Atencio-Ortiz, Pedro; Sanchez-Torres, German; Branch-Bedoya, John William
Evaluating supervised learning approaches for spatial-domain multi-focus image fusion

Evaluating supervised learning approaches for spatial-domain multi-focus image fusion

Pedro Atencio-Ortiz [a], German Sanchez-Torres [b] & John William Branch-Bedoya [c]

[a] Facultad de Ingenierías, Instituto Tecnológico Metropolitano, Medellín, Colombia. pedroatencio@itm.edu.co
[b] Facultad de Ingenierías, Universidad del Magdalena, Santa Marta, Colombia. gsanchez@unimagdalena.edu.co
[c] Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. jwbranch@unal.edu.co

Abstract

Image fusion is the generation of an image f that combines the most relevant information from a set of images of the same scene, acquired with different cameras or camera settings. Multi-Focus Image Fusion (MFIF) aims to generate an image f_e with extended depth-of-field from a set of images taken at different focal distances or focal planes, and it proposes a solution to the typical limited depth-of-field problem in an optical system configuration. A broad variety of works presented in the literature address this problem. The primary approaches found there are domain transformations and block-of-pixels analysis. In this work, we evaluate different systems of supervised machine learning applied to MFIF, including k-nearest neighbors, linear discriminant analysis, neural networks, and support vector machines. We started from two images at different focal distances and divided them into rectangular regions. The main objective of the machine-learning-based classification system is to choose the parts of both images that must be in the fused image in order to obtain a completely focused image. For focus quantification, we used the most popular metrics proposed in the literature, such as: Laplacian energy, sum-modified Laplacian, and gradient energy, among others. The evaluation of the proposed method considered classifier testing and fusion quality metrics commonly used in research, such as visual information fidelity and mutual information feature. Our results strongly suggest that the automatic classification concept satisfactorily addresses the MFIF problem.

Keywords: Multi-focus image fusion; image processing; supervised learning; machine learning.

Evaluando aproximaciones basadas en aprendizaje supervisado para la fusión en el dominio espacial de imágenes multi-foco

Resumen

La fusión de imágenes genera una imagen f que combina las características más relevantes de un conjunto de imágenes de la misma escena adquiridas con diferentes cámaras o configuraciones. La Fusión de Imágenes Multifoco (MFIF) parte de un conjunto de imágenes con diferente distancia focal para generar una imagen f_e con una profundidad de campo extendida. Lo que constituye una solución al problema de la profundidad de campo limitada en la configuración de un sistema óptico. La literatura muestra una amplia variedad de trabajos que abordan este problema. Las transformaciones de dominios y el análisis de bloques de píxeles son la base de los principales enfoques propuestos. En este trabajo se presenta una evaluación de diferentes sistemas de aprendizaje supervisado aplicados a MFIF, incluyendo k-vecinos más cercanos, análisis discriminante lineal, redes neuronales y máquinas de soporte vectorial. El método inicia con dos imágenes de la misma escena, pero con diferentes distancias focales que se dividen en regiones rectangulares. El objetivo principal del sistema de clasificación, que está basado en aprendizaje de máquina, es elegir las partes de ambas imágenes que deben estar en la imagen fusionada para obtener una imagen completamente enfocada. Para la cuantificación del enfoque se utilizaron las métricas más populares propuestas en la literatura como: la Energía Laplaciana, el Laplaciano Modificado por Suma y el Gradiente de Energía, entre otras. La evaluación del método propuesto incluye la fase de prueba de los clasificadores y las métricas de calidad de fusión utilizadas comúnmente en la investigación, tales como la fidelidad de la información visual y la característica de información mutua. Los resultados muestran que el concepto de clasificación automática puede abordar satisfactoriamente el problema MFIF.

Palabras clave: Fusión de imágenes mutifoco; procesamiento de imágenes; aprendizaje supervisado; aprendizaje de máquina.

1. Introduction

Traditional optical systems are limited by focus range, which means that not all objects in a scene appear clearly defined [1], so only the objects within the field of depth of the camera are focused and are perceived clearly, while the rest of the scene is blurred [2]. Thus, to generate an image with all objects adequately focused, the academic community used the idea of forming a synthetic image by fusing a set of images of the same scene. This is called Multi-Focus Image Fusion (MFIF). The MFIF process consists of merging multiple images with different focal planes to generate an image in which all objects appear sharp [3] without the introduction of any artifacts.

The application of MFIF is widespread and used to solve problems related to three-dimensional reconstruction [2,4], mobile image processing, microscopic imaging [5], and computer vision [6], among others.

To generate an extended-focus image from a sequence of partially focused images of the same scene, digital image descriptors called focus measures [7] are used in different approaches.

MFIF is generally divided into two major types of methods [1,8-9]:

- spatial-domain image fusion methods;
- frequency-domain image fusion methods.

The spatial-domain methods use measurable characteristics of spatial information of image pixels. They estimate these measures pixel by pixel or by using pixel sets [10,11]. The principal advantages of these methods are that they are easy to implement and require low computational complexity. However, they do require rich texture information in an image to generate good results and are usually weak in smooth image regions. Furthermore, grouping of pixels presents difficulties related to the correct determination of group size for quality maximization and to the presence of artifacts generated on border blocks [3].

Various approaches within the special-domain methods of MFIF have been applied. For example, K.L. Hua et al. [11] used random walks on graphs created from several feature sets of focus measures and color consistency to model local and global characteristics. This approach estimated the measures in each pixel locally for each input image and used them to maximize the global focus score and color consistency. On the other hand, a method based on sparse feature matrix decomposition using morphological filtering to extract salient features of original input images, was proposed in [12]. They used a pixel-wise methodology to fuse each sparse feature matrix estimation based on morphological filtering to generate the fused image.

Some works use hierarchical structures called QuadTree for recursively partitioning the pixel space of the image and decompose the input images into blocks of variable size [13,14]. Using a focus measure based on the sum-modified-Laplacian (SML), the method detects the focused regions. So, the resulting image is generated using the focused regions of input images. The main problem with partitioned methods is determination of the correct block size. Avoiding defocused regions in large regions or small blocks with low-contrast variance is the main challenge of QuadTree [14]. Similarly, the accuracy of the partitioning depends on searching deep in the tree, which has a direct impact on the computational cost. Other graph-based works are found in [15,16].

