A novel multi-focus image fusion algorithm based on feature extraction and wavelets

Rodrigo Nava\textsuperscript{a}, Boris Escalante-Ramírez\textsuperscript{b} and Gabriel Cristóbal\textsuperscript{c}

\textsuperscript{a}Posgrado en Ciencia e Ingeniería de la Computación, Universidad Nacional Autónoma de México, México
\textsuperscript{b}Dep. de Procesamiento de Señales, Facultad de Ingeniería, Universidad Nacional Autónoma de México, México
\textsuperscript{c}Instituto de Óptica “Daza de Valdés” (CSIC), Serrano 121, Madrid, 28006, Spain

ABSTRACT

Focusing cameras is an important problem in computer vision and microscopy. Due to the limited depth of field of optical lenses in CCD devices, there are sensors which cannot generate images of all objects with equal sharpness. Therefore, several images of the same scene have different focused parts. One way to overcome this problem is to take different in-focus parts and combine them into a single composite image which contains the entire focused scene. In this paper we present a multi-focus image fusion algorithm based on feature extraction and wavelets. Classical wavelet synthesis is known to produce Gibbs phenomenon around discontinuities. The approach of \textit{wavelet on the interval transform} is suitable to orthogonal wavelets and does not exhibit edge effects. Since Canny filter’s operator is a Gaussian derivative, a well known model of early vision, we used it to get salient edges and to build a decision map who determines which information to take and at what place. Finally, quality of fused images is assessed using both traditional and perceptual-based quality metrics. Quantitative and qualitative analysis of the results demonstrate higher performance of the algorithm compared to traditional methods.

Keywords: image fusion, wavelet transform, feature extraction, perceptual quality metrics

1. INTRODUCTION

At present time, image fusion is widely recognized as an important tool for improving performance in image-based applications such as remote sensing, machine vision, medical imaging, optical microscopy and so on. Image fusion allows merging images from multiple sensors or even multiple images from the same sensor. Its goal is to integrate complementary and redundant information to provide a composite image which could be used to better understanding of the entire scene.

Image fusion can be defined as the process by which two or more images or some of their features are combined together to form a single image more suitable for human and machine perception or further task such as segmentation, feature extraction and object recognition.\textsuperscript{1}

In this paper, we considered multi-focus images, i.e, situations where two or more images that depict the same scene will not be in-focus everywhere(if one object is in-focus, another one will be out of focus). This occurs because there are sensors which cannot generate images of all objects at various distances with equal sharpness. Another example of multi-focus images is found in optical microscopy. Limited depth of field is a common problem with conventional light microscopy. Since the specimen can be sectioned by moving the object along the optical axis, portions of the objects' surface outside the optical plane appear defocused in the acquired image.\textsuperscript{2} The advantages of multi-focus images can be fully exploited by merging the sharply focused regions into one image that will be in-focus everywhere.\textsuperscript{3,4}

\textsuperscript{1}Further author information: (Send correspondence to R.N.)
R.N.: E-mail: urielrnv@uxmcc2.iimas.unam.mx
B.E.: E-mail: boris@servidor.unam.mx, Telephone (52)55-56223213, Fax: (52)55-56223579
G.C.: E-mail: gabriel@optica.csic.es, Telephone (34)91-5616800(x2313), Fax: (34)91-5645557
In the previous literature, many techniques for multi-focus image fusion have been suggested. An overview on this issue can be found in.\textsuperscript{5,6} Some methods such as wavelet-based algorithms and the Hermite transform have shown better results for visual interpretation.\textsuperscript{7,8} Wavelet methods are widely employed, classical approach uses quadrature mirror filter banks (QMF) and Fourier transform (FT), but exhibits edge effects and artifacts due to aliasing in the fused image.\textsuperscript{9,10} In this work, we used a orthonormal wavelets on the interval construction in order to avoid some of disadvantages of the classical approach.\textsuperscript{11}

The existing multi-focus image fusion methods can be classified into several groups. A common classification is to distinguish three different levels according to the stage at which the fusion takes place: pixel, feature and decision level.\textsuperscript{12}

