FUSION OF VISUAL AND THERMAL IMAGES USING COMPLEX EXTENSIONS OF EMD

D. Looney[1] and D. P. Mandic[1]

[1]Imperial College London, UK

E-mail: {david.looney06,d.mandic}@imperial.ac.uk

ABSTRACT

Empirical mode decomposition (EMD) is a fully data driven technique for decomposing signals into their natural scale components. Given its ability to separate spatial frequencies, it is natural to consider EMD for image fusion. However the problem of uniqueness, caused by the empirical nature of the algorithm and its sensitivity to parameters, makes it difficult to perform fusion of data from heterogeneous sources. A recently proposed solution to this problem is to use complex extensions of EMD which guarantees the same number of decomposition levels, that is the uniqueness of the scales. A new fusion rule based on the inherent properties of the decompositions is proposed. The methodology is used to address the fusion of images from different modalities (visual and thermal).

Index Terms— empirical mode decomposition (EMD), complex-valued signal processing, image fusion, thermal imaging

1. INTRODUCTION

A significant challenge in data and information fusion is the ability to fuse images from different modalities. With the increasing availablitiy of different image acquisition devices, a wide number of techniques [1],[2] ,[3] have been applied to the problem. The fused image ideally retains all relevant information from the input images while suppressing any unwanted artifacts such as noise. Specifically, the increasing availability of thermal cameras has let to a development of fusion techniques to combine visual and thermal images. Examples include Wavelet [4] and principal component analysis (PCA) [5] based algorithms.

The recently proposed empirical mode decomposition EMD [6] is a fully data driven technique which decomposes the signal into narrow-band components called intrinsic mode functions (IMFs). Unlike Fourier or wavelet based methods that project signals onto a fixed basis set, EMD makes no prior assumptions about the data and as such it belongs to the class of Exploratory Data Analysis (EDA) techniques [7]. The original algorithm was successfully applied to a number of problems which require high resolution but are separable in the time-frequency domain, such as in ocean engineering [8],

biomedical signal processing [9],[10] and seismics [11]. Furthermore the algorithm does not make unrealistic assumptions of strict orthogonality, unlike PCA, and instead the decompositions are said to be locally orthogonal [6].

Because of its ability to separate spatial frequencies, it is natural to consider EMD for the problem of heterogeneous image fusion. In [12] the use of EMD was proposed for the fusion of visual and thermal images. It was shown how EMD compared favourably with wavelet and PCA based approaches, particularly in retaining edge-based information from the different image modalities. Although the work [12] yielded good results and clearly illustrated the potential of EMD, an automated scheme for fusion was not presented and it relied on visual inspection to determine the most informative components. Furthermore, no guarantee could be made on the number of IMFs. More recent work [13] considers a 2D (two-dimensional) extension for the related task of multispectral image fusion. Although fusion is performed in an automatic fashion, the solution does not address the problem of uniqueness.

The fully adaptive and empirical nature of EMD as well as a sensitivity to parameters (such as stopping criteria or envelope estimation), compromises the uniqueness of the decomposition. For this reason, signals with similar statistics often yield different numbers of IMFs. This makes a multiscale comparison of IMFs from different sources unfeasible. A possible solution to this problem is a limiting on the number of performed sifting operations. In this way, although the exact number of decompositions can be controlled, it may adversely affect the component properties so that they no longer satisfy IMF criteria. Furthermore, a guarantee on the number of IMFs is only a necessary requirement which does not uniquely facilitate a physically meaningful comparison between decompositions. Instead, a set of common frequency scales must be found between data sources.

