IMAGE FUSION USING A 3-D WAVELET TRANSFORM

S. G. Nikolov, D. R. Bull, C. N. Canagarajah, M. Halliwell, P. N. T. Wells

University of Bristol, UK

IMAGE FUSION

The successful fusion of images acquired from different modalities or instruments is of great importance in many applications, such as medical imaging, microscopic imaging, remote sensing, computer vision, and robotics. With 3-D imaging and image processing becoming widely used, there is a growing need for new 3-D image fusion algorithms capable of combining 3-D multimodality or multisource images. Such algorithms can be used in areas such as 3-D medical imaging (e.g. fusion of Magnetic Resonance (MR) and Computed Tomography (CT) images, fusion of MR and ultrasound (US) images), or 3-D microscopic imaging (e.g. fusion of Confocal Laser Scanning Microscopy (CLSM) and Scanning Acoustic Microscopy (SAM) images, fusion of MR and CLSM images).

Image fusion can be defined as the process by which several images, or some of their features, are combined together to form a single image. Let us consider the case where we have only two original images I_1 and I_2, which have the same size and which are already aligned using some image registration algorithm. Our aim is to combine the two input images into a fused image I. Image fusion can be performed at different levels of the information representation. Four different levels can be distinguished according to [1], i.e. signal, pixel, feature and symbolic levels. When fusion is done at pixel level the input images are combined without any pre-processing. To date, the results of pixel level image fusion in areas such as remote sensing and medical imaging are primarily intended for presentation to a human observer for easier and enhanced interpretation. Therefore, the perception of the fused image is of paramount importance when evaluating different fusion schemes. In the case of pixel level fusion, some generic requirements can be imposed on the fusion result: (a) the fused image should preserve, as closely as possible, all relevant information contained in the input images [2]; (b) the fusion process should not introduce any artifacts or inconsistencies, which can distract or mislead the human observer, or any subsequent image processing steps [2]; (c) in the fused image, irrelevant features and noise should be suppressed to a maximum extent. Pixel level fusion algorithms vary from very simple, e.g. image averaging, to very complex, e.g. Principal Component Analysis (PCA), pyramid based image fusion and wavelet transform (WT) fusion. Several approaches to pixel level fusion can be distinguished, depending on whether the images are fused in the spatial domain or they are transformed into another domain, and their transforms are fused. After the fused image is generated, it may be processed further and some features of interest may be extracted.

THE 3-D WAVELET TRANSFORM

The separable extension of the wavelet transform (WT) to three dimensions is straightforward and is explained in [3, 4, 5]. The three-dimensional separable wavelet transform (3-D WT) $W(x,y,z) = \omega(I(x,y,z))$ of the image $I(x,y,z)$ can be calculated by convolving the image with a pair of filters, one of which is a high-pass filter H, while the other is a low-pass filter L, along the z, y, and x directions. Thus, when applying the 3-D WT to volume data, the resulting filter cascade is constructed using seven oriented high-pass filters and one low-pass filter and the volume is decomposed into eight channels (subvolumes) at each band (scale) (Figure 1).

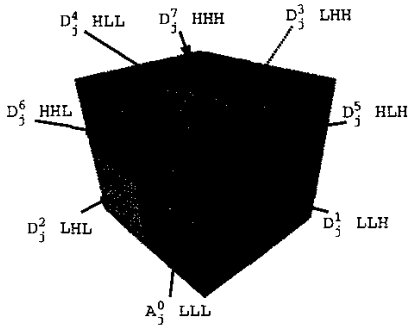

Figure 1: The eight different channels in one band of the 3-D WT. Each of the eight channels can be named using the following notation: $A_j^0 = LLL$, $D_j^1 = LLH$, $D_j^2 = LHL$, $D_j^3 = LHH$, $D_j^4 = HLL$, $D_j^5 = HLH$, $D_j^6 = HHL$, $D_j^7 = HHH$, where $j = 1, 2, \ldots, J$, L stands for a low-pass filter, H stands for a high-pass filter, and the filters are applied first along the z direction (right-most position), then along the y direction (middle position), and finally along the x direction (left-most position).

3-D WT IMAGE FUSION

The wavelet transform offers several advantages over similar pyramid based techniques when applied to image fusion: (a) the wavelet transform is a more compact representation than the image pyramid. This becomes of very great importance when it comes to fusion of 3-D and 4-D images. The size of the WT is the same as the size of the image. On the other hand, the size of a Laplacian pyramid, for instance, is $\frac{4}{3}$ of the size of the image; (b) the wavelet transform provides directional information, while the pyramid representation doesn't introduce any spatial orientation in the decomposition process [6]; (c) in pyramid based image fusion, the fused images often contain blocking effects in the regions where the input images are significantly different. No such artifacts are observed in similar wavelet based fusion results [6]; (d) images generated by wavelet image fusion have better signal to noise ratios (SNR) than images generated by pyramid image fusion, when the same fusion rules are used [7]. When subject to human analysis, wavelet fused images are also better perceived according to [6, 7].

