Root Mean Square Error (RMSE): The RMSE measure calculates the deviation of fused image from the original image. The lesser deviation shows the more similarity between the fused image and the original image. The RMSE value is calculated as:

$$RMSE = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (O(i, j) - F(i, j))^2} \tag{21}$$

where, M, N represents the size of the original image O and the fused image F .

Mutual Information (MI): MI calculates the similarity between the fused image and the original Image. The larger MI value indicates greater similarity between the fused and original images. MI is computes as:

$$MI = \sum_{l_1=1}^L \sum_{l_2=1}^L h_{O,F}(l_1, l_2) \log_2 \frac{h_{O,F}(l_1, l_2)}{h_O(l_1)h_F(l_2)} \tag{22}$$

where, h_O, h_F , and $h_{O,F}$ are original, fused, and joint, image histograms, respectively. L is number of intensity levels in the original and fused images.



Fig. 6 **a** Right focussed, **b** left focussed, **c** original image, **d** fused using RF

Peak Signal to Noise Ratio (PSNR): PSNR value is used to measure the resemblance between the original and fused image. PSNR value is computed as:

$$\text{PSNR} = 20 \log_{10} \frac{L^2}{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (O(i, j) - F(i, j))^2} \quad (23)$$

Correlation (Corr.): The Corr value indicates the similarity between the original image and fused image. Its values vary between 0 and 1. The larger correlation value indicates the greater similarity between the original and fused images. The value of Corr is calculated as:

$$\text{Cor} = \frac{2 \sum_{i=1}^M \sum_{j=1}^N O(i, j)F(i, j)}{\sum_{i=1}^M \sum_{j=1}^N (F(i, j))^2 + \sum_{i=1}^M \sum_{j=1}^N (O(i, j))^2} \quad (24)$$

Structural Similarity Index Measure (SSIM): SSIM measure computes the similarity between two images. The SSIM is calculated by combining a comparison of luminance, contrast, and structure. It is applied locally in an 8×8 square window. The window is moved pixel-by-pixel over the whole image. At each step, the local statistics and the SSIM index are calculated within the window. SSIM values range between [0 1]. Values close to 1 show

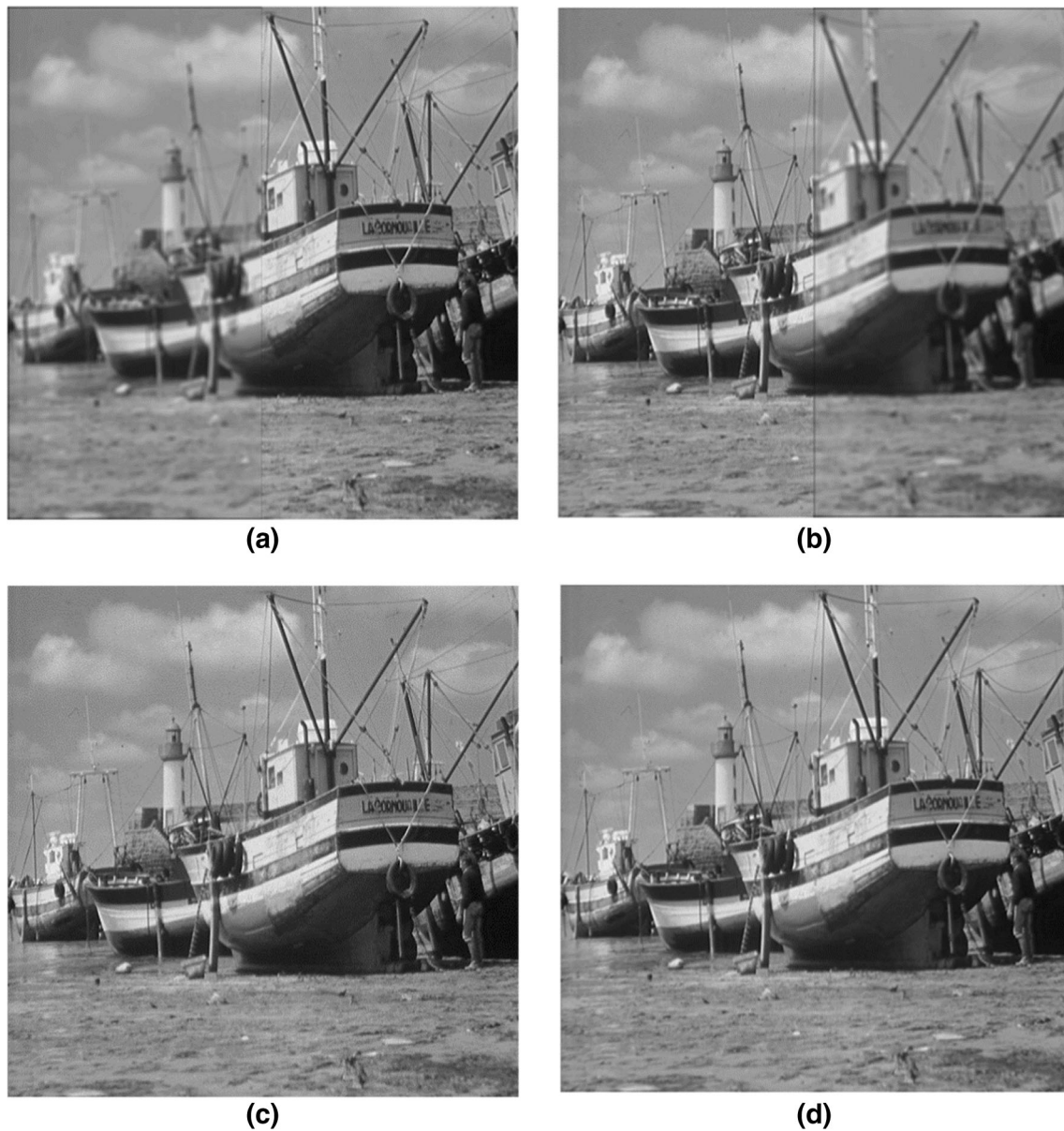


Fig. 7 a Right focussed, b left focussed, c original image, d fused using RF

the highest similarity between original and fused Images [28]. SSIM is calculated as:

$$SSIM(O, F) = \frac{(2\mu_O\mu_F + C_1)(2\sigma_{OF} + C_2)}{(\mu_O^2 + \mu_F^2 + C_1)(\sigma_O^2 + \sigma_F^2 + C_2)} \quad (25)$$

μ_O and μ_F are mean intensities of original and fused images, respectively. σ_O and σ_F are the standard deviations of original and fused images, respectively. σ_{OF} is combine standard deviation of fused and original image. Here, $C_1 = 6.5$ and $C_2 = 58.5$ are constants.

3 Experimental setup

Several experiments are conducted using a pair of multi-focus images to analyze the performance of the proposed scheme. Performance is evaluated using both synthesized and real blurred images. To obtain the synthesized blurred images, we blurred the left half of original image to form image I_1 , and right half of original image to form image, I_2 with the Gaussian blur of radius 5. Further, we blurred blocks of different sizes (80×80 and 64×64) of the original image to

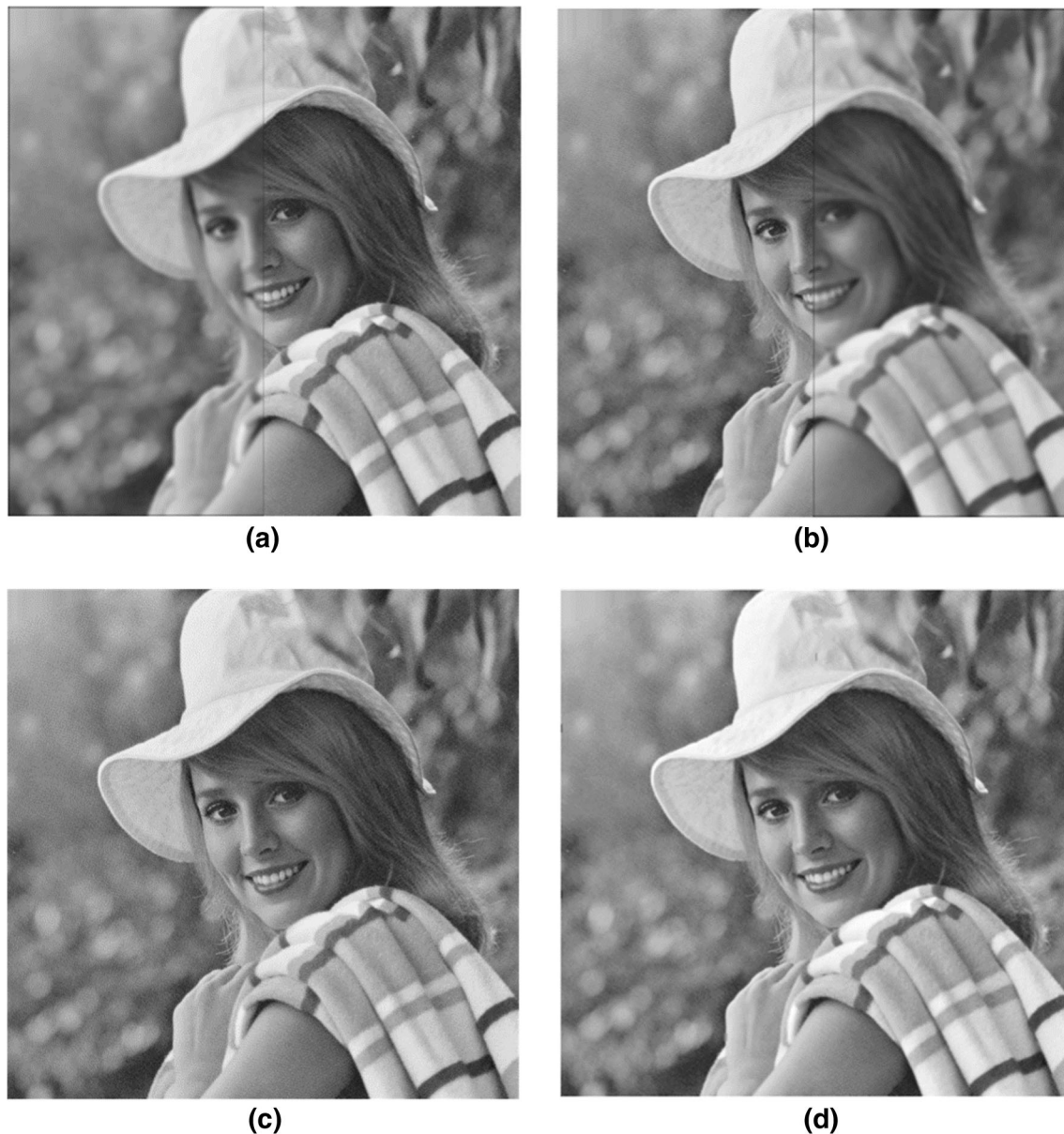


Fig. 8 **a** Right focussed, **b** left focussed, **c** original image, **d** fused using RF

