Extending Depth of Field by Intrinsic Mode Image Fusion

Harishwaran Hariharan, Andreas Koschan, and Mongi Abidi
Imaging, Robotics and Intelligent Systems Laboratory
University of Tennessee, Knoxville, TN-37996
hari@utk.edu

Abstract

Here, a versatile data-driven application independent method to extend the depth of field is presented. The principal contribution in this effort is the use of features extracted by Empirical Mode Decomposition, namely Intrinsic Mode Images, for fusion. The input images are decomposed into intrinsic mode images and fusion is performed on the extracted oscillatory modes, by means of weighing schemes that allow emphasis of focused regions in each input image. The fused image unifies information from all focal planes, while maintaining the verisimilitude of the scene. In order to validate the fusion performance of our method, we have compared our results with those of region-based and multiscale decomposition based fusion techniques. Several illustrative examples and objective comparisons are provided.

1. Introduction

When a scene is being imaged, it is desirable in certain applications to have all the objects of the scene to be in focus. Typically, lenses possess the virtue of limited depth of field (DOF) and this makes the acquisition of such an all-in-focus image difficult, especially under limited illumination conditions. This is a major issue in many imaging purposes, e.g. inspection of microscopic scenes and long range feature tracking. In multifocus fusion, the critical initiative is to obtain focal information from dissimilar focal planes in the scene and fuse them into an image where all the focal planes appear to be in focus. In other words, we simulate acquisition with a lens having an infinite depth of field. Figure 1 demonstrates an example of such an extension of depth of field.

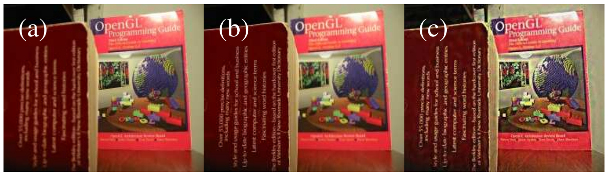


Figure 1. An example of multifocus fusion (a-b) input images where certain areas of the input images are in focus [Courtesy: 1], (c) multifocus fused image using proposed method,where both planes are in focus.

2. Related work

In the literature, various solutions to the problem are investigated by originating methods based on region based methods, multiscale decomposition (MSD), and learning methods. In region based method, the input images are initially divided into stacks of blocks [2,10] or into regions based on segmentation [3,4]. Based on a sharpness vote criterion, one region per stack is selected and a mosaic of such selected regions forms the fused image. The most commonly reported issues in this family of methods are blocking effects [5]. In MSD based methods, the input images are decomposed into multiscale coefficients; fusion rules are used in the selection or treatment of these coefficients and synthesized via inverse transforms to form the fused image [6, 7]. The most widely reported issues in this family are ringing effects and related distortions [5]. Furthermore, constructing a wavelet kernel that can adapt itself towards multiple applications is difficult. In cases with acutely restricted depth of field, such as in the case of microscopic multifocus fusion, it is imperative to address focal overlap between input images. In this effort, we discuss a data-driven general purpose multifocus fusion method that is capable of fusing data from varied applications, such as microscopic scene inspection and long range feature

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

tracking. The key contribution in our method is the use of intrinsic mode images in the fusion of multifocus images using the features extracted through Empirical Mode decomposition. In section 2, a basic description of Empirical Mode Decomposition (EMD) and details of our fusion algorithm are presented. Illustrative examples and comparisons are presented in Section 3 before drawing conclusions.

3. Intrinsic mode image fusion

In this effort, we present a method for fusing multifocus images by exploiting the potential of a relatively recent method, namely EMD, for analyzing nonlinear and non-stationary datasets developed by Huang *et al* [8]. This decomposition method is data-driven and application independent. The facet of decomposing a signal (in our case, a two dimensional image) into intrinsic mode images (IMIs) is employed in the fusion process. The motivation behind using EMD is that it is a local-global decomposition scheme which does not require a 'mother' kernel making it application independent. Since it involves no region segmentation and is a global method, border artifacts are absent. A typical EMD of an image employs a sifting process that elicits the finest oscillatory modes from the data, analogous to filtering particles through a set of fine to coarse sieves. A typical EMD of an image is shown in Figure 2 where the decomposition shows fine and superfine details of the image. The last IMF called the residue displays the trend of the data.