Another approach for dealing with the selection of focused regions in images is the segmentation approach. S. Li et al. [17] proposed a method that uses morphological filtering for rough segmentation of the images based on an initially estimated focus map. The method then uses the *image matting* technique to refine the segmentation results and a merging process to generate the final image. Another segmentation approach was proposed by M. Nejati et al. [18], based on a training and a testing phase. The training phase constructs a dictionary using focus information maps from local patches of source images. Each pixel from each input image is classified as in focus or not in focus. The final image results from pixel fusion according to a decision map that indicates which source image must be used to obtain the pixel intensity value. Overall, although the segmentation/optimization-based methods applied to the problem of image fusion generate adequate results, they involve a high computational cost.

Frequency-domain image fusion methods transform input images into a frequency domain representation where they are combined. One approach within this category is based on multi-scale decomposition. This is the most commonly reported approach to MFIF. For example, a shift and rotation invariant pyramid representation called Steerable Pyramid was applied by Z. Lin et al. [22]. Other frequency representations reported are Discrete Wavelet Transform [7,23] and Robust Principal Component Analysis [6]. Frequency-domain image fusion methods can be applied to multi-focus and multi-modal images with acceptable behavior. However, they add some noise level and cannot guarantee fidelity of the input image in the final image.

The MFIF problem can be stated as a classification problem where the classifier must decide in which source image the pixel or region has a high focus measure. J. Saeedi and K. Faez proposed a wavelet-based MFIF method which used a two-class Fisher classifier to group the regions into focused and defocused ones [19]. To reduce the number of misclassified regions due to uncertainty, they included a fuzzy logic algorithm. On the other hand, some works reported the use of other types of classifiers, such as neural networks, to tackle the MFIF problem [20,21].

The aim of this study is to compare various classification approaches to the MFIF problem. We selected four popular focus measures and proposed a new one based on morphological features; then we used these as classifier inputs.

This paper is organized as follows. In Section 2, we present the methodology used in this study, including the data and vector feature selection, training, and fusion stages. Section 3 presents the results obtained and comparisons. Finally, Section 4 concludes with a summary and future works.

2. Methodology

We establish the MFIF as a classification problem where the aim is to process a pair of input images, labeling their
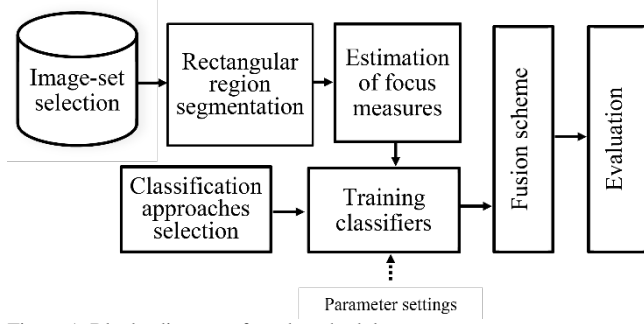
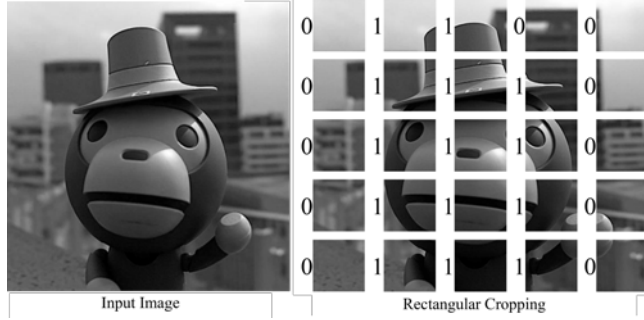Figure 1. Blocks diagram of used methodology.
Source: The authors.

Figure 2. The scheme used to form the subset of images.
Source: The authors.

regions as focused and defocused, in order to build up a final image that merges the focused regions. The methodology begins by selecting the image set used for the training and testing stages. After the image set was defined, we selected focus measures reported in the literature and machine-learning-based classifiers. The training process and the image fusion were the next methodological stages. Finally, we carried out an evaluation step to compare the behaviors of the selected classifiers. Figure 1 shows a block diagram of the methodology used.

2.1. Image-set selection and rectangular segmentation

Our initial set of images consisted of 30 pairs of multifocus images taken from public datasets [18,24] and a set of images of our own acquired in the laboratory. Every pair consisted of two images: a near-focused one and a far-focused one. Then, using rectangular cropping, a subset of 830 images was constructed and associated with a binary tag (*focused:1, defocused:0*) by human judgment. Figure 2 illustrates this process.

2.2. Focus measures and feature vector formation

In the MFIF context, a focus image operator is a local metric that quantifies the quality of focus in an image region. Ideally, when the region is perfectly focused, these operators must generate a maximum value [25] that decreases in a similar way when the image becomes blurred. Many focus metrics have been proposed by the scientific community. A typical focus metric should satisfy the following requirements [7]:

- independence of image content
- monotonicity of blur
- unimodality (only one maximum value)
- value variation according to degree of blurring
- minimal computation complexity
- robustness to noise

We selected four focus measures frequently used in reported works [7] and proposed a new focus measure based on morphological features.

Let us consider $f(x, y)$, the intensity level of a pixel (x, y). The selected focus measures are defined below:

2.2.1. Energy of Laplacian

The Energy of Laplacian (EOL) of an image f is computed as:

$$EOL = \sum_x \sum_y (f_{xx} + f_{yy})^2 \qquad (1)$$

Where

$$
\begin{aligned}
f_{xx} + f_{yy} = &-f(x-1, y-1) - 4f(x-1, y) \\
&- f(x-1, y+1) \\
&- 4f(x, y-1) + 20f(x, y) \\
&- 4f(x, y+1) \\
&- f(x+1, y-1) \\
&- 4f(x+1, y) \\
&- f(x+1, y+1)
\end{aligned} \qquad (2)
$$

2.2.2. Sum-modified Laplacian

The modified Laplacian is a proposal to avoid the cancellation trend of the second derivate in the EOL basic definition [25]. So, the SML is defined by equations (3) and (4):

$$
\begin{aligned}
\nabla^2{}_{ML} f(x, y) = &|2f(x, y) - f(x-\beta, y) \\
&- f(x+\beta, y)| \\
&+ |2f(x, y) - f(x, y-\beta) \\
&- f(x, y+\beta)|
\end{aligned} \qquad (3)
$$

where β is a spacing parameter addressing the accommodation of texture variation in the image and set to $\beta = 1$.

$$SML = \sum_{i=x-N}^{i=x+N} \sum_{j=x-N}^{j=x+N} \nabla^2{}_{ML} f(i, j) \, for \, \nabla^2{}_{ML} f(i, j) \geq T \qquad (4)$$

where T is a discrimination threshold value and N is the window size.