form the image I_1 and I_2 , with Gaussian blur of radius 7. Real blurred images are obtained from www.imagefusion.org. Since multi-focus images, have non-uniform magnitude of focussed pixels, as a result, the block-based fusion techniques may affect the quality of the fused image at the block boundary. To minimize this effect, we used suitable block size of 3×3 for real blurred images. Further, block size of 8×8 pixels is used for the synthetic blurred images. If we further reduce block size, the computational time increases without appreciable increase in image quality. To develop SVM-based fusion model, we computed the optimal value of model parameters ($c = 70$) that controls the model complexity and $\gamma = 0.05$ of RBF function. PNN-based fusion

model develops using the optimal value of spread parameter ($\sigma = 0.9$). PCA-based image fusion technique is employed by arranging input images into two column vectors as proposed in [5]. The values of empirical means, eigenvalues, and eigenvectors are calculated. The eigenvectors that have the larger eigen values are obtained. The normalized components P_1 and P_2 (i.e., $P_1 + P_2 = 1$) are calculated to obtain the fused image as:

$$F(x, y) = P_1 I_1(x, y) + P_2 I_2(x, y) \quad (26)$$

Fusion techniques are developed in Wavelet domain by decomposing the input images into detail and approximation bands as proposed in [5]. These details and

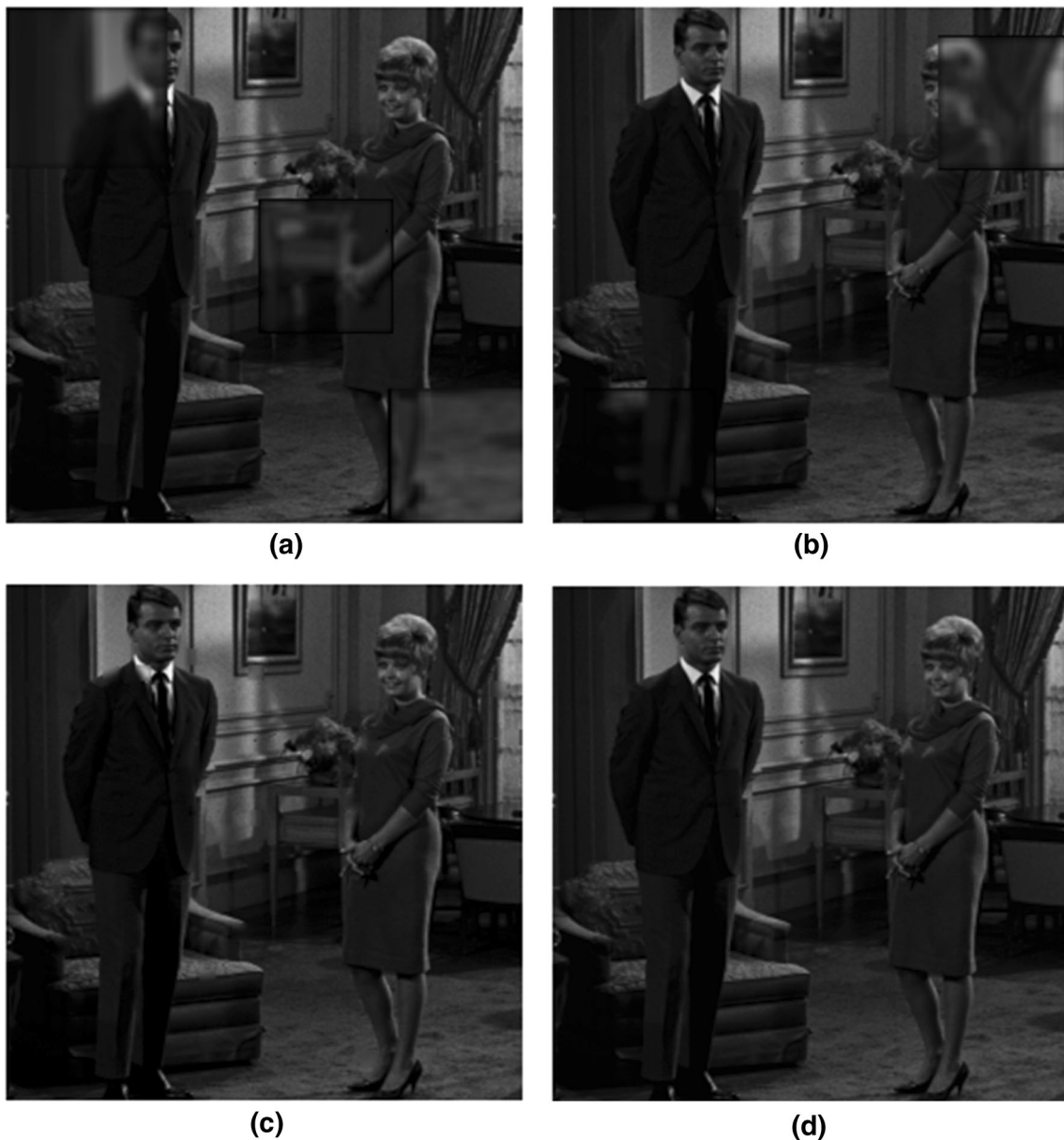


Fig. 9 Couple Image **a** blurred image 1, **b** blurred image 2, **c** original Image, **d** fused using RF

approximation coefficients are fused according to the fusion rule. The fused image is obtained by applying the inverse transform. Image fusion in Wavelet domain is performed using stationary Wavelet transform with “symlet” function of 2nd order. Input images are decomposed to first level.