- Pixel level fusion generates a fused image in which information content associated with each pixel is determined from a set of pixels in source images. Fusion at this level can be performed either in spatial or in frequency domain. However, pixel level fusion may conduct to contrast reduction.\textsuperscript{13}

- Feature level fusion requires the extraction of salient features which are depending on their environment such as pixel intensities, edges or textures. This similar features from the input images are fused. This fusion level can be used as a means of creating additional composite features. The fused image can also be used for classification or detection.\textsuperscript{14}

- Decision level is a higher level of fusion. Input images are processed individually for information extraction. The obtained information is then combined applying decision rules to reinforce common interpretation.\textsuperscript{15}

An important preprocessing step in image fusion is image registration. Corresponding pixel positions in the source images must refers to the same location. In this paper, we assumed that source images are already registered.

In the next section a new feature level fusion algorithm is proposed. Firstly, multiresolution wavelet is presented and then the detailed algorithm is given. Section 3 presents some experimental results. Finally, conclusions are given in section 4.

2. MULTI-FOCUS IMAGE FUSION

First, we give a brief description of the fusion methodology based on the multiresolution analysis (MRA). An image often contains physically salient features at many different scales or resolutions. There is strong evidence that the human visual system (HVS) processes information in a MRA fashion and it is especially sensitive to contrast changes.\textsuperscript{4} At the beginning of analysis all scales or resolutions have the same importance. MRA scheme decomposes input images into several coefficients each of which captures information present at a given scale. One major advantage of MRA is that spatial as well as frequency domain localization of an image is obtained simultaneously. The basic idea in MRA is to decompose input images at first by applying wavelet transform. Then the fusion operation on the transformed images is performed and finally the fused image is reconstructed by inverse transform that restores intensity values and spatial resolution. See Fig. 1.

In addition to this, wavelet transform of an image provides a multiple scale pyramid decomposition. There are four coefficients after each decomposition. These are Low-Low, Low-High, High-Low and High-High coefficients. The next decomposition stage operates only on Low-Low coefficient. See Fig. 1(a).

General theory of MRA assumes that signals or images have infinite length. If the MRA analysis is applied to images with finite support, undesirable edge effects will occur. However, filtering operations require values of the images outside its supported range. Images can be extrapolated by means of periodic, symmetric or polynomial based methods, which allow filters to be applied on the images $\in L^2(\mathbb{R})$. Also, special filters can be designed to replace original filters at the image borders.\textsuperscript{11} These filters are adapted to interval boundaries and do not require values outside interval. In this work we have implemented a polynomial method for orthogonal wavelets that appears in.\textsuperscript{16}
2.1 Fusion Algorithm

The proposed algorithm is shown in Fig. 2. We only show a bi-modal case but can be easily extended to a multi-modal case. The algorithm consists of the following steps:

1. Decomposition of the input images using wavelet on the interval transform in one or more levels. According to wavelet theory, the bigger decomposition level is, the smaller low frequency information contains. Justification of the number of decomposition levels appears in.\textsuperscript{17}

2. Extraction of salient features in each wavelet coefficient using the Canny’s filter. The edge detection process serves to simplify the analysis of images by drastically reducing the amount of data to be processed, while at the same time preserving useful structural information about object boundaries. We chose this filter because is based on Gaussian derivatives (GD) (see Fig. 3). Since receptive fields of simple cells in the primate visual cortex were well fit in the space and time domains by the GD model, filters based on GD are good operators for the detection of relevant image patterns at different spatial scales.\textsuperscript{18}
3. Build a decision map using Eq. 1.

\[
y_k^F = \begin{cases} 
  a_A^k & \text{if } a_A^k \in E_A^k \text{ and } a_B^k \notin E_B^k \\
  a_B^k & \text{if } a_B^k \in E_B^k \text{ and } a_A^k \notin E_A^k \\
  \max(a_A^k, a_B^k) & \text{if } a_A^k \in E_A^k \text{ and } a_B^k \in E_B^k \\
  \text{CICB} & \text{otherwise.}
\end{cases}
\]

where \(a_A^k\) and \(a_B^k\) are intensity values of decomposed input images at level \(k\), \(E_A^k\) and \(E_B^k\) are found edges of image A and B at level \(k\) respectively. \(y_k^F\) is intensity value of fused image in wavelet domain. CICB are context independent constant behavior operators. CICB is composed of operators which have the same behavior whatever the values of the information to combine and which are computed without any contextual or external information, i.e., \(\lor(a, b) \in \text{image } F(a, b) \geq \max(a, b)\).