2.2.3. Energy of image gradient

This operator is based on the concept of determining the local high-frequency variations. The Energy of Gradient (EOG) can be computed as:

$$EOG = \sum_x \sum_y (f_x^2 + f_y^2) \qquad (5)$$

Where

$$f_x = f(x+1,y) - f(x,y) \qquad (6)$$

$$f_y = f(x,y+1) - f(x,y) \qquad (7)$$

2.2.4. Spatial Frequency

Spatial frequency (*SF*) is a modification of the energy of image gradient operator. It is defined as:

$$SF = \sqrt{RF^2 + CF^2} \qquad (8)$$

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

$$RF = \sqrt{\frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=2}^{N} [f(x,y) - f(x,y-1)]^2} \qquad (9)$$

$$CF = \sqrt{\frac{1}{M \times N} \sum_{x=2}^{M} \sum_{y=1}^{N} [f(x,y) - f(x-1,y)]^2} \qquad (10)$$

2.2.5. Energy of morphological features

We propose a new focus measure based on the metric defined in [17] and the concept of the energy of the image. We named this focus measure as the *Energy of Morphological Features* (*EMF*). We used a combination of bottom-hat and top-hat operations to first extract salient local features and then calculate the sum of maximum values for a window.

$$EMF(W_{M \times N}^I) = \frac{1}{M \times N} \sum_{i=1}^{N} \sum_{j=1}^{M} max(T_{i,j}, B_{i,j})^2 \qquad (11)$$

Let *T* and *B* be top-hat and bottom-hat morphological operations, respectively, centered at the pixel with coordinates (*i,j*). We describe the *EMF* of a local window *W* of an input image *I*. Finally, we build up a feature vector $\vec{\gamma}$ composed of the ratio of each focus measure of input images A and B. Figure 3 shows the structure of the feature vector.

$\frac{Eol_A}{Eol_B}$	$\frac{Sml_A}{Sml_B}$	$\frac{Eog_A}{Eog_B}$	$\frac{Sf_A}{Sf_B}$	$\frac{Emf_A}{Emf_B}$

Figure 3. Vector feature structure.
Source: The authors.

2.3. Classifier selection

Because the MFIF has been established as a classification problem, we selected the most representative classification approaches in the literature: Linear Discriminant Analysis (LDA) [26], naïve Bayes [27-29], k-nearest neighbors (k-NN) [30], random forest, multilayer perceptron (MLP) [31], and support vector machine (SVM) [32]. The two classes for region classification were focused and defocused regions, represented as a binary tag.

2.4. Training procedure

We created a binary classification dataset using the previously created subset of images with binary tags and the feature vector formation scheme described in Section 2.2.

With this dataset, each binary classifier was trained and tested using cross-validation. Then, the best score classifier of each classification technique was saved for the posterior fusion scheme.

For this purpose, the Python with Scikit-learn [33] and PyBrain [34] libraries were used to configure, train, and test different classifiers.

2.5. Fusion scheme

The fusion scheme proposed in this work is composed of two main stages. In the first stage, a binary mask is generated with information about which regions of both images have high focus (Fig. 4). In the second stage, the fused image is obtained using input images *A* and *B* and the binary mask obtained in the first stage (Fig. 5).

2.5.1. Stage 1 – Binary mask generation

The first stage begins by moving a rectangular window over a pair of input images and iteratively describing each window as a feature vector. Then, the ratio $\vec{\gamma}$ of these two vectors (Fig. 3) is used as an input for a trained classifier which returns a binary label (*focused:1, defocused:0*) for the highest focus in image *A* or *B* respectively. This binary label is used to build a binary mask *Z* of resolution equal to the moving window size. This mask contains information about the highly focused regions in both input images.

$$Z \binom{x_0: x_1,}{y_0: y_1} = C(\vec{\gamma}_{x_0:x_1,y_0:y_1}) \qquad (12)$$

Equation (21) describes stage 1, where *Z* is the binary mask matrix, the range $(x_0: x_1, y_0: y_1)$ is the image portion of the sliding window, *C* is the binary classifier, and $\vec{\gamma}_{x_0:x_1,y_0:y_1}$ is the ratio of the feature vector of input images *A* and *B* for the sliding window in the range $(x_0: x_1, y_0: y_1)$.

A scheme of the generation of the binary mask *Z* in stage 1 is shown in Figure 4.

2.5.2. Stage 2 – Image fusion

It is possible to generate a fused image using the binary mask obtained in stage 1, but because of the rectangular

nature of the moving window, a high number of artifacts can be generated in regions containing borders of objects (see the rough behavior of borders and holes in Z in Figure 4). An improved result can be obtained by smoothing the binary mask.

This stage begins by applying the morphological operation *close* to group individual pixels and fill holes in the binary mask. Then a low-pass filter (*median blur*) is applied to delete random noise and smooth the borders of the binary mask. Both operations used rectangular kernels of the same size. Finally, using the smoothed binary mask Z', input images A and B are cropped to generate the fused image F. This process is illustrated in Figure 5.