4 Results and discussion

In this section, we have compared the qualitative and quantitative performance of the proposed fusion scheme with other approaches based on Wavelet [27], PCA [27], SVM [29], and PNN [17].

The qualitative performance is demonstrated in the Figs. 6, 7, 8, 9 for the images Barbara, Boat, Elaine, and Couple, respectively. These figures show that, for synthesized blurred images, the fused images using the proposed scheme have produced better quality and they have greater similarity with the original images. Further, Figs. 10, 11, 12 highlight the visual quality of the real blurred images fused using our RF-based scheme. It is observed that the fused image obtained through our scheme have better quality than the Wavelet-, PNN-, and PCA- based fusion techniques. From Fig. 10 it is observed that the fused Leaf’s image obtained through PCA approach keep the blurring effects. On the other hand, our scheme has produced better quality fused image than the

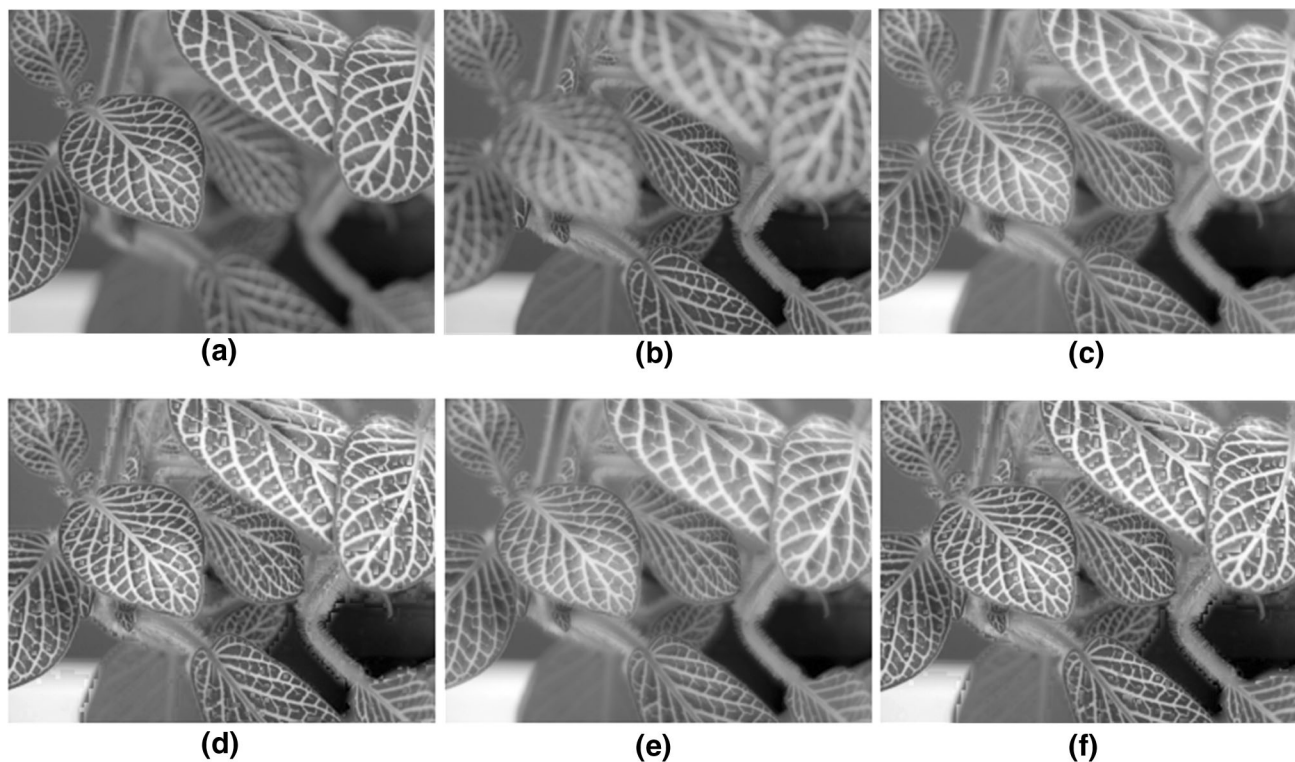


Fig. 10 a Blurred image 1, b blurred image 2, c fused using Wavelet, d fused using PNN, e fused using PCA, f fused using RF

