Block-based Feature-level Multi-focus Image Fusion

Abdul Basit Siddiqui, M. Arfan Jaffar
National University of Computer and Emerging Sciences
Islamabad, Pakistan
{basit.siddiqui,arfan.jaffar}@nu.edu.pk

Ayyaz Hussain, Anwar M. Mirza
National University of Computer and Emerging Sciences
Islamabad, Pakistan
{ayyaz.hussain,anwar.m.mirza}@nu.edu.pk

Abstract— In recent times, the applications of image processing have grown immensely. Usually due to limited depth of field of optical lenses especially with greater focal length, it becomes impossible to obtain an image where all the objects are in focus. Image fusion deals with creating an image in which all the objects are in focus. Thus it plays an important role to perform other tasks of image processing such as image segmentation, edge detection, stereo matching and image enhancement. In this paper, a novel feature-level multi-focus image fusion technique has been proposed which fuses multi-focus images using classification. Ten pairs of multi-focus images are first divided into blocks. The optimal block size for every image is found adaptively. The block feature vectors are fed to feed forward neural network. The trained neural network is then used to fuse any pair of multi-focus images. We have also presented the results of extensive experimentation performed to highlight the efficiency and utility of the proposed technique.

Keywords— Fusion, Optimal Block, NN

I. INTRODUCTION

Image fusion is a sub-field of image processing in which more than one images are fused to create an image where all the objects are in focus. Image fusion is of significant importance due to its application in medical science, forensic and defense departments. The process of image fusion is performed for multi-sensor and multi-focus images of the same scene. Multi-sensor images of the same scene are captured by different sensors whereas multi-focus images are captured by the same sensor. In multi-focus images, the objects in the scene which are closer to the camera are in focus and the farther objects get blurred. Contrary to it, when the farther objects are focused then closer objects get blurred in the image. To achieve an image where all the objects are in focus, the process of images fusion is performed either in spatial domain or in transformed domain. Spatial domain includes the techniques which directly incorporate the pixel values. In transformed domain, the images are first transformed into multiple levels of resolutions. An image often contains physically relevant features at many different scales or resolutions. Multi-scale or multi-resolution approaches provide a means to exploit this fact [1]. After applying certain operations on the transformed images, the fused image is created by taking the inverse transform.

Image fusion is generally performed at three different levels of information representation including pixel level, feature level and decision level [2]. In pixel-level image fusion, simple mathematical operations such as max (maximum) or mean (average) are applied on the pixel values of the source images to generate fused image. However these techniques usually smooth the sharp edges or leave the blurring effects in the fused image. In feature level multi-focus image fusion, the source images are first segmented into different regions and then the feature values of these regions are calculated. Using some fusion rule, the regions are selected to generate the fused image. In decision level image fusion, the objects in the source images are first detected and then by using some suitable fusion algorithm, the fused image is generated.

A number of image fusion techniques have been presented in the literature. In addition of simple pixel level image fusion techniques, we find the complex techniques such as Laplacian Pyramid [3], fusion based on PCA [4], discrete wavelet (DWT) based image fusion [5], Neural Network based image fusion [6] and advance DWT-based image fusion [7]. These techniques have different merits and demerits such as linear wavelets like Haar wavelet during the image decomposition does not preserve the original data [8]. Similarly due to low-pass filtering process of wavelets, the edges in the image become smooth and hence the contrast in the fused image is decreased.

In this paper, we have proposed a new method for multi-focus image fusion. The proposed method is discussed in section II. In section III, the quantitative measures used to evaluate the performance of the proposed method are described. Section IV covers the experimentation details and section V concludes the study.

II. PROPOSED METHOD

We used an image set of 10 different images to train our neural network. Every image is first divided into number of blocks. The block size plays an important role in distinguishing the blurred and un-blurred regions from each other. To accomplish this task, we run the algorithm for finding adaptive block size using Genetic Algorithm introduced in [9] for every image in the image set. Xinman Zhang et. al. [9] used a population of chromosomes where every chromosome represents the width and length of the block. After dividing the images into blocks, the feature values of every block of all the images are calculated and a features
file is created. A sufficient number of feature vectors are used to train the neural network. The trained neural network is then used to fuse any set of multi-focus images. Image dataset, Feature selection and the proposed algorithm are discussed in the following sections.