Several wavelet based techniques for fusion of 2-D images have been described in the literature [8, 9, 6, 10, 7, 2]. All of these publications study only the case of 2-D image fusion. In this paper, we have extended some of the above mentioned 2-D fusion schemes to 3-D images. New 3-D fusion schemes are also presented. The 3-D WT image fusion algorithms described in this study have been used to combine both phantom (texture and non-texture) images (Figure 5) and multimodality (CT and MR) images (Figures 6 and 7).

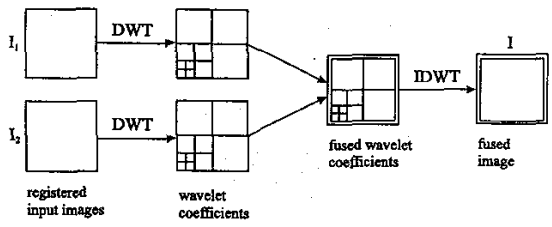

Figure 2: Fusion of the WT of two images.

The general idea of all wavelet based image fusion schemes is that the wavelet transforms ω of the two registered input images $I_1(x,y,z)$ and $I_2(x,y,z)$ are computed and these transforms are combined utilising some kind of fusion rule ϕ (Figure 2). Then, the inverse wavelet transform ω^{-1} is computed, and the fused image I is reconstructed:

$$\begin{aligned} I(x,y,z) &= \omega^{-1}(\phi(\omega(I_1(x,y,z)),\omega(I_2(x,y,z)))) \\ &= \omega^{-1}(\phi(W_1(x,y,z),W_2(x,y,z))) \quad (1) \end{aligned}$$

The fusion rule ϕ is actually a set of fusion rules ϕ_j^c, where $j = 1,..,J$ and $c = 1,\ldots,7$, which define the fusion of each pair of corresponding channels for each band.

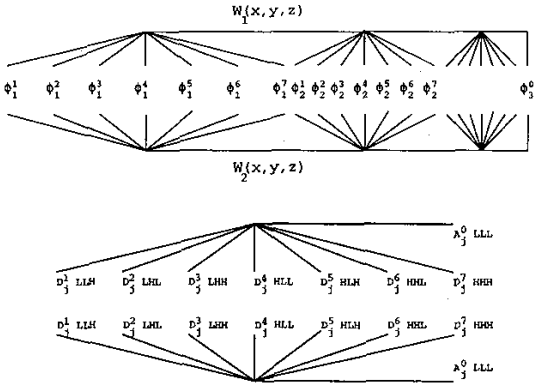

Figure 3: A WT fusion diagram (top) - the upper part of the diagram shows the wavelet decomposition of the first 3-D image, while the lower part - the wavelet decomposition of the second 3-D image; WT fusion diagram notation (bottom) - one band of the WT is shown together with the symbols for each channel (see Figure 1).

Thus, a wavelet transform fusion diagram (WTFD) can be constructed (Figure 3), which defines and illustrates the fusion process of the two three-dimensional wavelet transforms. Figure 4 shows an example of a wavelet transform fusion diagram.

A number of fusion rules can be used to combine the wavelet coefficients of two 3-D wavelet transforms. Some fusion rules (1,2,3) have been suggested by other authors to combine 2-D images. Here we have given their 3-D equivalents. Other, more advanced fusion schemes (4,5,6) are proposed in this study. Some of them have been used by the authors to fuse 3-D phantom and medical images.

1. fusion by averaging [10] (see Figure 5 (middle left and right)) - for each band of decomposition and for each channel the wavelet coefficients of the two images are averaged, i.e. $D_j^c I = (D_j^c I_1 + D_j^c I_2)/2$, $A_J^0 I = (A_J^0 I_1 + A_J^0 I_2)/2$, where $j = 1,..,J$ and $c = 1,\ldots,7$.

2. fusion by maximum [6, 10, 2] (see Figure 5 (bottom left) and Figure 7 (bottom)) - for each band of decomposition and for each channel, the maximum of the respective wavelet coefficients is taken, i.e. $D_j^c I = \max(D_j^c I_1, D_j^c I_2)$,

$A_J^0 I = \max(A_J^0 I_1, A_J^0 I_2)$, where $j = 1,..,J$ and $c = 1,\ldots,7$. An better option is $A_J^0 I = (A_J^0 I_1 + A_J^0 I_2)/2$.