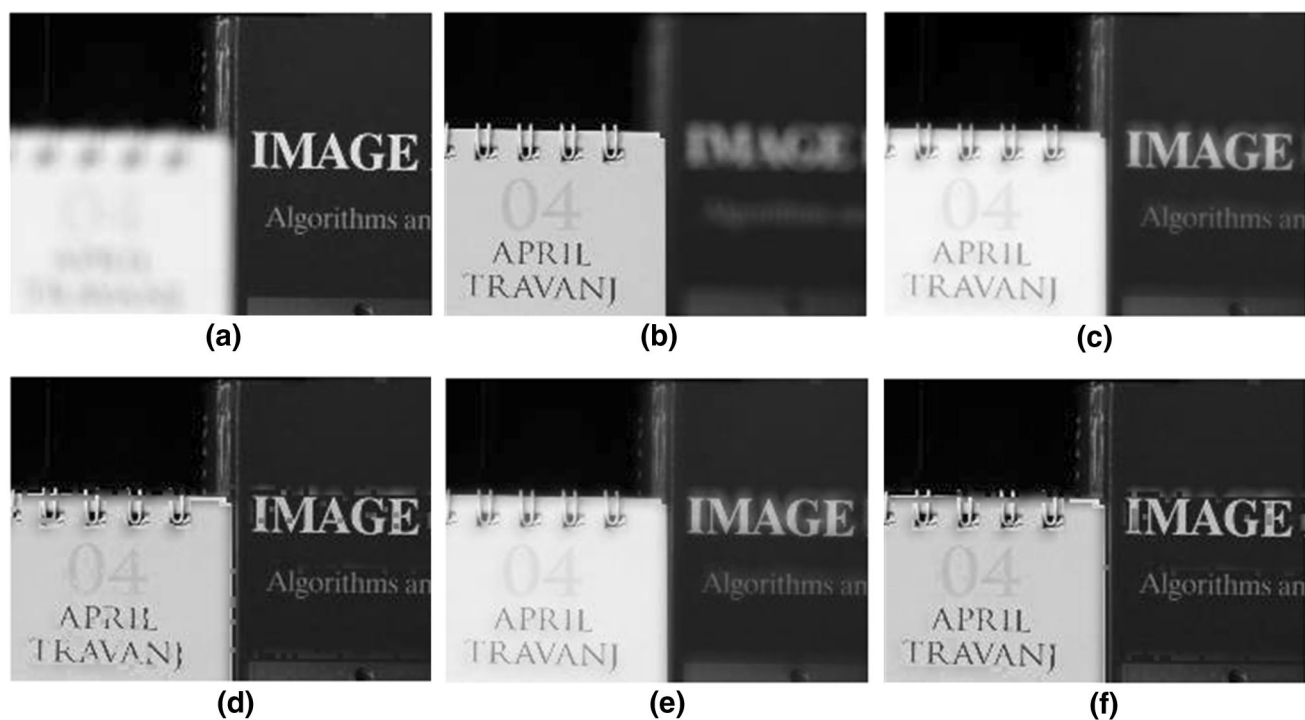


Fig. 11 a Blurred image 1, b blurred image 2, c fused using Wavelet, d fused using PCA, e fused using SVM, f fused using RF

images fused using PNN- and Wavelet-based approaches. We observed in Fig. 12 the blurring effects in fused images obtained through the Wavelet and PCA approach.

However, it is noted from this figure that the proposed scheme has better quality-fused images than previous Wavelet-, PCA-, and SVM-based approaches. It is clear

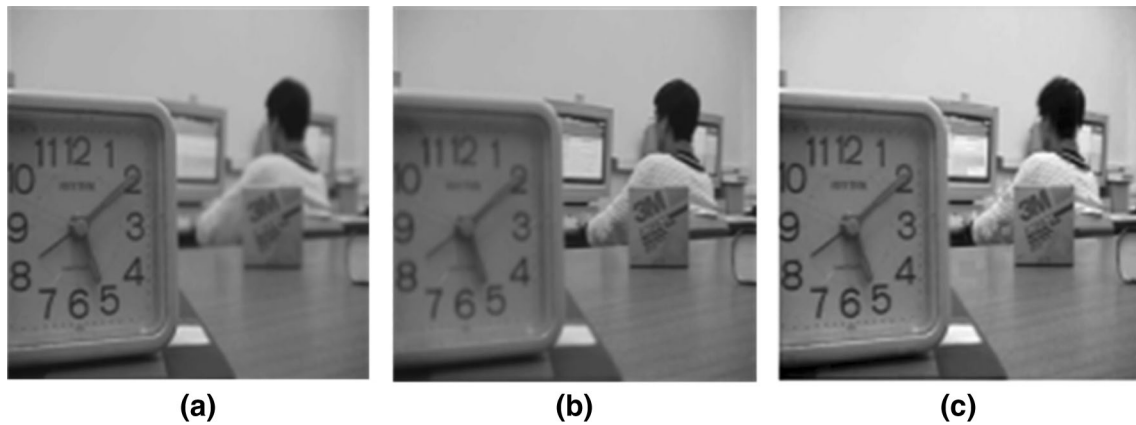


Fig. 12 **a** Blurred image 1, **b** blurred image 2, **c** fused using RF

Table 1 Performance comparison of the proposed scheme with SVM [27] and PNN [17]

Images	Models	Acc	Sn	Sp	F-Score	MCC
Elaine	SVM	94.04	92.77	95.31	0.9397	0.677
	PNN	94.53	95.31	93.75	0.945	0.691
	Proposed	98.83	98.88	98.78	0.988	0.703
Barbara	SVM	91.92	95.17	88.67	0.922	0.696
	PNN	69.14	54.69	83.59	0.639	0.825
	Proposed	97.29	97.41	97.17	0.973	0.698
Boat	SVM	94.214	93.457	94.97	0.942	0.681
	PNN	92.19	92.97	91.41	0.923	0.681
	Proposed	98.97	98.88	99.07	0.990	0.703
Lena	SVM	89.99	87.40	92.58	0.897	0.655
	PNN	72.66	92.19	53.13	0.771	0.671
	Proposed	97.78	97.95	97.61	0.978	0.700
Cameraman	SVM	89.45	88.09	90.82	0.8931	0.658
	PNN	96.88	93.75	99.99	0.968	0.678
	Proposed	98.14	98.44	97.85	0.982	0.701
Couple	SVM	92.35	97.25	98.23	0.9723	0.536
	PNN	93.25	92.36	97.25	0.9623	0.640
	Proposed	96.48	98.25	96.99	0.985	0.703

from Table 1 that our scheme has exhibited sufficient higher Acc values of 98.97, 97.78, 98.14, and 96.48 for Elaine, Barbara, Boat, Boat, Lena, Cameraman, and Couple images, respectively. However, for these images, PNN and SVM approaches give relatively lower values of Acc measure.