A. Creating Image Dataset

In the proposed method, we first created an image-set of 10 grayscale images. This image-set is shown in the Fig. 1. These images are taken from different image processing websites. The images are of different scenes and backgrounds. For every image in the set, we created its two versions of the same size. In the first version, some of the regions are randomly selected in the left half of the image and are blurred. A similar process is performed in the right half of the image in the second version. The blurred versions are generated by Gaussian blurring of radius 1.5. For the image set of 10 grayscale images, we created 20 versions of blurred images. For proposed method experimentation, we resized all the images into 480x640 resolutions.

![Image Set](image.png)

Fig. 1 Image set used to train neural network

B. Features Selection

In feature-level image fusion, the selection of different features is an important task. In multi-focus images, some of the objects are clear (in focus) and some objects are blurred (out of focus). The blurred objects in an image reduce its clearness. We have used five different features to characterize the information level contained in a specific portion of the image. This features set includes Variance, Energy of Gradient, Contrast Visibility, Spatial Frequency and Canny Edge information. From the Fig. 2, we observe how the blurriness of increasing Gaussian radius affects the clearness of the image. The values of the features in the image against blurriness of different degrees are given in Table 1.

We observe from Fig. 1 that with the increasing blurriness, the clearness of the image is reduced and the identification of different objects in the image becomes difficult. Features values given in Table 1 show the degradation of the original image. Especially the values of energy of gradient, spatial frequency and edge information are significantly reduced.

1) Contrast Visibility: It calculates the deviation of a block of pixels from the block’s mean value. Therefore it relates to the clearness level of the block. The visibility of the image block is obtained using equation (1).

\[
VI = \frac{1}{m \times n} \sum_{(i,j) \in B_k} \frac{|I(i,j) - \mu_k|}{\mu_k}
\]  

(1)

Here \(\mu_k\) and \(m \times n\) are the mean and size of the block \(B_k\) respectively.

2) Spatial Frequency: Spatial frequency measures the activity level in an image. It is used to calculate the frequency changes along rows and columns of the image. Spatial frequency is measured using equation (2).

\[
SF = \sqrt{(RF)^2 + (CF)^2}
\]  

(2)

Where

\[
RF = \sqrt{\frac{1}{m \times n} \sum_{j=2}^{n} \sum_{i=2}^{m} [I(i,j) - I(i,j-1)]^2}
\]

and

\[
CF = \sqrt{\frac{1}{m \times n} \sum_{i=2}^{m} \sum_{j=2}^{n} [I(i,j) - I(i-1,j)]^2}
\]

![Blurred Images](image.png)

Figure 2: BARM image (a) Original Image (b) blurred with radius 0.5 (c) blurred with radius 1.0 and (d) blurred with radius 1.5

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>FEATURE VALUES WITH INCREASING BLURRINESS IN BARM IMAGE SHOWN IN FIGURE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure</td>
<td>Variance</td>
</tr>
<tr>
<td>1 (a)</td>
<td>195.74</td>
</tr>
</tbody>
</table>

Identify applicable sponsor/s here. If no sponsors, delete this text box.

(sponsors)
Here $I$ is the image and $m \times n$ is the image size. A large value of spatial frequency describes the large information level in the image and therefore it measures the clearness of the image.

3) Variance: Variance is used to measure the extent of focus in an image block. It is calculated using equation (3)

$$\text{Variance} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (I(i, j) - \mu)^2$$  \hspace{1cm} (3)

Here $\mu$ is the mean value of the block image and $m \times n$ is the image size. A high value of variance shows the greater extent of focus in the image block.

4) Energy of Gradient (EOG): It is also used to measure the amount of focus in an image. It is calculated using equation (4).

$$\text{EOG} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} (f_i^2 + f_j^2)$$

where $f_i = f(i + 1, j) - f(i, j)$
and $f_j = f(i, j + 1) - f(i, j)$

Here $m$ and $n$ represent the dimensions of the image block. A high value of energy of gradient shows greater amount of focus in the image block.

5) Edge Information: The edge pixels can be found in the image block by using Canny edge detector. It returns 1 if the current pixel belongs to some edge in the image otherwise it returns 0. The edge feature is just the number of edge pixels contained within the image block.

C. Proposed Algorithm

Stepwise working of the proposed method is given under.