4. Inverse Wavelet on the Interval transform of fused image \(y_k^F\).

![Figure 3. Canny operator in 3D](image)

### 3. EXPERIMENTAL RESULTS

We have performed various experiments using different orthogonal wavelets in order to compare the proposed algorithm with traditional fusion methods using a toolbox developed in Java language called **FUSION LAB** and described in. Fig. 4 shows a snapshot of the evaluation tool. We have separated our experiments in two classes:

- Multi-focus image fusion with *ground truth*. Images with a reference image (distortion free).
- Multi-focus image fusion without *ground truth*. Images without a reference image.

![Figure 4. Fusion Lab developed with Java 1.5 and Eclipse IDE](image)
3.1 Evaluation of multi-focus image fusion with ground truth

In experiment #1, we considered gray scale images of size 512×512 (see Fig. 5). The ground truth image where both the apple and the napkin container are in-focus is shown in Fig. 5(a), Fig. 5(b) has the napkin container blurred and the apple in-focus and Fig. 5(c) the other way around.

![Figure 5. Input images for apple and napkin container experiment.](image)

In scheme 1 we fused input images using Haar wavelet with two levels of decomposition*. Larger values of low-low and detail coefficients were selected for reconstructing the fused image (MAX-MAX). In this scheme we did not perform any feature extraction.

In scheme 2 we fused input images with the same rule of previous scheme (MAX-MAX) and we performed feature extraction with parameters: filter = Prewitt, \( \sigma = 0.6 \), low-threshold = -100 and high-threshold = -10.

Fig. 6(a) shows the decision map obtained from image A and Fig. 6(b) shows the decision map obtained from image B for scheme 2.

Fused image result of scheme 1 is shown in Fig. 7(a) and fused image result of scheme 2 is shown in Fig. 7(b). We compared both fused images with ground truth image in order to assess our proposed algorithm.22 Table 1 shows quantitative results.

![Figure 6. Decision Maps for proposed algorithm](image)

---

*Haar coefficients \( a_0 = \frac{1}{\sqrt{2}} \), \( a_1 = \frac{1}{\sqrt{2}} \)
In the fifth column we used a perceptual quality metric based on structural similarity (SSIM). In scheme 3 we fused input images using DB2 wavelet with one level of decomposition. Average values of low-low and detail coefficients were selected for reconstructing the image one (AV-AV). In this scheme we did not perform any feature extraction.

In scheme 4 we fused input images with the same rule of previous scheme (AV-AV) and we performed feature extraction.

\[ a_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, \quad a_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \quad a_2 = \frac{3-\sqrt{3}}{4\sqrt{2}} \text{ and } a_4 = \frac{1-\sqrt{3}}{4\sqrt{2}}\]
Fig. 9(a) shows the decision map obtained from image A and Fig. 9(b) shows the decision map obtained from image B for scheme 4. Fused image of scheme 3 is shown in Fig. 7(a) and Fig. 10(b) shows fused image of scheme 4. The quantitative results appear in Table 2.

![Decision maps for scheme 4](image)

Figure 9. Decision maps for scheme 4

![Fused images](image)

Figure 10. Fused images of human white blood cells experiment.