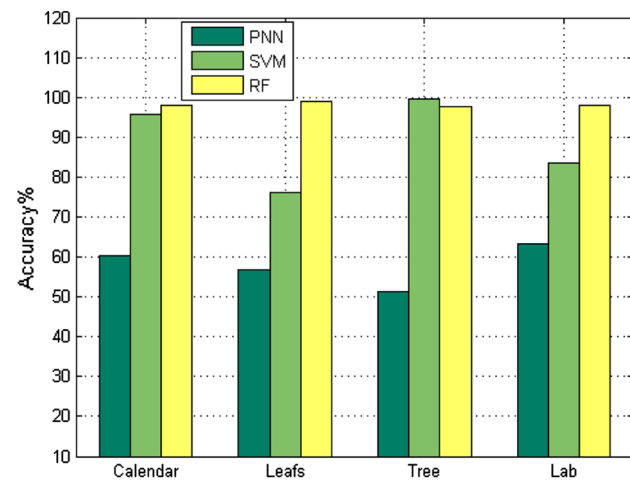
Our scheme has demonstrated higher Sn values of 98.88, 97.41, 98.88, 97.95, 98.44, and 98.25 for these images. On the other hand, for these images, PNN and SVM approaches give relatively lower values of Sn measure.

Table 2 Comparison of proposed scheme with Wavelet [27], PCA [25], SVM [29], and PNN [17]

Image	Model	PSNR	MI	RMSE	Corr	SSIM
Elaine	Wavelet	39.62	5.520	2.60	0.998	0.997
	PCA	36.48	4.016	3.63	0.997	0.9520
	SVM	40.06	19.88	2.41	0.999	0.986
	PNN	41.35	20.16	2.08	0.999	0.990
Barbara	Proposed	47.87	21.50	0.98	0.999	0.997
	Wavelet	37.72	4.86	3.17	0.998	0.994
	PCA	30.74	3.60	7.07	0.992	0.955
	SVM	33.28	19.55	5.33	0.995	0.974
Boat	PNN	32.19	14.19	6.04	0.994	0.963
	Proposed	38.13	21.4 0	3.05	0.998	0.992
	Wavelet	38.12	4.65	3.09	0.998	0.995
	PCA	34.50	3.55	4.59	0.995	0.963
Cameraman	SVM	37.94	18.60	3.23	0.998	0.985
	PNN	39.33	18.08	2.75	0.998	0.988
	Proposed	46.57	20.33	1.20	0.999	0.997
	PCA	32.39	4.15	6.04	0.995	0.968
Couple	Wavelet	32.71	4.41	5.83	0.996	0.977
	SVM	31.75	6.25	6.54	0.995	0.973
	PNN	38.15	6.78	3.13	0.999	0.990
	Proposed	41.39	6.86	2.16	0.999	0.994
Lena	PCA	37.37	4.40	3.22	0.998	0.979
	Wavelet	39.31	5.41	2.59	0.998	0.996
	SVM	38.34	18.59	2.97	0.998	0.984
	PNN	37.35	16.41	3.32	0.998	0.986
Couple	Proposed	42.74	20.93	1.79	0.999	0.996
	PCA	32.29	3.25	5.01	0.9886	0.9549
	Wavelet	26.17	2.45	10.34	0.9455	0.7843
	SVM	30.65	4.89	5.26	0.9825	0.9726
Couple	PNN	36.25	4.35	4.02	0.9566	0.9825
	Proposed	41.62	6.14	2.02	0.9980	0.9941

Table 3 Performance comparison of the proposed scheme with PCA, Wavelet, PNN, and SVM approaches for real images

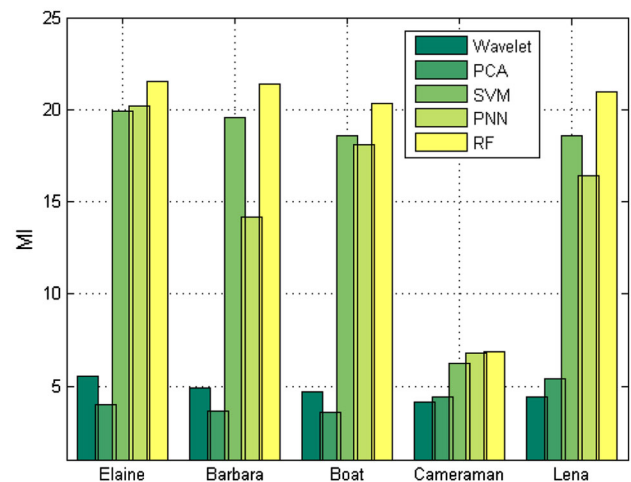
Image	Method	SF	MG	STD
Leafs	PCA	13.959	0.017	40.781
	Wavelet	20.727	0.051	6.970
	PNN	23.241	0.036	43.539
	SVM	22.071	0.001	44.101
	Proposed	27.014	0.037	44.921
Lab	PCA	16.393	0.192	44.891
	Wavelet	19.830	0.006	0.0660
	PNN	19.834	0.197	45.672
	SVM	34.700	0.190	44.922
	Proposed	20.519	0.226	45.693
Calendar	PCA	12.887	0.316	61.702
	Wavelet	13.203	0.017	1.551
	PNN	11.635	0.327	62.254
	SVM	22.891	0.317	62.645
	Proposed	23.914	0.353	64.617
Tree	PCA	21.423	0.107	34.409
	Wavelet	37.087	0.276	3.9555
	PNN	24.002	0.336	39.875
		28.000	0.137	38.306
	Proposed	32.961	0.358	39.953