It was proposed in our previous work [14] to utilise recent complex extensions of the algorithm [15] [16] to address the problem of uniqueness. This was achieved by finding a set of frequency scales common to each of the image sources. Fusion of images with different focus points was subsequently achieved in an automated fashion by comparing the local variance of the real and imaginary components of the IMFs. In this paper, we extend the theory by proposing a new fusion

rule based on a more natural property of the complex IMFs which facilitates fusion in a block by block fashion. The proposed fusion rule is therefore more computationally light and lends itself to real time operation. We demonstrate the effectiveness of the algorithm by applying it to the fusion of thermal and visual images and making a comparison with other techniques.
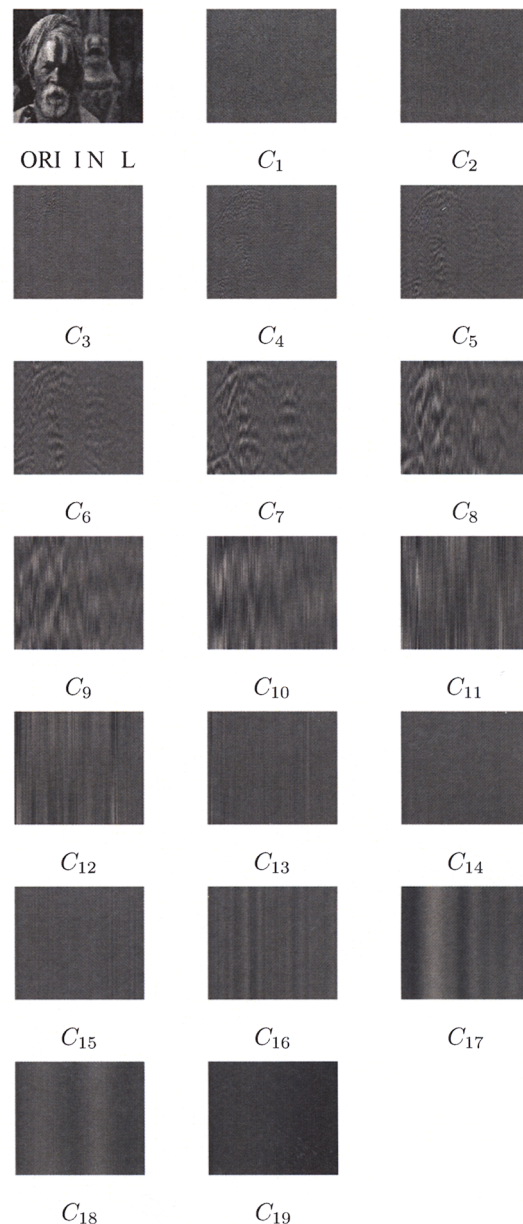
2. THE EMD ALGORITHM

Empirical mode decomposition [6] is a data driven time-frequency technique which adaptively decomposes a signal, by means of a process called the sifting algorithm, into a finite set of AM/FM modulated components. These components, called "intrinsic mode functions" (IMFs), represent the oscillation modes embedded in the data. By definition, an IMF is a function for which the number of extrema and the number of zero crossings differ by at most one, and the mean of the two envelopes associated with the local maxima and local minima is approximately zero. By EMD the signal $x(k)$ is expressed as

$$x(k) = \sum_{i=1}^{M} c_i(k) + r(k) \qquad (1)$$

where $c_i(k)$, $i = 1, \ldots, M$, is the set of IMFs and $r(k)$ is the residual. The first IMF is obtained as follows [6].

1. Let $\tilde{x}(k) = x(k)$;

2. Identify all local maxima and minima of $\tilde{x}(k)$;

3. Find an "envelope," $e_{min}(k)$ (resp. $e_{max}(k)$) that interpolates all local minima (resp. maxima);

4. Extract the "detail," $d(k) = x(k) - (1/2)(e_{min}(k) + e_{max}(k))$;

5. Let $\tilde{x}(k) = d(k)$ and go to step 2); repeat until $d(k)$ becomes an IMF.

Once the first IMF is obtained, the procedure is applied recursively to the residual $r(k) = x(k) - d(k)$ to obtain all the IMFs. An example of EMD is shown in Fig. 2 which illustrates the decomposition process. An image is converted into a vector by concatenating the rows. The IMFs obtained from the decomposition of the vector have been reconverted into the original 2D form, creating a set of scale images which are shown in Fig. 2. Note how each of 'Image Modes' represents the frequency scales within the image. The higher index IMFs contain high frequency detail such as noise and the image edges while slowly oscillating effects such as illumination are contained within the low index IMFs.