<table>
<thead>
<tr>
<th>scheme</th>
<th>PSNR [dB]</th>
<th>MSE</th>
<th>SNR$_W$</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>#3</td>
<td>30.086</td>
<td>9.8117e$^{-4}$</td>
<td>7.696</td>
<td>0.965</td>
</tr>
<tr>
<td>#4</td>
<td>44.391</td>
<td>3.638e$^{-5}$</td>
<td>16.431</td>
<td>0.995</td>
</tr>
</tbody>
</table>

### 3.2 Evaluation of multi-focus image fusion without ground truth

Among the quality metrics, the mean square error (MSE) and signal to noise ratio (SNR) are widely employed because they are easy to calculate and usually they have low computational cost, but these objective metrics require a reference image together with the processed image in order to evaluate the visibility of artifacts. This imposes obvious limitations on those applications where such metric cannot be used. Non-reference metrics are much more difficult to define, since the metrics are not relative to the original image by providing an absolute value associated to a given image. Here, we used a non-reference quality metric based on information that appears in.\(^21\)
For experiment #3 we used images by courtesy of Dr. Shutao Li (see Fig. 11). Input image A Fig. 11(a) has letters in-focus and drink blurred. Fig. 11(b) the other way around.

In scheme 5 we fused input images using SYM4 wavelet with one levels of decomposition\(^1\). We use (MAX-MAX) rule for merging coefficients. In this scheme we did not perform any feature extraction.

In scheme 6 we fused input images with the same rule of previous scheme (MAX-MAX) and we perform feature extraction with parameters: filter = Prewitt, \(\sigma = 0.7\), low-threshold = 15 and high-threshold = 25.

![Input image A](image1.png) ![Input image B](image2.png)

Figure 11. Input images for experiment #3.

Fused image result of scheme 5 is shown in Fig. 12(a) and fused image result of scheme 6 is shown in Fig. 12(b). The resulting performance metric for both schemes are given in Table 3.

![Fused image for scheme 5](image3.png) ![Fused image for scheme 6](image4.png)

Figure 12. Results for drink and letters experiment.

<table>
<thead>
<tr>
<th>Fusion</th>
<th>(I^q(A,F))</th>
<th>(I^q(B,F))</th>
<th>(M^q_{FAB})</th>
<th>(NM^q_{FAB})</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>2.277</td>
<td>2.181</td>
<td>4.459</td>
<td>0.320</td>
</tr>
<tr>
<td>#2</td>
<td>2.363</td>
<td>2.155</td>
<td>4.519</td>
<td>0.324</td>
</tr>
</tbody>
</table>

Table 3. Results of the fusion performance for experiment #3. Mutual Information uses \(\alpha = 0.8\)

In the last experiment #4 we fused three artificial images taken in our laboratory with a digital Reflex camera. Input images are shown in Fig. 14(a). These input images have distinct focal planes of the same scene.

\(^1\)SYM4 coefficients \(a_0 = -0.1294, a_1 = 0.2241, a_2 = 0.8365\) and \(a_3 = 0.4830\)
In scheme 7 we fused images using DB2 wavelet with one level of decomposition and MAX-MAX criteria for merging coefficients. In this scheme we did not perform any feature extraction.

In scheme 8 we fused images with the same criteria of scheme 7 but we performed feature extraction with parameters: filter = Sobel, $\sigma = 0.4$, low-threshold $= 0$ and high-threshold $= 25$.

![Input images A, B, C](image1)

Figure 13. Input images for experiment #4.

Fused image result of scheme 7 is shown in Fig. 14(a) and fused image result of scheme 8 is shown in Fig. 14(b).

![Fused images of scheme 7, 8](image2)

Figure 14. Fused images for experiment #4.

4. CONCLUSIONS

In this paper we proposed a feature level fusion algorithm that incorporates a Gaussian derivative model of early vision and we developed an evaluation tool called Fusion Lab. We evaluated the performance of the proposed algorithm. Based on results showed in Tables 1 to 3 and based on visual perception we can conclude that the feature level fusion method proposed performs better than tradition fusion methods. A possible explanation is due to the fact, that the feature level method preserves the information from the edges.

ACKNOWLEDGMENTS

This work has been sponsored by the following grants: UNAM PAPIIT IN106608, TEC2007-67025/TCM; TEC2006-28009-E and PI040765.
REFERENCES