1. Find the optimal block size for each set of $LF_i$ and $RF_i$ using [9]. $LF_i$ is the left-focused and $RF_i$ is the right-focused versions of the $i$th image in the dataset discussed in section (II-A). Here $i = 1, 2, 3, ..., 10$.

2. Divide the versions $LF_i$ and $RF_i$ of every image in the dataset into $K$ number of blocks of size $M \times N$.

3. Create the features file for all $LF_{ij}$ and $RF_{ij}$ according to the features discussed in section (II-B). Here $j = 1, 2, 3, ..., K$. For all $i$, there are two sets of features values for every block $j$ named as $FSLF_{ij}$ and $FSRF_{ij}$ each of which contains five feature values. Subtract the features values of block $j$ of $LF_i$ from the corresponding feature values of block $j$ of $RF_i$ and include this pattern in features file. Normalize the feature values between [0 1].

4. Assign the class value to every block $j$ of $ith$ image. If block $j$ is visible in $LF_i$ then assign it class value 1 otherwise give it a class value -1. In case of class value -1, block $j$ is visible in $RF_i$.

5. Create a neural network with adequate number of layers and neurons. Train the newly created neural network with adequate number of patterns selected from features file created in step 2.

6. By using the trained neural network, identify the clearness of all the blocks of any pair of multi-focus images to be fused.

7. Fuse the given pair of multi-focus images block by block according to the classification results of the neural network such that

$$\begin{align*}
\text{Output of NN for block } j \\
\text{if } > 0, \text{ Select } j \text{ from left - focused image} \\
\text{if } < 0, \text{ Select } j \text{ from right - focused image}
\end{align*}$$

The block diagram of the proposed method is shown in Fig. 3.

III. Quantitative Measures

There are different quantitative measures which are used to evaluate the performance of the fusion techniques. We used three measures Root Mean Square Error (RMSE), Peak Signal to Noise Ratio (PSNR) and Mutual Information (MI) when the reference image is available. For blind image fusion (reference image is not available), we used Entropy, Standard Deviation (SD) and Spatial Frequency (SF). These measures are defined in Table 2.
**Fig. 3** Block diagram of the proposed method

**TABLE II.** PERFORMANCE METRICS USED FOR IMAGE FUSION

<table>
<thead>
<tr>
<th>Metric</th>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RMSE</strong></td>
<td>$RMSE = \sqrt{\frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} [R(i, j) - F(i, j)]^2}$</td>
<td>Calculates the deviation between the pixel values of reference image and fused image. A lesser value shows good fusion results. Here $R, F$ are the reference and fused images respectively. $mn$ is the image size.</td>
</tr>
<tr>
<td><strong>PSNR</strong></td>
<td>$PSNR = 20 \log_{10} \left( \frac{L}{\sqrt{\frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} [R(i, j) - F(i, j)]^2}} \right)$</td>
<td>Determines the degree of resemblance between reference and fused image. A bigger value shows good fusion results. $L$ denotes to number of gray level in the image.</td>
</tr>
<tr>
<td><strong>MI</strong></td>
<td>$MI = \sum_{i=1}^{m} \sum_{j=1}^{n} h_{RF}(i,j) \log \frac{h_{RF}(i,j)}{h_{R}(i,j)h_{F}(i,j)}$</td>
<td>Determines how much information the fused image retrieved from the input source images. A bigger value shows good fusion results. Here $h_{RF}, h_{R}, h_{F}$ are the joint, reference and fused images histograms respectively.</td>
</tr>
<tr>
<td><strong>Entropy</strong></td>
<td>$H = - \sum_{i=0}^{L-1} h_{F}(i) \log_{2} h_{F}(i)$</td>
<td>Quantifies the quantity of information contained in the fused image. A bigger value shows good fusion results. Here $h_{F}$ is the normalized histogram of fused image and $L$ is the number of gray levels.</td>
</tr>
</tbody>
</table>
Here $h_{F}$ is the normalized histogram of fused image and $L$ is the number of gray levels.

$$SD = \sqrt{\frac{1}{L} \sum_{i=0}^{L} (i - i')^2 h_{F}(i)}$$  where $i' = \sum_{i=0}^{L} i h_{F}$

Measures the contrast in the fused image. A well contrast image has high standard deviation.