Fig. 1 Illstration o t e si ting process ia EMD oran original image (top le t Note o eac o t e IMFs C_1 - C_{19} represents t e re ency scales it in t e image T e ig er index IMFs contain ig re ency detail s c as noise and t e image edges ile slo ly oscillating e ects s c as ill - mination are contained it in t e lo index IMFs

3. PROBLEM OF UNIQUENESS

Due to the fully adaptive and empirical nature of EMD, as well as a sensitivity to parameters, signals with similar statistics often yield different decompositions both in terms of their number and their properties. This is known as the problem of uniqueness and it is a significant obstacle in applications of EMD. To demonstrate the problem of uniqueness, we consider a sinusoid with additive white Gaussian noise (AWGN) given by $\cos(k) + v(k)$ where $\cos(k)$ denotes the sinusoid and $v(k)$ the noise. The $M = 5$ extracted IMFs are given in Fig. 2(a). Note that the original noise-free sinusoid corresponds to the second IMF and that a standard fusion algorithm would be to retain this component and omit the rest. Performing EMD on the same sinusoid corrupted by different AWGN but with the same statistics as before (mean,variance), we obtain a different number of IMF components. Seven of the nine IMFs are shown in Fig. 2(b).

Note that in addition to a change in the number of IMFs, the signal of interest is contained within different components (C_2, C_3 and C_4) compared to the previous decomposition. This specific phenomenon whereby similar modes appear across different frequency scales is known as scale-mixing which even recent advances in stopping criteria cannot fully address [17].

The issue of uniqueness is a common problem that can be addressed by placing bounds on the number of performed sifting operations. However this is clearly sub-optimal as it affects the properties of decompositions. Other solutions include Ensemble EMD [18] and genetic based algorithms [19]. However their computational complexity is a significant disadvantage. In short it is clear that in order to use EMD for fusion, the following problems must first be addressed.

- The number of IMFs from each source must be equal

- The IMF attributes (frequency) from each source should also be equivalent to enable a meaningful comparison

The solution proposed in [14] is to decompose data from heterogeneous sources simultaneously, thus finding a set of common scales which are unique to the sources being analysed. This not only guarantees the number of IMFs but also ensures a physically meaningful comparison is possible. Recent hyper dimensional extensions of empirical mode decomposition [20],[15],[16] make it possible to determine such a set of common scales.

4. COMPLEX EXTENSIONS OF THE EMD ALGORITHM

As the need for multi-dimensional signal processing has become more apparent, several extensions of EMD have been

(a)   (b)
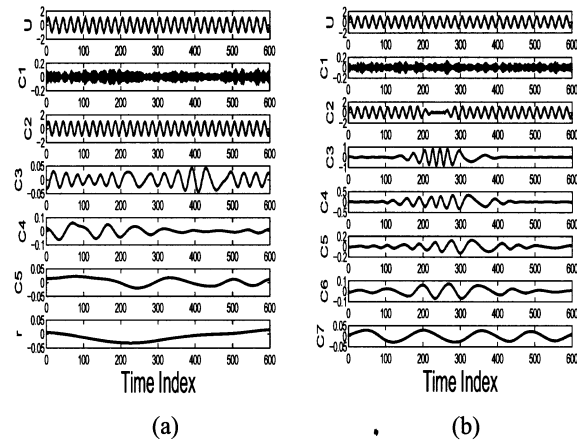
Fig. 2. EMD applied to a sinusoid corrupted by white Gaussian noise (denoted by U) and the extracted IMFs (where the kth IMF is denoted by C_k and r denotes the residual). The signals in (a) and (b) are both signals with similar statistics, yet the EMD algorithm has produced a different number of decompositions, $M = 5$ in (a) and $M = 9$ (of which 7 are shown) in (b). Consequently, the original signal of interest (sinusoid) can be located in different modes for different decompositions, C_2 in (a) and C_2, C_3 and C_4 in (b).

recently developed: "Complex Empirical Mode Decomposition" [20], "Rotation Invariant Empirical Mode Decomposition" [15] and "Bivariate Empirical Mode Decomposition" [16].