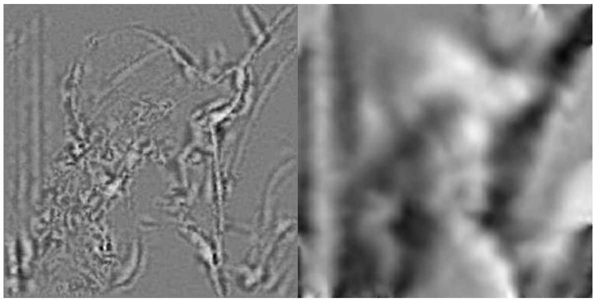
As per the intrinsic mode definition, the decomposition method can simply employ the envelopes defined by the local maxima and minima of the data individually. For clarity, we explain the EMD process for a channel matrix before being generalized for multiple image fusion. The extrema of an image channel *X* are identified and all local maxima (or minima) are interpolated by a cubic spline to form the maxima (or minima) envelope. The pixel-wise mean of the envelopes, m_1, is subtracted from the data r_0 for the first component h_1. For the first iteration, r_0 is the original image matrix *X*.

$$h_1 = r_0 - m_1 \text{, where } r_0 = X \tag{1}$$

In the second sift, h_1 is considered as the data where m_{11} is the mean of the h_1 envelopes.

$$h_{11} = h_1 - m_{11} \tag{2}$$

(a) Original Image (b) IMI 1



(c) IMI 2 (d) Trend Image (IMF 12)

Figure 2. A typical EMD of an image into IMIs (a) original image used for decomposition (b) IMI 1 (Super fine details) (c) IMI 2 (fine details) and (d) trend image (approximation). Intermediate IMIs are not shown due to space considerations.

The sifting is continued *k* times till the first IMI is obtained.

$$h_{1k} = h_{1(k-1)} - m_{1k} \tag{3}$$

We designate $c_1 = h_{1k}$ as the first IMF. To retain the physical meanings of the IMI, a standard deviation based stopping criterion is used. Sifting is ceased if the standard deviation, *SD*, computed from two consecutive sifting results is negligible.

$$SD = \sum_{t=0}^{T} \left[\frac{\left| \left(h_{1(k-1)} - h_{1(k)} \right) \right|^2}{h_{1(k-1)}^2 (t)} \right] \tag{4}$$

The isolated intrinsic mode image, c_1 contains the finest scale of the signal and we separate c_1 from the data.

$$r_1 = r_0 - c_1 \tag{5}$$

The subsequent data called the residue, r_1, still holds lower frequency information. In the next iteration, the residue r_1 is treated as the new data in place of r_0 and sifted. This procedure is repeated on all the subsequent residues (r_j's), to realize a set of IMIs.

$$r_1 - c_2 = r_2, \ldots\ldots, r_{n-1} - c_n = r_n \tag{6}$$

The sifting through residuals can be stopped when the residue becomes monotonic containing no IMIs. Based on the nature of the IMI, experiments were conducted to utilize these bases towards image fusion. Fusion is achieved during the Empirical Mode Synthesis. Given N input images, the fused image F is synthesized as follows,

$$F = \sum_{p=1}^{C_h} \sum_{q=1}^{M-1} \alpha_{pq} \sum_{i=1}^{N} c_{pqi} + \sum_{p=1}^{C_h} \alpha_{pM} \sum_{i=1}^{N} \gamma_{pMi}$$
$$1 \geq \alpha_{p1} \geq \alpha_{p2} \geq \ldots \geq \alpha_{pM} \geq \frac{1}{N}, \forall p=1,2,3 \tag{7}$$

where, M is the level of decomposition at which the residues becomes monotonic. The central idea of our fusion is emphasizing the superfine and fine details in a set of input images by weighing the corresponding IMIs with α_{pq}*'s* and α_{pM}. This emphasizes of the focused area in the input images. Our method is extendable to grayscale or color image datasets ($C_h = 1$ *or 3).*

4. Experimental Results

In our experiments, we have performed fusion and related analyses on various datasets from different imaging applications, varying from microscopic to longer range data sets. Our method assumes coregistered images. We have compared our method with a region based and multiscale decomposition based (MSD) method. In the region based technique, multiple size windows were used to select areas in focus as discussed in [2]. For MSD based fusion, we have implemented the fusion algorithm, due to Frechette and Ingle [7]. This method is selected as it is designed for fusion with multiple frames and has many parallels with our method. This method uses the coiflet wavelet (level 2) family which is reported suitable for multifocus fusion. In the example in Figure 3 (a-c), a series of images from a large chamber scanning electron microscope (LC-SEM) are shown, where various planes of the image are in focus. In Figure 3 (d) an image fused using the region selection method is shown. Prominent border artifacts are seen under close inspection, highlighted by the red arrows. In Figure 3(e), an image fused using MSD fusion is shown. While a good rendition of the scene is obtained, there are ringing effects upon close inspection. In Figure 3(f), an image fused using the proposed method is shown. Border artifacts are substantially reduced and a crisp overall perspective of the scene is obtained. In our implementation, we use $\alpha_{p1,2} = 1$, $\alpha_{p(q,\ldots,M)} = 1/N \, \forall$ $p = 1,2,3$, $SD = 0.1$ which emphasizes the fine and superfine details in the intrinsic mode images which unify and emphasize the areas in focus.