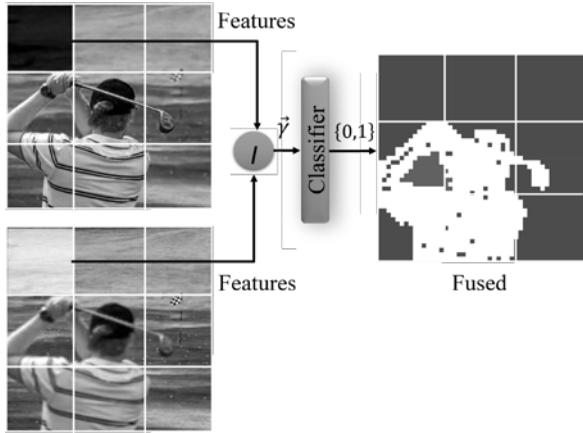$$F = A \times Z'_A + B \times Z'_B \tag{13}$$

Figure 4. Scheme of the generation of binary mask Z in stage 1.
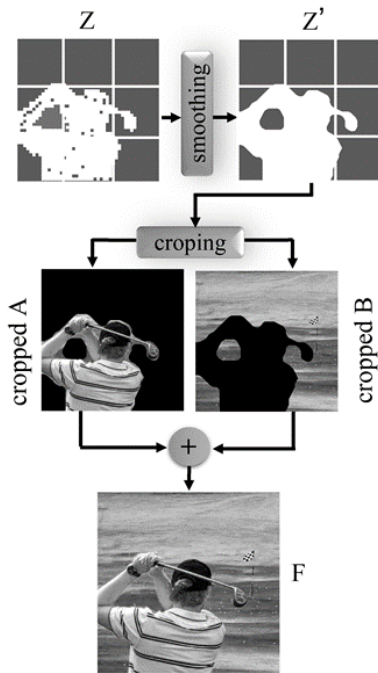Source: The authors.

Figure 5. Scheme of stage 2 for the generation of the fused image F.
Source: The authors.

Equation (22) describes stage 2, where F is the image fused using the proposed scheme, A and B are the input images, Z'_A and Z'_A are the pixels of the smoothed mask belonging to images A and B respectively, and $\times$ is an element-wise multiplication of two matrices.

2.6. Fusion quality measure

The fusion quality metric reflects the quality of visual information of a fused image obtained from a set of input images [35]. Evaluation of image fusion algorithms has become an important issue due to the different complexity characteristics used by several proposed approaches. Typically, the way in which the quality of the fused images is measured is by the mean of experts who score them. This approach does not offer a general way to evaluate approaches automatically and implies a costly effort. However, for objective evaluation of fusion results, we use three fusion-quality metrics including Visual Information Fidelity for Fusion (VIFF) [36], Petrovic's metric based on edge information ($Q^{AB/F}$) [35], and Feature Mutual Information (FMI) [37].

2.6.1. Visual information fidelity for fusion

VIFF is founded on the Visual Information Fidelity (VIF) quality metric based on Natural Scene Statistics theory, which measures the visual information by computing mutual information between different models estimated from images. These models are in the wavelet domain and include Gaussian Scale Mixture, the distortion model, and the Human Visual System.

The common procedure for VIF estimation is to divide the images into k sub-bands, each of which is divided in turn into b blocks. The mutual information between the different models is estimated and VIF can be stated as:

$$VIF(I_r, I_d) = \frac{\sum_k \sum_b \left(1 + \frac{g_{k,b}^2 \cdot \left(\sigma_{k,b}^r\right)^2}{\left(\left(\sigma_{k,b}^d\right)^2 - g_{k,b}^2 \cdot \left(\sigma_{k,b}^r\right)^2 + \sigma_N^2\right)}\right)}{\sum_k \sum_b \left(1 + \frac{\left(\sigma_{k,b}^r\right)^2}{\sigma_N^2}\right)} \tag{14}$$

2.6.2. Edge-based fusion performance

This metric is based on the idea of quantifying the important information preserved in the fused image. The important information is associated with edge information, and therefore the metric measures the amount of edge information transferred from input images to the fused image, using a Sobel operator to get the relative edge strength and orientation between the input and fused images [35].

Let A be an input image and let the pixel $A(i,j)$ have an edge strength $g_A(i,j)$ and orientation $\alpha_A(i,j)$ defined as:

$$g_A(i,j) = \sqrt{S_A^x(i,j)^2 + S_A^y(i,j)^2} \tag{15}$$

$$\alpha_A(i,j) = \tan^{-1}\left(\frac{S_A^x(i,j)}{S_A^y(i,j)}\right) \tag{16}$$

where $S_A^x(i,j)$ and $S_A^y(i,j)$ are the result of applying a horizontal and vertical Sobel template centered on $A(i,j)$. The relative edge strength (G^{AF}) and orientation (Δ^{AF}) of the image A with respect to a fused image are defined as:

$$G^{AF}(i,j) = \begin{cases} \dfrac{g_F(i,j)}{g_A(i,j)}, & g_A(i,j) > g_F(i,j) \\ \dfrac{g_A(i,j)}{g_F(i,j)}, & Otherwise \end{cases} \tag{17}$$

$$\Delta^{AF}(i,j) = 1 - \frac{|\alpha_A(i,j) - \alpha_F(i,j)|}{\frac{\pi}{2}} \tag{18}$$

The edge strength and orientation preservation values can be derived by:

$$Q_g^{AF}(i,j) = \frac{\Gamma_g}{1 + e^{k_g(G^{AF}(i,j) - \sigma_g)}} \tag{19}$$

$$Q_\alpha^{AF}(i,j) = \frac{\Gamma_\alpha}{1 + e^{k_g(\Delta^{AF}(i,j) - \sigma_\alpha)}} \tag{20}$$

where the constants Γ_g, k_g, σ_g, Γ_α, k_α, and σ_α determine the shape of sigmoid functions used to form the edge strength and orientation preservation value [35].