Complex EMD, the first extension to be introduced, is derived from the inherent relationship between the positive and negative frequency components of a complex signal and the properties of the Hilbert transform. The idea behind this approach is rather intuitive: first note that a complex signal has a two-sided, asymmetric spectrum. The complex signal can therefore be converted into a sum of analytic signals by a straightforward filtering operation that extracts the opposite sides of the spectrum. Direct analysis in $\mathbb{C}$ can subsequently be achieved by applying standard EMD to both the positive and negative frequency parts of the signal. Although it retains important properties of univariate EMD, such as its behaviour as a dyadic filter bank, it is difficult to interpret the meaning of extracted IMFs and the approach is not suitable for extensions to higher dimensions.

By comparison, rotation invariant EMD (RIEMD) operates completely within $\mathbb{C}$ based on the direct application of complex splines. The complex envelope approximation is characterised by the definition of suitable extrema in $\mathbb{C}$. This is not straightforward since $\mathbb{C}$ is not an ordered field [21], and RIEMD uses a locus where the angle of the complex-valued first derivative becomes zero, that is, it is based on a change in the phase of the signal. This definition is equivalent to the

extrema of the imaginary part, that is, for a complex signal $Z(t)$ (for convenience we here use a continuous time index t)

$$\angle \dot{Z}(t) = 0 \Rightarrow \angle \{\dot{x}(t) + j \cdot \dot{y}(t)\} = 0$$

$$\Rightarrow tan^{-1} \frac{\dot{y}(t)}{\dot{x}(t)} = 0 \Rightarrow \dot{y}(t) = 0 \qquad (2)$$

Unlike complex EMD, the IMFs of rotation invariant EMD possess a physical meaning as is illustrated by its analysis of real world complex quantities such as wind data [15].

Bivariate EMD [16] operates in a similar fashion to rotation invariant EMD. By projecting the data in k directions, the approach can consider extrema in several directions and construct a 3D tube to enclose them. The local mean is defined as the center of the tube which is described, at each interval, by k points (with each point being associated with a specific direction). Assuming $k = 4$ directions, the center of the tube at a point is given by either the barycenter of the four points or the intersection of straight lines passing through the middle of the tangents. RIEMD and bivariate EMD are equivalent for the case of $k = 2$. In addition the real and imaginary components of the IMFs tend to approximate standard real IMFs themselves

Like rotation invariant EMD, bivariate EMD generates an equal number of IMFs for the real and imaginary parts making it suitable for the analysis of multi-dimensional data. Since bivariate EMD can operate for $k > 2$, we will use this method for image fusion.

5. TOWARDS FUSION OF VISUAL AND THERMAL IMAGES

It was proposed in [14] to regard data from multiple sources as a single multi-dimensional entity which, for two sources, facilitates the use of a complex extension of EMD (either [15] or [16]). Under the assumption that the analysis determines common oscillations on a scale by scale level, the complex extensions can be used to determine which source contains locally the most information by performing a suitable comparison between the real and imaginary components of the IMFs. This is because salient features such as edges or lines will be highlighted by large variations in the amplitudes of the scales. The fusion rule used in [14] was based on the local variance of the IMFs. Here, it is instead proposed to base the decision on instantaneous amplitude, a natural choice given the nature of the decomposition. This is similar to wavelet-based fusion which retains decompositions which correspond to the largest coefficients. For rigour, simulations were initially performed on artificially generated data sets so as to illustrate the potential of bivariate EMD in finding "common data scales." Next, an automatic fusion algorithm using instantaneous amplitude is described and applied to a real world fusion problem, that of fusing thermal and visual images.