IV. EXPERIMENTS AND RESULTS

Image fusion is performed in two different situations. In first situation, the reference image is available and in second situation, the reference image is not available (blind image fusion). We exercised the feed forward neural network with different number of hidden layers and with different number of neurons on each layer. We found the best results with one hidden layer of 30 neurons. The learning rate $\alpha$ and the threshold for mean square error are kept as 0.01. The results of the proposed technique are compared with different existing methods including DWT, aDWT, PCA and Laplacian Pyramid based image fusion techniques. The performance of an existing Probabilistic Neural Network based technique is also compared with the results of the proposed technique. The experimentation results are obtained when the reference image is available and when it is not available (blind image fusion). In order to evaluate the performance of the proposed technique, the results for four different pairs of multi-focus images are obtained including balloon, lab, clock and pepsi can images.

A. PNN-based Image Fusion and the Proposed Technique Variation

Shutao Li et. al. proposed a probabilistic neural network based technique to perform multi-focus image fusion. They trained a neural network to classify the selection of blocks of the two source images to generate the fused image. In their technique, the source images were divided into the blocks of fixed size of 32x32. The block size is an important factor to achieve good fusion results. Fixed size blocks may separate the blurred and un-blurred regions within one pair of multi-focus images but for some other pair of multi-focus images, the contents of the image block are partially blurred. The size of the block varies image to image because different images have different blurred regions. In the proposed method, for every pair of multi-focus images, an optimal block size is found using the technique given in [9]. We have used five different features to calculate the clearness of a block more accurately as compared to three features used by Shutao Li et. al.

A major difference between the proposed method and the existing PNN-based image fusion technique is the training of the neural network. Shutao Li et. al. in their technique create and train a new neural network for every pair of multi-focus images which is really time consuming. In the proposed method, we trained the neural network using the block features of ten different pair of multi-focus images. Once the classifier is obtained then it can be used to fuse any pair of multi-focus images.

B. Quantitative Assessments and the Visual Comparison

1) When the reference image is available: For balloon and lab images, the reference images are available. A visual comparison is shown in figures 4 and 5 for balloon and lab images. Both the balloon and lab images are of size 480x640.
TABLE III. RESULTS OF QUANTITATIVE MEASURES FOR BALLOON AND LAB IMAGES

<table>
<thead>
<tr>
<th>Method</th>
<th>Balloon</th>
<th>Lab</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>MI</td>
</tr>
<tr>
<td>DWT</td>
<td>5.1025</td>
<td>12.6004</td>
</tr>
<tr>
<td>aDWT</td>
<td>5.0781</td>
<td>12.6122</td>
</tr>
<tr>
<td>Laplacian</td>
<td>2.8222</td>
<td>16.2871</td>
</tr>
<tr>
<td>PNN</td>
<td>0.2634</td>
<td>19.3674</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.1967</td>
<td>24.3553</td>
</tr>
</tbody>
</table>

2) When the Reference Image is not available (Blind Image Fusion): We have used clock and pepsi can images in this category. The clock and pepsi can image sizes are 256x256 and 512x512 respectively. When the reference image is not available then the performance of the fusion process is evaluated on the basis of different set of quantitative measures. These measures are described in the section (3). Fig. 6 & 7 provide a visual comparison of the proposed technique with the existing techniques.

In case of pepsi can image, block effects are visible at the table edge in the fused image generated by PNN-based technique. In PNN-based technique, the source images are divided into parts using a block of fixed size and hence it leaves the block effects in the fused image. Table 4 provides some statistics to evaluate the performance of the proposed technique when the reference image is not available.
V. CONCLUSION

In this paper, a feature-level block-based multi-focus image fusion technique is proposed. A feed forward neural network is first trained with the block features of ten pairs of multi-focus images. A feature set including spatial frequency, contrast visibility, edges, variance and energy of gradient is used to define the clarity of the image block. Block size is determined adaptively for each image. The trained neural network is then used to fuse any pair of multi-focus images. Experimentation results show that the proposed technique performs better than the existing techniques.

By finding the block size adaptively, the blurred and unblurred regions within the source images are optimally identified. As a result of it, the proposed technique performs better. In the proposed technique, only one neural network is created whereas in PNN-based image fusion [6], neural network for every pair of multi-focus images is created which is really time consuming.

REFERENCES