The edge preservation value is defined as:

$$Q^{AF}(i,j) = Q_g^{AF}(i,j)Q_\alpha^{AF}(i,j) \tag{21}$$

Finally, for input images A and B, the final weighted performance measure $Q^{AB/F}$ with respect to the fused image F is estimated as:

$$Q^{AB/F} = \frac{\sum_{n=1}^N \sum_{m=1}^M [Q^{AF}(i,j)w^A(i,j) + Q^{BF}(i,j)w^B(i,j)]}{\sum_{n=1}^N \sum_{m=1}^M (w^A(i,j) + w^B(i,j))} \tag{22}$$

2.6.3. Feature mutual information

Mutual Information (MI) is derived from information theory and quantifies the amount of information obtained about one variable from another variable. Thus, FMI quantifies the amount of image features transferred from the source images into the fused image. Gradient maps represent the images' features because they contain information about edges, directions, texture, contrast, and pixel neighborhoods [37]. Like classic approaches used to estimate MI, FMI estimation is based on the calculation of the joint probability distribution functions. Assuming the intensity pixels of the fused image $F(x,y)$ and input images $A(z,w)$ and $B(z,w)$, the methods use the normalized values of gradient magnitude image features, like marginal distributions, and therefore the amount of feature information in the fused image F from input images A and B is given by [37]:

$$I_{FA} = \sum_{f,a} P_{FA}(x,y,z,w) \log_2 \frac{p_{FA}(x,y,z,w)}{p_F(x,y) \cdot p_A(z,w)} \tag{23}$$

$$I_{FB} = \sum_{f,a} P_{FB}(x,y,z,w) \log_2 \frac{p_{FB}(x,y,z,w)}{p_F(x,y) \cdot p_B(z,w)} \tag{24}$$

where P_{FA} and P_{FB} are the joint distribution between the fused image and each input image. p_F, p_A, and p_B are the marginal distributions. The FMI is defined as:

$$FMI_F^{AB} = I_{FA} + I_{FB} \tag{25}$$

Based on [37-39], the normalized FMI can be obtained as:

$$FMI_F^{AB} = \frac{I_{FA}}{H_F + H_A} + \frac{I_{FB}}{H_F + H_B} \tag{26}$$

where H_F, H_A, and H_B are the histogram-based entropies of the images A, B, and F.

3. Experiments and results

In this section, we present detailed experimental settings used to evaluate the performance of the classification approach for MFIF from two perspectives: classifier performance and fusion quality. The first evaluates the learning ability of a classifier to decide whether or not an image is focused. The second measures the fusion result generated by a classifier of the information contained in the input images. Finally, we discuss the results exhibited.

3.1. Classifier performance

We selected five binary classifiers and trained them with the parameters shown in Table 1 and a dataset of 830 images labeled as binary for focused or defocused cases.

Training parameters were obtained experimentally, selecting the set of parameters that allowed us to obtain the best results.

Table 1.
Parameters used for training binary classifiers

Binary Classifier	Parameters	Technology
LDA	N/A	Scikit-learn
k-NN	$k = 3$	Scikit-learn
Gaussian Naïve Bayes	N/A	Scikit-learn
RBF-SVM	Gamma = 0.5; C = 1; Kernel = rbf	Scikit-learn
FFN-MLP	HL neurons = 50; HL activation function: $Tanh$; OL activation function: softmax	PyBrain

HL: Hidden Layer; OL: Output Layer
Source: The authors

For each classifier, we progressively changed the dataset size and used the k-fold cross-validation technique [40] with $k = 10$ in each iteration, to measure the learning and test precision/error. This allowed us to evaluate the performance variation (standard deviation) of each classifier. Table 2 shows the mean and standard deviation of precision of training and testing for each classifier. The highest scores are obtained by k-NN for training precision and by MLP for test precision. In contrast, the lowest precision is obtained by using Naïve Bayes in both the training and test stages.
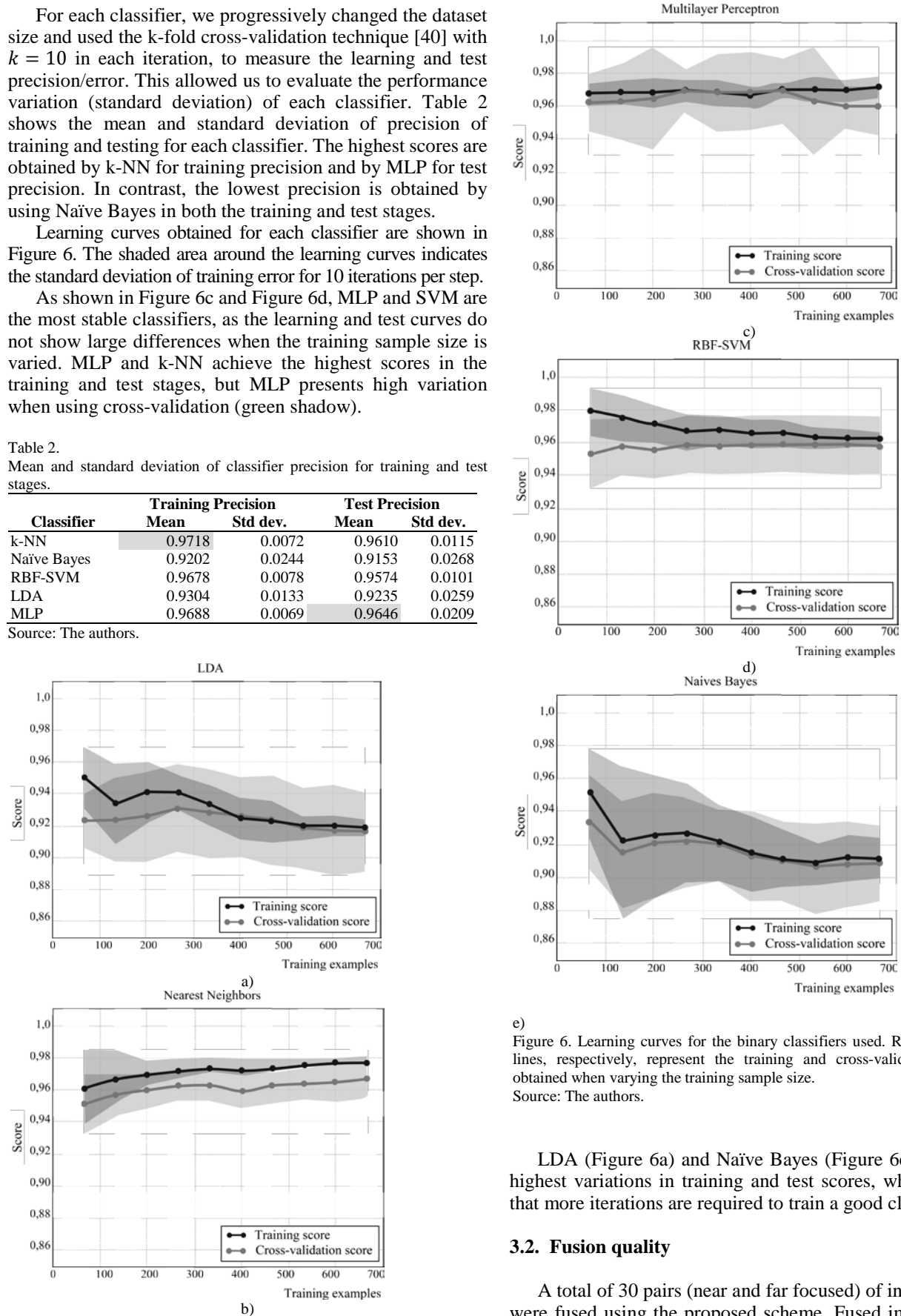
Learning curves obtained for each classifier are shown in Figure 6. The shaded area around the learning curves indicates the standard deviation of training error for 10 iterations per step.