5.1. Test Sets

In the first experiment, we constructed two signals from a set of three sinusoidal components. These signals are visible in the top panels (denoted by $\mathbb{R}\{U\}$ and $\mathbb{I}\{U\}$) of Fig. 3. Common to each test signal were the frequencies of two of the sinusoids, although their amplitude and phase were different. A third high frequency sinusoid was added to $\mathbb{R}\{U\}$ only. Performing bivariate EMD on the complex signal constructed from each of these signals ($\mathbb{R}\{U\} + j\mathbb{I}\{U\}$), a set of complex IMFs were obtained, for which the real and imaginary values of which are shown in Fig. 3.

Observe the high frequency sinusoid, contained only within $\mathbb{R}\{U\}$, in the real part of the first IMF. On the other hand, the imaginary component, $\mathbb{I}\{C1\}$, shows comparatively less high frequency content. The methodology can determine, however, common frequency scales as is evident by comparing $\mathbb{R}\{C2\}$ and $\mathbb{I}\{C2\}$ as well as $\mathbb{R}\{C3\}$ and $\mathbb{I}\{C3\}$. Note that the approach is robust to changes in scale amplitude as well as phase and also that the individual components approximate standard real IMFs. To address the problem of uniqueness,
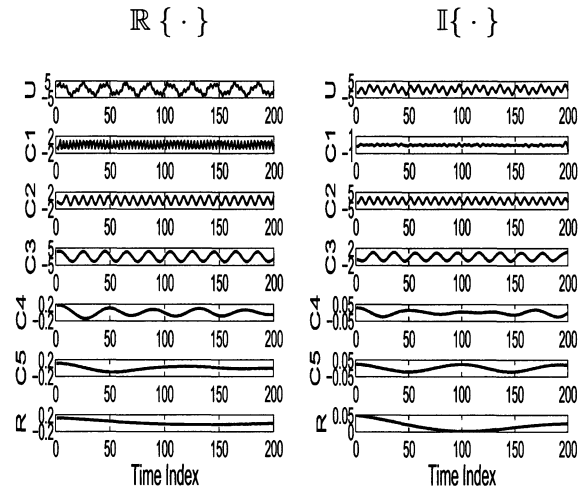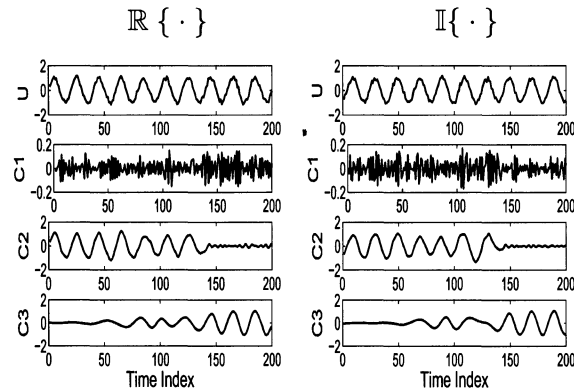

Fig. 3. Simultaneous decomposition of two test signals ($\mathbb{R}\{U\}$ and $\mathbb{I}\{U\}$) is achieved using bivariate EMD. The real part of the IMF components is given on the left side, and the imaginary part on the right side. Note that where common frequencies exist between $\mathbb{R}\{U\}$ and $\mathbb{I}\{U\}$), they appear simultaneously in the real and imaginary IMFs (C_2 and C_3).

in the next experiment we performed bivariate EMD on a complex signal, the real and imaginary components of which are shown respectively in the top panels (denoted by $\mathbb{R}\{U\}$ and $\mathbb{I}\{U\}$) of Fig. 4. The real component was a sinusoid corrupted by AWGN. The imaginary component was the same sinusoid, shifted by some arbitrary phase value, corrupted by different AWGN but with the same statistics as the AWGN affecting the real component. The set of complex IMFs are

displayed component-wise on Fig. 4.