**Fig. 13** Performance comparison of proposed scheme with PNN and SVM in terms of Acc

Similarly, our scheme has sufficient improved values than SVM and PNN approaches in terms of Sp, *F*-score, and MCC measures. It is pertinent to mention that proposed scheme observe little bit lower Sp and MCC values of 97.85 and 0.698 for Cameraman and Barbara images, respectively. However, values of Acc, Sn, and *F*-Score calculated by proposed scheme are much higher than that of PNN for the same images.

Table 4 Performance comparison of the proposed scheme with Wavelet-based recent approaches [11]

Image	Model	PSNR	MI	RMSE
Cameraman	Four-band Wavelet	33.238	3.036	5.554
	Vector Wavelet	31.808	3.100	6.548
	Multi-band vector Wavelet	34.423	3.610	4.846
	Proposed	41.390	6.860	2.160
Lena	Four-band Wavelet	34.575	2.937	4.762
	Vector Wavelet	34.162	3.181	4.993
	Multi-band vector Wavelet	36.968	3.654	3.615
	Proposed	42.740	20.930	1.790

**Fig. 14** Performance comparison of the proposed approach in terms of MI

It can also be observed from Table 1 that for Elaine, Boat, and Lena images the performance of proposed scheme is much better than PNN and SVM. Overall result of our proposed scheme is better.

Table 2 mimics the performance comparison of proposed scheme with previous PCA-, Wavelet-, PNN-, and SVM-based approaches for different blurred images. It is depicted from Table 2 that our proposed scheme has sufficient higher PSNR values of 47.87, 38.13, 46.57, 41.39, 42.74, and 41.62 for Elaine, Barbara, Boat, Cameraman, Lena, and Couple images, respectively. However, for these images, PCA-, Wavelet-, PNN-, and SVM-based approaches give relatively lower values of PSNR measure. Our scheme has yielded quite lower RMSE values for these images, as well. On the other hand, for these images, PCA-, Wavelet-, PNN-, and SVM-based approaches have given relatively higher values of RMSE. Similarly, in terms of MI, Corr, and SSIM measures, our scheme has demonstrated significant improvement than that of PCA-,

Wavelet-, PNN-, and SVM-based approaches. The higher values of MI and Corr highlight that our scheme has transferred more effectively useful information from multi-focus images to fused images. The higher value of SSIM indicates that images fused through the proposed scheme preserve structure similarity more effectively. Figure 14 visually shows that our proposed scheme has higher MI value for various images.

Table 3 shows the comparison of proposed scheme with PCA, Wavelet, PNN, and SVM approaches for real blurred images. It is observed from the table that, our proposed scheme has sufficient higher SF values of 27.01, 20.51, 23.91 and 32.96, for Leafs, Lab, Calendar, and Tree, images, respectively. However, for these images, PCA, Wavelet, PNN, and SVM approaches give relatively lower values of SF measure. This indicates our scheme preserves spatial information more effectively than PCA, Wavelet, PNN, and SVM approaches. Our scheme has given quite higher MG values for these images. While, for the same images, PCA, Wavelet, PNN, and SVM approaches observed relatively lower values of MG. Similarly, proposed scheme has significant improved values of STD than that of PCA, Wavelet, PNN, and SVM approaches. Figure 13 shows that our proposed scheme has more accuracy value as compared to SVM and PNN approaches for various real blurred images.

Table 4 shows the performance comparison of the proposed scheme with recent Wavelet-based image fusion approaches of four-band Wavelet, vector Wavelet, and multi-band vector Wavelet proposed in [11]. It is observed from Table that proposed scheme outperforms these Wavelet-based approaches in terms of PSNR, MI, and RMSE measures.

It is clear from Table 4 that our proposed scheme has noteworthy higher PSNR values of 41.39 and 42.74 for Cameraman and Lena images, respectively. However, for these images, four-band Wavelet, vector Wavelet, and multi-band vector Wavelet approaches give relatively lower values of PSNR. Similarly proposed scheme has provided higher MI values that that of four-band Wavelet, vector Wavelet, and multi-band vector Wavelet approaches. The proposed scheme has yielded lowest RMSE values of 2.16 and 1.79 for the same images. On the other hand, for these images, Wavelet-based recent approaches, give relatively higher values of RMSE. Figure 14 indicates that our proposed scheme has produced very good results in terms of MI than previous fusion approaches.

5 Conclusion

In this paper, we have studied the performance of RF based ensemble scheme for multi-focus image fusion in terms of

various quality measures. The proposed scheme has effectively incorporated both feature and decision level information. The feature-level information is extracted from both spatial and frequency domains. These informative features are tailored, during RF algorithm, for useful ensemble of diverse type of random trees. The predicted labels of ensemble trees are aggregated to construct the image fusion. The proposed scheme has yielded improved performance on various standard images and real blurred images. Our scheme has demonstrated better-fused image than feature-level approaches in PCA and wavelet transform. Moreover, our novel ensemble approach has given better-fused images than SVM- and PNN-based approaches under various qualitative and quantitative measures.