As shown in Figure 6c and Figure 6d, MLP and SVM are the most stable classifiers, as the learning and test curves do not show large differences when the training sample size is varied. MLP and k-NN achieve the highest scores in the training and test stages, but MLP presents high variation when using cross-validation (green shadow).

Table 2.
Mean and standard deviation of classifier precision for training and test stages.

Classifier	Training Precision		Test Precision	
	Mean	Std dev.	Mean	Std dev.
k-NN	0.9718	0.0072	0.9610	0.0115
Naïve Bayes	0.9202	0.0244	0.9153	0.0268
RBF-SVM	0.9678	0.0078	0.9574	0.0101
LDA	0.9304	0.0133	0.9235	0.0259
MLP	0.9688	0.0069	0.9646	0.0209

Source: The authors.

a)

b)

c)

d)

e)
Figure 6. Learning curves for the binary classifiers used. Red and green lines, respectively, represent the training and cross-validation scores obtained when varying the training sample size.
Source: The authors.

LDA (Figure 6a) and Naïve Bayes (Figure 6e) have the highest variations in training and test scores, which means that more iterations are required to train a good classifier.

3.2. Fusion quality

A total of 30 pairs (near and far focused) of input images were fused using the proposed scheme. Fused images were

evaluated using the fusion quality metrics detailed in Section 2.6. The mean fusion quality for each classifier over 30 cases is shown in Table 3.

The Naïve Bayes classifier obtained the highest quality score for the VIFF and Qabf metrics, and MLP obtained the highest quality score for the FMI metric among all classifiers.

The high mean fusion quality achieved by MLP is expected because this classifier obtained the highest test score, as shown in Table 2. On the contrary, Naïve Bayes achieved the lowest training and test scores (Table 2) out of

all classifiers, but the highest fusion quality according to two out of three metrics among all classifiers (Table 3).

Some examples of the fusion results of the classifiers with the best results (Naive Bayes and MLP) are presented in Figure 7. Subtle differences in the masks generated from both classifiers can be observed. A qualitative judgment about this result is that MLP generates fewer artifacts than Naive Bayes.

Differences in fusion quality can be observed near the borders of objects, where artifacts are present due to the ambiguous nature of focus of the borders of objects in a natural scene (Fig. 8).

Figure 7. Results of image fusion scheme proposed for classifiers with the best fusion results: Naive Bayes and MLP. Column 1) input image 1; column 2) input image 2; column 3) Naive Bayes: near focus plane mask; column 4) MLP: near-focus plane mask.
Source: The authors.

4. Conclusions

The results obtained in the previous section show that the automatic classification concept can satisfactorily address the MFIF problem; that is, a classification scheme can be used to decide whether an image is focused or defocused based on local features. Thus, using this classification scheme, a fused image can be generated in which most regions or pixels are focused. The main contribution of this work is that it compares different classifiers in an MFIF scheme, thus evaluating which one obtains the best results in both learning and fusion stages. However, this work only evaluates the main classifiers found in the literature, and other classifiers not evaluated in this work may obtain better results.

Table 3.
Average fusion quality obtained with different classification methods

Quality Metric	FFN-MLP	LDA	Naïve Bayes	k-NN	RBF-SVM
VIFF	0.9258	0.9258	0.9259	0.9253	0.9246
Qabf	0.7572	0.7509	0.7573	0.7569	0.7564
FMI	0.6089	0.6083	0.6084	0.6083	0.6078

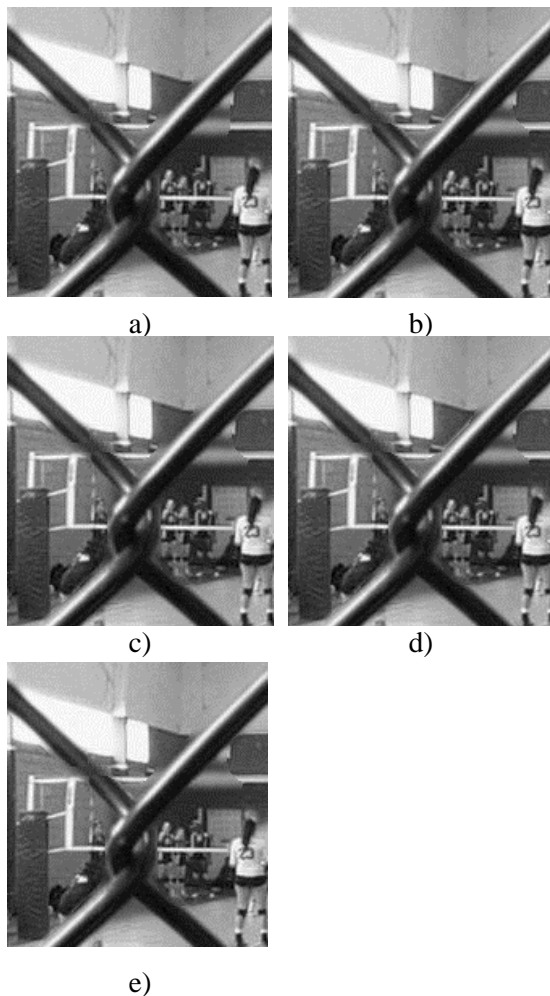Source: The authors.

a)

b)

c)

d)

e)

Figure 8. Magnified regions of fusion results obtained by a) FFN-MLP; b) k-NN; c) LDA; d) Naïve-Bayes; and e) RBF-SVM.
Source: The authors.