Note that by design most of the noise is contained within the first IMF and the signal of interest (sinusoid) is spread across subsequent IMF components. This result demonstrates that although the approach can be affected by the problem of scale mixing, it is irrelevant as it occurs simultaneously in each channel as indicated by components C_2 and C_3. Thus, its robustness facilitates a meaningful comparison between scales even if scale-mixing occurs, and forms the basis for the proposed image fusion algorithm.

Fig. 4. Simultaneous decomposition of two test signals ($\mathbb{R}\{U\}$ and $\mathbb{I}\{U\}$) is achieved using bivariate EMD. Although affected by the problem of scale mixing, it is irrelevant as it occurs simultaneously in each component (C_2 and C_3).

5.2. EMD Image Fusion

Simultaneous decomposition of a set of visual and thermal images is proposed as follows. The images are divided into a set of N × N blocks. Consider two such image blocks, A and B, which refer respectively to the visual and thermal images. The rows of each of the images are concatenated so as to construct two vectors ($v1$ and $v2$) and, using bivariate EMD, the complex vector $v1 + jv2$ is decomposed into M complex IMFs. Separating the IMFs into their real and imaginary components and reconverting each into their original 2D form gives a set of M scale images for both A and B, denoted by A_i and B_i for $i = 1, \ldots, M$. This is illustrated in Fig. 5. The fused image block, F, is then given by

$$F = \sum_{i=1}^{M} [\alpha_i A_i + \beta_i B_i] \qquad (3)$$

where α_i and β_i are weighted coefficients which satisfy $\alpha_i + \beta_i = 1$. The values for the coefficients are determined by comparing the total instantaneous amplitude for each scale (denoted by $IA\{A_i\}$ for A_i and $IA\{B_i\}$ for B_i).

If $IA\{A_i\} \approx IA\{B_i\}$ then α_i and β_i are both set to 0.5. If, however, $IA\{A_i\} > IA\{B_i\}$ then $\alpha_i = 1$ and $\beta_i = 0$.
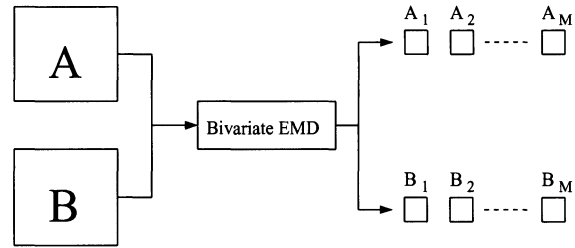
Fig. 5. Simultaneous decomposition of two image blocks using bivariate EMD.

Similarly, $\alpha_i = 0$ and $\beta_i = 1$ if $IA\{A_i\} < IA\{B_i\}$. This methodology is applied to each block. Thus, the fused image retains locally dominant features (in this case determined by instantaneous amplitude) at each scale.

5.3. Existing Fusion Methods

Before giving fusion results using the EMD-based methodology described above, it is first necessary to outline some existing fusion procedures using PCA and wavelets so as to provide comparisons.

- Pixel averaging across the thermal and visual images, highlighted by [12], can be used as a straightforward benchmark for image fusion.

- Image fusion using PCA is implemented as follows in our analysis. The image pixels from the visual and thermal images are arranged as columns of a matrix P. Eigenvalue analysis is performed on the correlation matrix of P. The matrix P is projected in the direction of strongest linear variance using the largest principal component, which provides the fusion output.

- Using the wavelet transform, the visual images are decomposed into an equal number of levels using a preselected basis function. For our analysis, a straightforward operation based on the decomposition level was selected [4] to perform fusion. Decompositions corresponding to the largest coefficients at each level are retained.

6. RESULTS

Fig. 6 shows visual and thermal images obtained from [12]. On the one hand, the visual image clearly depicts a pair of scissors partially obstructed by a plastic covering. On the other hand, the scissors is fully visible in the thermal image but overall it lacks the definition of the visual image. Additionally it is more susceptible to noise.