3. high/low fusion [10] (see Figure 5 (bottom right)) - the high frequency information is kept from one image while the low frequency information is kept from the other, e.g. $D_j^c I = D_j^c I_1$ and $A_J^0 I = A_J^0 I_2$, where $j = 1,..,J$ and $c = 1,\ldots,7$.

4. composite fusion - various combinations of the different channels of $W_1(x,y,z)$ and $W_2(x,y,z)$ are composed, where some channels are taken from the first WT and some from the second.

5. fusion by denoising (hard or soft thresholding) - the wavelet coefficients of the high frequency components $D_j^7 I_k$ are thresholded by either hard or soft thresholding, where $j = 1,..,J$, $c = 1,\ldots,7$ and $k = \{1,2\}$. The main goal of thresholding is to remove the noise in the input images. In hard thresholding the absolute values of all wavelet coefficients are compared to a fixed threshold τ. If the magnitude of the coefficient is less than the threshold, the coefficient is replaced by zero:

$$\tilde{D}_j^c I = \begin{cases} 0 & \text{if } D_j^c I < \tau \\ D_j^c I & \text{otherwise} \end{cases} \qquad (2)$$

Soft thresholding (see Figure 6 (bottom)) shrinks all the wavelet coefficients towards zero:

$$\tilde{D}_j^c I = sign(D_j^c I)(|D_j^c I| - \tau)_+ \qquad (3)$$

6. fusion of the graphs formed from the WT maxima - the WT maxima (the multiscale edges of the 3-D image) can be linked to construct graphs. These graphs can be combined instead of combining all the wavelet coefficients. This fusion technique is based on the results reported in [11, 12].

Several other 2-D WT image fusion algorithms have been proposed in the literature, which are based on some of the principles of visual perception, e.g. fusion using an area based selection rule with a consistency verification [6] or contrast sensitivity fusion [7]. Since some of these methods have been designed specifically to improve the interpretation of fused 2-D images, their three-dimensional analogues are difficult to construct.

Fusion diagrams can be used to illustrate more complex WT fusion schemes where one or several filters are applied to each band of the two wavelet transforms prior to fusion (Figure 4).

Many multimodality images are made up of both smooth and textured regions. Such images can be segmented in terms of smooth and textured regions

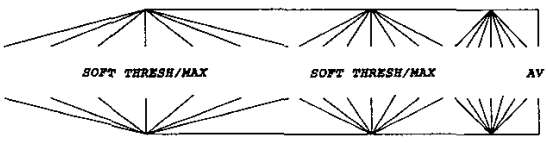

Figure 4: Wavelet transform fusion by soft thresholding and maximum.

by analysing their wavelet transforms [13], and depending on each pair of regions to be combined (i.e. smooth with smooth region, smooth with textured region, textured with textured region), different fusion rules can be used. Several examples of 3-D WT image fusion are presented in this paper. Figure 5 shows the fusion of two 3-D phantom images, i.e. a solid cube and a 3-D texture which consists of a grid of lines parallel to the three axes. Cross-sections of the input images and the fused image are displayed in Figure 5. The application of 3-D WT image fusion to combine multimodality (CT and MR) images is illustrated in Figures 6 and 7.

In the image fusion literature only a few measures for quantitative evaluation of different image fusion schemes have been proposed [7, 6]. Generally, the construction of the 'perfect' fused image $\bar{I}$ is an ill-defined problem, since in most cases the optimal combination is not known in advance. However, in very special cases, some measures for image fusion evaluation can give meaningful results. The change in the SNR is used as a measure in [7] to compare different fusion schemes, i.e. pyramid and wavelet based fusion. A performance measure defined as the standard deviation of the difference image between the 'perfect' fusion result $\bar{I}$ formed manually by 'cutting and pasting', and the fused image I, is suggested in [6] to evaluate the fusion of a pair of multifocus test images. Both of these measures are very much application dependent and they cannot be used to compare any kind of image fusion results. With 3-D multimodality imaging routinely used in practice, new measures for quantitative evaluation of 3-D image fusion algorithms are needed.