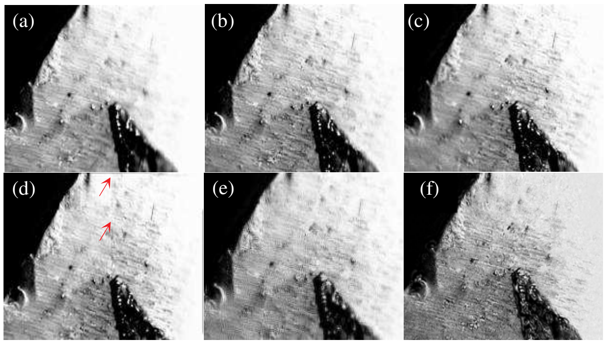


Figure 3. Comparison of different fusion methods (a-c) a few input microscopic (LC-SEM) images (note various sections of the input images are blurred due to the narrow depth of field), (d) fusion by region selection method [2] (Blocking effects shown with arrows) (e) fusion using MSD based fusion [7] and (f) fusion using proposed fusion method.

(a) Tire Tread Focused Image (b) Soil Focused Image

(c) MSD Based Fusion (d) Proposed Fusion

Figure 4. An example of multifocus fusion (a) input showing focused tire treads (f) input image showing focused soil (c) fusion using MSD based fusion [7] (h) fusion by proposed method.

In Figure 4 (a, b), input images from a terrain modeling application are shown. In Figure 4(a), the tread of the tire is in focus. In Figure 4 (b), the soil is in focus. In 4(c), fusion using a MSD fusion is shown. The result from the region selection method is not shown due to space and resolution considerations. In figure 4 (c), we see that while an overall rendition of the tire and soil appear together in the scene there is a considerable blurring due to ringing effects. The fused image using the proposed method, in Figure 4(d), is much sharper and details from both planes are prominently visible. Since improvement in fusion quality is hard to visually validate, we performed objective evaluations of the fused images as well. The images fused using region selection; MSD, and EMD fusion were evaluated using the Tenengrad sharpness measure [9]. The Tenengrad sharpness measure T for a fused image F is obtained by,

$$T = \sum_{i=1}^{m} \sum_{j=1}^{n} \sqrt{F_x^2(x,y) + F_y^2(x,y)}, \qquad (8)$$

where $m \times n$ is the total number of pixels in F and x and y denote directional gradient operations. The objective results are consistent with visual inspection and show that our method produces images with better overall sharpness. The results of some of the objective tests are summarized in Table 1. The images fused using our method have the most measured sharpness when compared against the other methods. The Macroscopic dataset is not shown due to space limitations.

Table 1. An objective comparison of fusion results.

	Tire-Soil	**Microscopic**	**Macroscopic**
MSD	2.47E+09	1.45E+09	1.28E+09
Region Selection	2.71E+09	3.71E+09	1.91E+09
Proposed	5.60E+09	6.40E+09	4.23E+09

5. Conclusions

A data-driven and multifocus fusion method, where Intrinsic Mode Images are used in the weighted synthesis of an all-in-focus image, is presented. Our method capitalized on using information from the different oscillatory modes of an image, while retaining the visual verisimilitude of the scene. We demonstrate multifocus fusion on datasets from different applications. Illustrative examples are presented along with comparisons. Our IMI fusion method outperformed the competing algorithms regarding sharpness in all our experiments.

Acknowledgements

This research was supported by the DOE URPR under grant DOE-DE-FG02-86NE37968 and BWXT-Y12 #4300056316.

References

[1] A. Goshtasby. Fusion of multi-focus images to maximize image information. *Defense and Security Symposium*, Orlando, Florida, 2006, pp. 17-21.

[2] R. Redondo, F. Sroubek, S. Fischer, and G. Cristobal. Multifocus fusion with multisize windows. *Proc. of SPIE*, 2005, pp. 410-418.

[3] D. Fedorov, B. Sumengen, and B. S. Manjunath. Multi-focus imaging using local focus estimation and mosaicking. *IEEE ICIP* 2006, pp. 2093-2096.

[4] Z. W. Liao, S. X. Hu, and Y. Y. Tang. Region-based multi-focus image fusion based on Hough transform and wavelet domain HMM. *Proc of International Conference on Machine Learning and Cybernetics*, 9:5490-5495, 2005.

[5] Z. Zhang and R. S. Blum. A categorization of MSD based image fusion schemes with a performance study for a digital camera application. *Proc. of the IEEE*, vol. 87, 1999, pp. 1315-1326.

[6] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and C. N. Canagarajah. Region-based image fusion using complex wavelets. *Proc. of International Conference on Information Fusion*, 2004, 555-562.

[7] S. Frechette, V. K. Ingle. Gradient based multi-focus video image fusion. *Proc. of IEEE Conference on AVSBS*, 2005, pp. 486-492.

[8] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London. A.,* 454:903-995, 1998.

[9] E. P. Krotkov, *Active Computer Vision by Cooperative Focusing*. USA: Springer Verlag, 1989.

[10] I. De and B. Chanda. A simple and efficient algorithm for Multifocus Image Fusion using Morphological Wavelets. *IEEE Transactions in Signal Processing*, vol. 86, 2006, pp. 924-936.