The most important conclusions obtained from this work are as follows:

- Classification precision does not guarantee fusion quality; that is, a classifier can have a low training and testing score but a high fusion quality.
- Naïve Bayes is an example of the latter. It is well known that Naïve Bayes is a robust classifier, and thus, subject to overfitting problems. Thus, it can perform well with small amounts of training data, which is the case found in this work. It is possible that this can be related to the number of artifacts a classifier generates and therefore the fusion quality.
- The MLP classifier shows good results in both the classification stage (training and test) and the fusion stage (fusion quality).
- Artifacts are generated near the borders of objects in the scene due to the spatial nature of the scheme used and the rectangular shape of the moving window. A good classifier generates fewer artifacts in these regions.
- MFIF is not a trivial problem, and a specific classifier can directly affect the fusion quality obtained using the proposed scheme in this work.
- Both the size of the moving window and kernels for morphological operations directly affect the results. Thus, future work should focus on evaluating the impact of these parameters on fusion quality obtained by this scheme.

Furthermore, the scheme proposed in this work could be adapted to address multi-focus scenes that present more than two (near and far) focal depths. An optimization technique based on region-growing could thus be used to generate a better smoothed mask Z' that encloses the objects that appear in the scene more precisely.

References

[1] Zhang, B., Lu, X., Pei, H., Liu, H., Zhao, Y. and Zhou, W., Multi-focus image fusion algorithm based on focused region extraction. Neurocomputing, 174, pp. 733-748, Jan. 2016. DOI: 10.1016/j.neucom.2015.09.092.

[2] Favaro, P., Mennucci, A. and Soatto, S., Observing shape from defocused images. Int. J. Comput. Vis., 52(1), pp. 25-43, 2003. DOI: 10.1023/A:1022366408068.

[3] Xiao, J., Liu, T., Zhang, Y., Zou, B., Lei, J. and Li, Q., Multi-focus image fusion based on depth extraction with inhomogeneous diffusion equation. Signal Process., 125, pp. 171-186, Aug. 2016. DOI: 10.1016/j.sigpro.2016.01.014.

[4] Saeed, A. and Choi, T.-S., A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise. Pattern Recognit., 41(6), pp. 2200-2225, 2008. DOI: 10.1016/j.patcog.2007.12.014.

[5] Song, Y., Li, M., Li, Q. and Sun, L., A new wavelet based multi-focus image fusion scheme and its application on optical microscopy, in: Robotics and Biomimetics, 2006. ROBIO '06. IEEE Int. Conf., 2006, pp. 401-405. DOI: 10.1109/ROBIO.2006.340210.

[6] Wan, T., Zhu, C. and Qin, Z., Multifocus image fusion based on robust principal component analysis. Pattern Recognit. Lett., 34(9), pp. 1001-1008, 2013. DOI: 10.1016/j.patrec.2013.03.003.

[7] Huang, W. and Jing, Z., Evaluation of focus measures in multi-focus image fusion. Pattern Recognit. Lett., 28(4), 2007. DOI: 10.1016/j.patrec.2006.09.005.

[8] Xin, W., You-Li, W., and Fu, L., A New Multi-source image sequence fusion algorithm based on SIDWT, in: 2013 Seventh International Conference on Image and Graphics, 2013, pp. 568-571. DOI: 10.1109/ICIG.2013.119.

[9] Yu, B. et al., Hybrid dual-tree complex wavelet transform and support vector machine for digital multi-focus image fusion, Neurocomputing, 182, pp. 1-9, Mar. 2016. DOI: 10.1016/j.neucom.2015.10.084.

[10] Haghighat, M.B.A., Aghagolzadeh, A. and Seyedarabi, H., Multi-focus image fusion for visual sensor networks in DCT domain. Comput. Electr. Eng., 37(5), pp. 789-797, Sep. 2011. DOI: 10.1016/j.compeleceng.2011.04.016.

[11] Hua, K.-L., Wang, H.-C., Rusdi, A.H. and Jiang, S.-Y., A novel multi-focus image fusion algorithm based on random walks, J. Vis. Commun. Image Represent., 25(5), pp. 951-962, Jul. 2014. DOI: 10.1016/j.jvcir.2014.02.009.

[12] Li, H., Li, L. and Zhang, J., Multi-focus image fusion based on sparse feature matrix decomposition and morphological filtering, Opt. Commun., 342, pp. 1-11, May 2015. DOI: 10.1016/j.optcom.2014.12.048.

[13] Bai, X., Zhang, Y., Zhou, F. and Xu, B., Quadtree-based multi-focus image fusion using a weighted focus-measure. Inf. Fusion, 22, pp. 105-118, Mar. 2015. DOI: 10.1016/j.inffus.2014.05.003.

[14] De, I. and Chanda, B., Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure, Inf. Fusion, 14, pp. 136-146, 2015. DOI: 10.1016/j.inffus.2012.01.007.

[15] Rajagopalan, A.N. and Chaudhuri, S., An MRF model-based approach to simultaneous recovery of depth and restoration from defocused images. IEEE Trans. Pattern Anal. Mach. Intell., 21(7), pp. 577-589, Jul. 1999. DOI: 10.1109/34.777369.

[16] Xu, N., Tan, K., Arora, H. and Ahuja, N., Generating omnifocus images using graph cuts and a new focus measure, in: Proc. 17th Int. Conf. Pattern Recognition, 2004. ICPR 2004, 4, pp. 697-700, 2004. DOI: 10.1109/ICPR.2004.1333868.

[17] Li, S., Kang, X., Hu, J. and Yang, B., Image matting for fusion of multi-focus images in dynamic scenes. Inf. Fusion, 14, pp. 147-162, 2013. DOI: 10.1016/j.inffus.2011.07.001.

[18] Nejati, M., Samavi, S. and Shirani, S., Multi-focus image fusion using dictionary-based sparse representation. Inf. Fusion, 25, pp. 72-84, 2015. DOI: 10.1016/j.inffus.2014.10.004.

[19] Saeedi, J. and Faez, K., Fisher classifier and fuzzy logic based multi-focus image fusion. In: Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE Int. Conf., pp. 420-425, 2009. DOI: 10.1109/ICICISYS.2009.5357648.