CONCLUSION

In this paper we have presented a new approach to 3-D image fusion using a 3-D separable wavelet transform. Several known 2-D WT fusion schemes have been extended to handle 3-D images and some new image fusion schemes (i.e. fusion by hard and soft thresholding, composite fusion, fusion of the WT maxima graphs) have been proposed. The goal of this paper is to present the new framework for 3-D image fusion using the wavelet transform, rather than to compare the results of the various

fusion rules. Wavelet transform fusion diagrams have been introduced as a convenient tool to visually describe different image fusion schemes. A very important advantage of using 3-D WT image fusion over alternative image fusion algorithms is that it may be combined with other 3-D image processing algorithms working in the wavelet domain, such as 'smooth versus textured' region segmentation [13], volume compression [3, 5], where only a small part of all wavelet coefficients are preserved, and volume rendering [3, 5], where the volume rendering integral is approximated using multiresolution spaces. The integration of 3-D WT image fusion in the broader framework of 3-D WT image processing and visualisation is the ultimate goal of the present study. This work was funded by UK EPSRC Grant #GR/L53373.

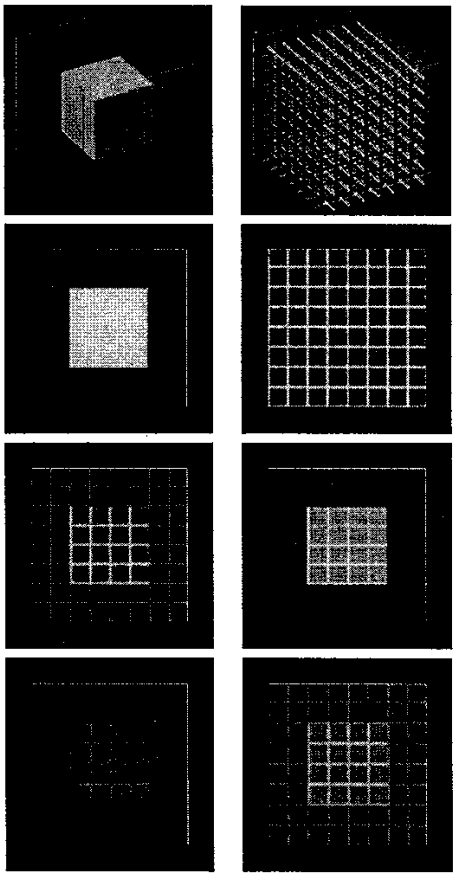

Figure 5: Wavelet transform fusion (Daubechies 4 wavelet) of two 3-D phantom images: first input image (top left) and second input image (top right); cross-sections at $y = 32$: first input image (second left); second input image (second right); fused image using averaging (third left); fused image using weighted averaging ($\alpha_1 = 0.75$) (third right); fused image using maximum (bottom left); fused image using the high frequency wavelet coefficients of the second image and the low frequency wavelet coefficients of the first image (bottom right). In all cases the fusion has been done in the wavelet domain. Volume size: $64 \times 64 \times 64$ voxels.

REFERENCES

[1] Abidi M. A. and Gonzalez R. C., editors, 1992, *"Data Fusion in Robotics and Machine Intelligence"*. Academic Press.

[2] Rockinger O., 1996, "Pixel-level fusion of image sequences using wavelet frames". In Mardia K. V., Gill C. A., and Dryden I. L., editors, *Proceedings in Image Fusion and Shape Variability Techniques, Leeds, UK*, Leeds University Press, 149–154.

[3] S. Muraki, 1992, "Approximation and rendering of volume data using wavelet transforms". In *Proceedings of the IEEE Visualization '91 Conference*, IEEE Press, 21–28.

[4] Press W. H., Teukolsky S. A., Vetterling W. T., and Flannery B. P., 1992, *"Numerical Recipes in C"*. Cambridge University Press, Cambridge, UK.

[5] Gross M., 1994, *"Visual Computing"*. Springer-Verlag, Berlin, Germany.

[6] Li H., Manjunath B. S., and Mitra S. K., 1995, "Multisensor image fusion using the wavelet transform". *Graphical Models and Image Processing*, 57(3), 235–245.

[7] Wilson T. A., Rogers S. K., and Myers L. R., 1995, "Perceptual based hyperspectral image fusion using multiresolution analysis". *Optical Engineering*, 34(11), 3154–3164.

[8] Koren I., Laine A., and Taylor F., 1995, "Image fusion using steerable dyadic wavelet transforms". In *Proceedings 1995 IEEE International Conference on Image Processing, Washington D.C.*, IEEE, 232–235.

[9] Chipman L. J. and Orr T. M., 1995, "Wavelets and image fusion". In *Proceedings 1995 IEEE International Conference on Image Processing, Washington D.C.*, IEEE, 248–251.

[10] Le Moigne J. and Cromp R. F., 1996, "The use of wavelets for remote sensing image registration and fusion". Technical Report TR-96-171, NASA Goddard Space Flight Center.

[11] Mallat S., 1989, "A theory for multiscale signal decomposition: The wavelet representation". *IEEE Transactions on PAMI*, 11(7), 674–693.