It is anticipated that that our scheme can be used as a useful tool for other applications of multi-focus image fusion. In future, we intend to extend our study for medical image fusion.

Acknowledgments This research work is financially supported by the Higher Education Commission of Pakistan under the indigenous PhD scholarship No. 17-5-4(Ps4-101) HEC/Ind-Sch-2007.

References

- Mitianoudis N, Stathaki T (2007) Joint fusion and blind restoration for multiple image scenarios with missing data. *Comput J* 50(6):660–673
- Susperregi L, Arruti A, Jauregi E, Sierrac B, Martínez-Otzeta JM, Lazkanoc E, Ansuategui A (2013) Fusing multiple image transformations and a thermal sensor with Kinect to improve person detection ability. *Eng Appl Artif Intell* 26(8):1980–1991
- Zhu H, Liu M, Ji H, Li Y (2010) Combined invariants to blur and rotation using Zernike moment descriptors. *Pattern Anal Appl*. doi:10.1007/s10044-009-0159-9:309-319
- Dai X, Zhang H, Liu T, Shu H, Luo L (2014) Legendre moment invariants to blur and affine transformation and their use in image recognition. *Pattern Anal Appl* 17(2):311–326
- Naidu VPS, Raol JR (2008) Pixel-level image fusion using wavelets and principal component analysis. *Def Sci J* 58(3):338–352
- Hamzaa AB, Heb Y, Krime H, Willskyd A (2005) A multiscale approach to pixel-level image fusion. *Integr Comput Aided Eng* 12:135–146
- Burt PJ, Adelson EH (1983) The Laplacian pyramid as a compact image code. *IEEE Trans Commun* 31(4):532–540
- Matsopoulos GK, Marshall S, Brunt JNH (1994) Multiresolution morphological fusion of MR and CT images of the human brain. *IEE Proc Vision Image Signal Proc* 141(3):137–142
- Burt PJ, Kolczynski RJ (1993) Enhanced image capture through fusion. In: Fourth international conference on computer vision 1993, pp 173–182
- Mumtaz A, Choi TS, Majid A, Mumtaz A (2010) Image fusion algorithm based on dual tree complex wavelet transform and support vector machine. In: International Bhurban conference on applied sciences & technology, Islamabad, Pakistan, pp 197–202
- Lan Y, Ren H, Zhang Y (2013) Multi-band vector wavelet transformation based multi-focus image fusion algorithm. *J Softw* 8(1):208–217

12. Goodman TNT, Lee SL, Tang WS (1993) Wavelets in wandering subspaces. *Trans Am Math Soc* 338(2):639–654
13. Piella G (2003) A general framework for multiresolution image fusion: from pixels to regions. *Inf Fusion* 4:259–280
14. S-h Zhao, Xue-zhi F, Kang G-d, Ramadan E (2002) Multi-source remote sensing image fusion based on support vector machine. *Chin Geogr Sci* 12:244–248
15. Mamatha SG, Rahim SA, Raj CP (2012) Feature-level multi-focus image fusion using neural network and image enhancement. *Global J Comput Sci Technol Graph Vis* 12(10):16–23
16. Pagidimarry M, Babu KA (2011) An all approach for multi-focus image fusion using neural network. *Int J Comput Sci Telecommun* 2(8):23–29
17. Li S, Kwok JT, Wang Y (2002) Multifocus image fusion using artificial neural networks. *Pattern Recogn Lett* 23:985–997
18. Li S, Kwok JT, Wang Y (2004) Fusing images with multiple focuses using support vector machines. *IEEE Trans Neural Netw* 15(6):1555–1561
19. Wang X-Y, Zhang B-B, Yang H-Y (2012) Active SVM-based relevance feedback using multiple Classifiers ensemble and features reweighting. *Eng Appl Artif Intell* 26(2013):368–381
20. Ali S, Majid A, Khan A (2014) IDM-PhyChm-Ens: intelligent decision-making ensemble methodology for classification of human breast cancer using physicochemical properties of amino acids. *Amino Acids* 46(4):977–993
21. Breiman L (1996) Bagging predictors. *Mach Learn* 26:123–140
22. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
23. Jiwu H, Shi YQ, Xianhua D (1999) A segmentation-based image coding algorithm using the features of human vision system. *J Image Graph* 4(5):400–404
24. Eskicioglu AM, Fisher PS (1993) Image quality measures and their performance. *IEEE Trans Commun* 43(12):2959–2965
25. Canny J (1986) A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell* 8(6):679–698
26. Zhang Y, Ge L (2009) Efficient fusion scheme for multi-focus images by using blurring measure. *Digit Signal Proc* 19(2):186–193
27. Naidu VPS, Raol JR (2008) Pixel-level Image fusion using wavelets and principal component analysis. *Defen Sci J* 58(3):338–352
28. Klonus S, Ehlers M (2009) Performance of evaluation methods in image fusion. In: 12th International conference on information fusion, Seattle, WA, USA, July 6–9 2009
29. Heng C, LI Jie, Weile Z (2006) A novel support vector machine-based multifocus image fusion algorithm. In: International conference on communications, circuits and systems proceedings, 2006, pp 500–504