[20] Li, S., Kwok, J. and Wang, Y., Multifocus image fusion using artificial neural networks. Pattern Recognit. Lett., 23, pp. 985-997, 2002. DOI: 10.1016/S0167-8655(02)00029-6.

[21] Wang, Z., Ma, Y. and Gu, J., Multi-focus image fusion using PCNN. Pattern Recognit., 43(6), pp. 2003-2016, Jun. 2010. DOI: 10.1016/j.patcog.2010.01.011.

[22] Liu, Z., Tsukada, K., Hanasaki, K., Ho, Y.K. and Dai, Y.P., Image fusion by using steerable pyramid. Pattern Recognit. Lett., 22(9), pp. 929-939, 2001. DOI: 10.1016/S0167-8655(01)00047-2.

[23] Santhosh, J., Ketan, B. and Anand, S., Application of SiDWT with extended PCA for multi-focus images, in Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014. Int. Conf., pp. 55-59, 2014. DOI: 10.1109/MedCom.2014.7005975.

[24] Savic, S., Multifocus image fusion based on empirical mode decomposition, in: 20th Int. Electrotechnical and Computer Science Conf., 2011, pp. 91-94.

[25] Nayar, S.K. and Nakgawa, Y., Shape from focus, IEEE Trans. Pattern Anal. Mach. Intell., 16(8), pp. 824-831, 1994. DOI: 10.1109/34.308479.

[26] Balakrishnama, G., Linear discriminant analysis – A brief tutorial, 1998.

[27] Cheeseman, P. and Stutz, J., Bayesian classification (autoclass): Theory and results, in: Advances in knowledge discovery and data mining, Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R., Eds., Menlo Park, CA, USA: American Association for Artificial Intelligence, 1996, pp. 153-180.

[28] Tsangaratos, P. and Ilia, I., Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. CATENA, 145, pp. 164-79, Oct. 2016. DOI: 10.1016/j.catena.2016.06.004.

[29] Domingos, P. and Pazzani, M., Beyond independence: Conditions for the optimality of the simple Bayesian classifier, in: Machine Learning, 1996, pp. 105-112.

[30] Villa-Medina, J.L., Boqué, R. and Ferré, J., Bagged k-nearest neighbours classification with uncertainty in the variables. Anal. Chim. Acta, 646(1–2), pp. 62-68, Jul. 2009. DOI: 10.1016/j.aca.2009.05.016.

[31] Orhan, U., Hekim, M. and Ozer, M., EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. Expert Syst. Appl., 38(10), pp. 13475-13481, Sep. 2011. DOI: 10.1016/j.eswa.2011.04.149.

[32] Kong, Y.B., Lee, E.J., Hur, M.G., Park, J.H., Park, Y.D. and Yang, S.D., Support vector machine based fault detection approach for RFT-30 cyclotron. Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip., 834, pp. 143-148, Oct. 2016. DOI: 10.1016/j.nima.2016.07.054.

[33] Pedregosa, F. et al., Scikit-learn: machine learning in Python. J. Mach. Learn. Res., 12, pp. 2825-2830, 2011.

[34] Schaul, T. et al., PyBrain. J. Mach. Learn. Res., 11, pp. 743-746, 2010.

[35] Xydeas, C.S. and Petrovic, V., Objective image fusion performance measure. Electron. Lett., 36(4), pp. 308-309, Feb. 2000. DOI: 10.1049/el:20000267.

[36] Han, Y., Cai, Y., Cao, Y. and Xu, X., A new image fusion performance metric based on visual information fidelity. Inf. Fusion, 14(2), pp. 127-135, 2013. DOI: 10.1016/j.inffus.2011.08.002

[37] Haghighat, M.B.A., Aghagolzadeh, A. and Seyedarabi, H., A non-reference image fusion metric based on mutual information of image features, Comput. Electr. Eng., 37(5), pp. 744-756, Sep. 2011. DOI: 10.1016/j.compeleceng.2011.07.012.

[38] Kvalseth, T.O., Entropy and correlation: some comments. IEEE Trans. Syst. Man Cybern., 17(3), pp. 517-519, May 1987. DOI: 10.1109/TSMC.1987.4309069.

[39] Hossny, M., Nahavandi, S. and Creighton, D., Comments on 'Information measure for performance of image fusion'. Electron. Lett., 44(18), pp. 1066-1067, Aug. 2008. DOI: 10.1049/el:20081754.

[40] Duda, R., Hart, P. and Stork, D., Pattern classification, 2nd ed., 2001.

P.S. Atencio-Ortiz, received the BSc. Eng in Systems Engineering in 2009 from the Universidad del Magdalena, Colombia, the MSc. degree in Systems Engineering in 2012 from the Universidad Nacional de Colombia. Medellin, Colombia. Currently, he is a full professor in the Faculty of Engineering at Instituto Tecnológico Metropolitano and PhD student in systems engineering at Universidad Nacional de Colombia. Medellin, Colombia. His research interests include: digital image processing; computer vision; and computational intelligence techniques.
ORCID: 0000-0002-9473-400X

G. Sanchez-Torres, received a BSc. Eng in Systems Engineering in 2005 from the Universidad del Magdalena, Colombia, a MSc. degree in Systems Engineering in 2006, and a PhD degree in Systems Engineering in 2012, both from the Universidad Nacional de Colombia in Medellin, Colombia. Currently, he is a full professor in the Faculty of Engineering, Universidad del Magdalena, Santa Marta, Colombia. His research interests include: digital image processing; 3D surface reconstruction, computer vision; computational intelligence techniques; and computational simulation and modeling.
ORCID: 0000-0002-9069-0732

J.W. Branch-Bedoya, received a BSc. Eng in Mining and Metallurgy Engineering in 1995, a MSc degree in Systems Engineering in 1997, and a PhD degree in Systems Engineering in 2007, all of them from the Universidad Nacional de Colombia in Medellin, Colombia. Currently, he is a full professor in the Computing and Decision Sciences Department, Facultad de Minas, Universidad Nacional de Colombia. His research interests include: automation; computer vision; digital image processing; and computational intelligence techniques.
ORCID: 0000-0002-0378-028X