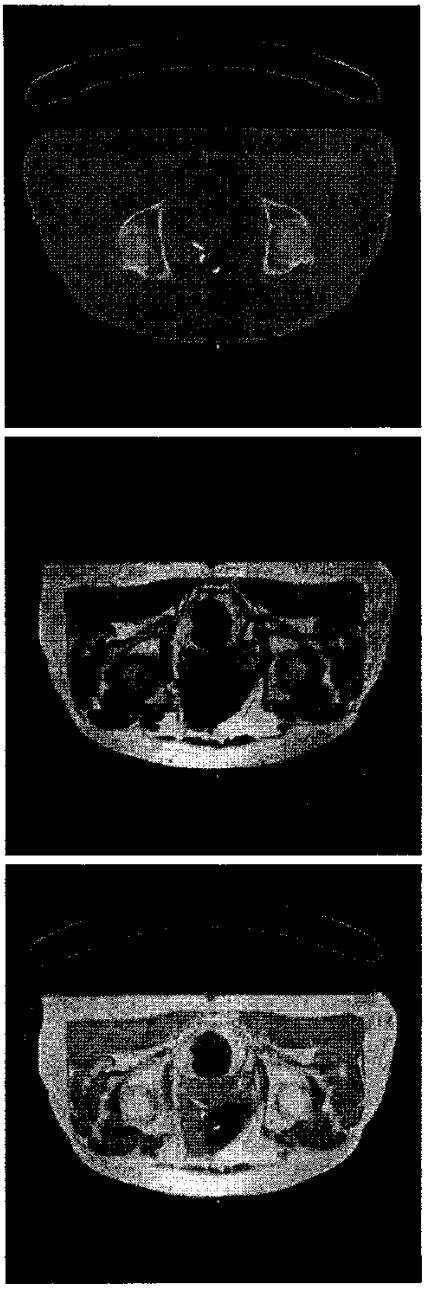

Figure 6: Wavelet transform fusion of two 3-D images of a human pelvis: CT image (top), MR image (middle), and WT fusion by soft thresholding and maximum, Daubechies 4 wavelet (bottom). Cross-sections at $z = 80$ are displayed. Isotropic volume size: $470 \times 470 \times 140$ voxels. Voxel size: 1^3mm. Images courtesy of the Bristol Oncology Centre (BOC) and the Bristol Royal Infirmary (BRI), Bristol, UK.

[12] Mallat S. and Zhong S., 1992, "Characterization of signals from multiscale edges". *IEEE Transactions on PAMI*, 14(7), 710–732.

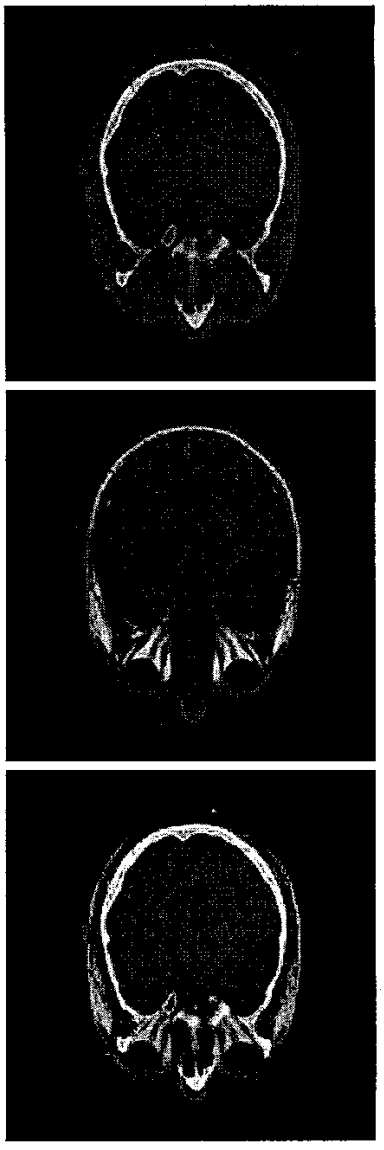

Figure 7: Fusion of two 3-D images of a human head: Frozen CT image (top), MR (T1) image (middle), WT fusion by maximum, Symmlet 6 wavelet (bottom). Cross-sections at $z = 60$ are displayed. Volume size: $256 \times 256 \times 164$ voxels. The input images are part of the Visible Human Project Collection and are courtesy of the National Library of Medicine, USA.

[13] Porter R. and Canagarajah N., 1996, "A robust automatic clustering scheme for image segmentation using wavelets". *IEEE Transactions on Image Processing*, 5(4), 662–665.