Fig. 7 shows the fused image using averaged pixel values. Clearly this is suboptimal as it is equivalent to a low pass

filtering operation and therefore distorts edges and other high frequency detail. Fig. 7 shows the results using PCA which, firstly, is not a great improvement over pixel averaging and, secondly, contains no colour information.

Fig. 9 and Fig. 10 shows respectively the fusion results using the wavelet based approach and the proposed EMD approach, which are both clearly superior to the results obtained using straightforward pixel averaging and PCA. Note that the results are similar due to the similarity of the two fusion schemes. However, the fully data driven approach of EMD means that high frequency content can be more accurately analysed and compared. This is illustrated by Fig. 10 which retains sharp edges and detail while also maintaining an overall reduction in high frequency artifacts compared to the wavelet approach. Note for example, the 'block artifacts' around the markers and the left sticker in Fig. 9.
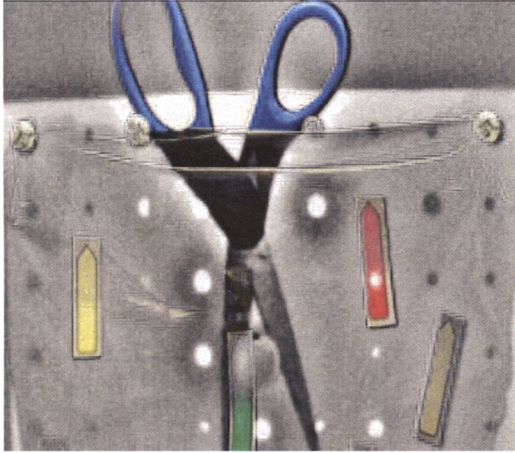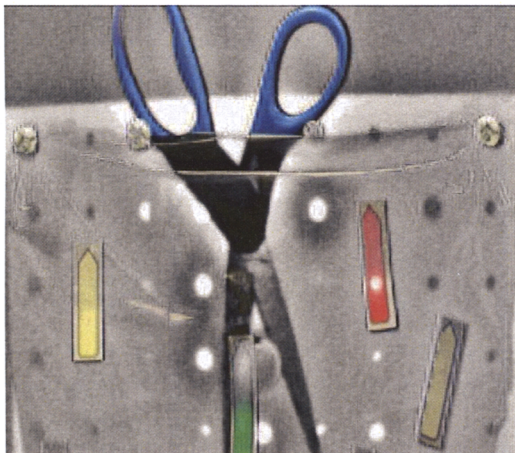

Fig. 7. An fused image achieved by performing pixel averaging.

Fig. 8. An fused image achieved by PCA.

Fig. 6. A visual image (above) and a thermal image (bottom).

7. CONCLUSION

The potential of higher dimensional EMD for the purposes of fusing visual and thermal images has been illustrated. A set of common frequency scales can be determined by simultaneously decomposing sources using the complex valued extensions of the algorithm which facilitates heterogeneous fusion. In this paper, a fusion rule has been proposed based on the instantaneous amplitudes of the components, a fundamental property of the decomposition. The proposed methodology facilitates a "block by block" comparison between image inputs. It has been demonstrated by simulations on real world data how it gives a superior performance over PCA and straightforward pixel averaging, and is matched only by a wavelet fusion scheme. However, the data driven nature of EMD means it has a greater potential for the analysis of high frequency content and its "block by block" methodology facilitates a near real time fusion operation. In future work, it is hoped to explore other fusion schemes that can take full advantage of the local nature of the algorithm.

Fig. 9. Wavelet based fusion

Fig. 10. Fusion using complex extensions of EMD

8. REFERENCES

[1] L. P. Yaroslavsky, B. Fishbain, A. Shteinman, and S. Gepshtein, "Processing and fusion of thermal and video sequences for terrestrial long range observation systems," in *Proceedings of the 7th Annual International Conference of Information Fusion*, 2004, pp. 848–855.

[2] D. A. Fay, A. M. Waxman, M. Aguilar, D. B. Ireland, J. P. Racamato, W. D. Ross, W. W. Streilein, and M. I. Braun, "Color visualization, target learning and search," in *Proceedings of the 3rd Annual International Conference of Information Fusion*, 2000, pp. 215–219.

[3] D. A. Socolinsky and L. B. Wolff, "Multispectral image visualization through first-order fusion," *IEEE Transactions on Image Processing*, vol. 11, no. 8, pp. 923–931, August 2002.

[4] H. Li, B. S. Manjunathand, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graphic Models and Image Processing*, vol. 57, no. 3, pp. 235–245, 1995.

[5] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis, "Face recognition by fusing thermal infrared and visible imagery," *Image and Vision Computing*, vol. 24, no. 7, pp. 727–742, 2006.

[6] N. E. Huang, Z. Shen, S. R. Long, M. L. Wu, H. H. Shih, Z. Quanan, 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,"

Proceedings of the Royal Society A, vol. 454, pp. 903–995, 1998.

[7] J.W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, Reading, MA, 1977.

[8] M. Datig and T. Schlurmann, "Performance and limitations of the Hilbert–Huang transformation (HHT) with an application to irregular water waves," *Ocean Engineering*, vol. 31, pp. 1783–1834, October 2004.

[9] T. M. Rutkowski, R. Zdunek, and A. Cichocki, "Multichannel EEG brain activity pattern analysis in time-frequency domain with nonnegative matrix factorization support," in *International Congress Series*, 2007, vol. 1301, pp. 266–269.

[10] J. Dauwels, T. M. Rutkowski, F. Vialatte, and A. Cichocki, "On the synchrony of empirical mode decompositions with application to electroencephalography," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 473–476.

[11] R. R. Zhang, S. Ma, E. Safak, and S. Hartzell, "Hilbert–Huang transform analysis of dynamic and earthquake motion recordings," *Journal of Engineering Mechanics*, vol. 129, pp. 861–875, 2003.

[12] H. Hariharan, A. Gribok, M. A. Abidi, and A. Koschan, "Image fusion and enhancement via empirical mode decomposition," *Journal of Pattern Recognition Research (JPRR)*, vol. 1, no. 1, pp. 16–32, 2006.

[13] X. Xu, H. Li, and A. N. Wang, "The application of BEMD to multi-spectral image fusion," in *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, 2007, vol. 1, pp. 448–452.

[14] D. Looney and D. Mandic, "Multi-scale image fusion using complex extensions of EMD," *Submitted to IEEE Signal Processing Letters*, 2008.

[15] M. Umair Bin Altaf, T. Gautama, T. Tanaka, and D. P. Mandic, "Rotation invariant complex empirical mode decomposition," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007, vol. 3, pp. 1009–1012.

[16] G. Rilling, P. Flandrin, P. Goncalves, and J.M. Lilly, "Bivariate empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 14, pp. 936–939, 2007.

[17] B. Xuan, Q. Xie, and S. Peng, "EMD sifting based on bandwidth," *IEEE Signal Processing Letters*, vol. 14, no. 8, pp. 537–540, 2007.

[18] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," Tech. Rep. 193, Center for Ocean-Land-Atmosphere Studies, 2004.

[19] Y. Kopsinis and S. McLaughlin, "Investigation of the empirical mode decomposition based on genetic algorithm optimization schemes," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2007, vol. 3, pp. 1397–1400.

[20] T. Tanaka and D. P. Mandic, "Complex empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 14, no. 2, pp. 101–104, Feb 2007.

[21] D. P. Mandic, M. Golz, A. Kuh, D. Obradovic, and T. Tanaka, *Signal Processing Techniques for Knowledge Extraction and Information Fusion*, Springer, 2008.