

Multispectral Multisensor Image Fusion Using Wavelet Transforms

By George Lemeshevsky

Open-File Report 99-626

U.S. Department of the Interior
U.S. Geological Survey

Multispectral multisensor image fusion using wavelet transforms

George P. Lemeshefsky
U.S. Geological Survey
Reston, VA 20192

ABSTRACT

Fusion techniques can be applied to multispectral and higher spatial resolution panchromatic images to create a composite image that is easier to interpret than the individual images. Wavelet transform-based multisensor, multiresolution fusion (a type of band sharpening) was applied to Landsat thematic mapper (TM) multispectral and coregistered higher resolution SPOT panchromatic images. The objective was to obtain increased spatial resolution, false color composite products to support the interpretation of land cover types wherein the spectral characteristics of the imagery are preserved to provide the spectral clues needed for interpretation. Since the fusion process should not introduce artifacts, a shift invariant implementation of the discrete wavelet transform (SIDWT) was used. These results were compared with those using the shift variant, discrete wavelet transform (DWT). Overall, the process includes a hue, saturation, and value color space transform to minimize color changes, and a reported point-wise maximum selection rule to combine transform coefficients. The performance of fusion based on the SIDWT and DWT was evaluated with a simulated TM 30-m spatial resolution test image and a higher resolution reference. Simulated imagery was made by blurring higher resolution color-infrared photography with the TM sensors' point spread function. The SIDWT based technique produced imagery with fewer artifacts and lower error between fused images and the full resolution reference. Image examples with TM and SPOT 10-m panchromatic illustrate the reduction in artifacts due to the SIDWT based fusion.

Keywords: image fusion, shift invariant wavelet transform, multispectral, TM

1. INTRODUCTION

Image classification techniques can be applied to the moderate spatial resolution multispectral imagery of the Landsat thematic mapper (TM) to produce thematic maps of land surface cover. One classification method (i.e., supervised classification) requires extracting training pixels that represent the spectral signatures of the land cover types. This process is facilitated by having field-collected information about cover type. In some cases, cover type may be identified by visual interpretation of the multispectral (XS) imagery, displayed as a false composite on a red, green, blue (RGB) color monitor. In conducting photointerpretation of an XS image, an analyst may use spectral clues based on prior knowledge and experience with spectral reflectance characteristics of ground cover types [1]. To augment this process, the analyst may use panchromatic (P) imagery of higher spatial resolution to gather additional clues related to spatial information (i.e., size, shape, orientation, and texture.)

Interpretation may be improved by using a composite image created from XS and coregistered higher spatial resolution P data. The process of combining image data to create a composite having the best characteristics of both is called merging, fusion, or sharpening of the lower resolution image [2]. There are many reported techniques for multi-image fusion, and much of the earlier reported work was with lower spatial resolution XS data (e.g., SPOT and TM) and higher resolution SPOT P; for a complete description see [2].

Email address: glemeshe@usgs.gov

The fusion result should be a meaningful combination of the data, and to be effective, requires radiometric correlation between the two images [2]. The composite should be free of artifacts and preserve, as much as possible, the spectral characteristics (radiometric fidelity) of the original data to facilitate interpretation based on spectral clues and knowledge of the spectral signatures of cover types from the original data [1].

The fusion technique described herein transfers the higher spatial frequency information of a P image to improve the spatial resolution of an XS image. The sharpening/fusion/merging examples are for three of the 30-m ground-projected instantaneous field of view (GIFOV) TM bands and a 10-m GIFOV SPOT P image. The fused result can be displayed as an RGB image. The fusion process makes use of the discrete wavelet transform (DWT), specifically a shift invariant implementation, for multiresolution image decomposition.

2. BACKGROUND

A brief summary of reported image fusion techniques that use multiresolution image pyramids or wavelets follows.

Burt [3] described an image fusion technique where a pair of coregistered images were first decomposed into multiresolution Gaussian, Laplacian image pyramids and the spatial details of the Laplacian images were combined, per pixel, on the basis of a “select the maximum amplitude” rule. Image reconstruction with the combined samples produced the fused result. The decomposition process is exactly reversible; each image can be recovered without error from its image pyramid representation.

Toet [4] described image fusion using multiresolution ratio image pyramids. Similar to Gaussian pyramid techniques, this process is reversible without loss. Decomposition was based on a ratio, in contrast to difference, between the image and a lowpass filtered image. Spatial details, as represented in the ratio images, were combined with a “maximum” selection rule.

Burt and Kolczynski [5] described image fusion based on gradient image pyramids. Two modes, averaging and selection, were used to combine directionally sensitive spatial details of the two respective gradient image pyramids. The mode determined for each pixel was based on the correlation between spatial patterns of the two gradient pyramids and computed from neighboring samples. A salience measure (local energy) determined which image to select.

Iverson and Lersch [6] reported image sharpening by using image pyramids and an adaptive technique for predicting, rather than selecting, higher resolution spatial details needed to increase spatial resolution. Relying on the local correlation between the Laplacian pyramid representation of the two images (P and XS), a multilayer feedforward neural network (NN) was trained to generate the spatial details of the lower resolution image given neighboring samples from the P image. At higher resolutions, output of the trained network produced the data for sharpening.

Aloisi and Grabit [7] described a method based on the discrete wavelet transform (DWT) for fusion of SPOT XS and higher resolution P images. Coregistered images were decomposed, and spatial details (as represented by DWT coefficients) needed to enhance the lower resolution image were estimated using an NN-implemented model to estimate XS coefficients. Similar to the technique of [6], an NN was trained, at lower resolution, to generate XS coefficients from neighboring P coefficients. The NN was then used to synthesize the higher resolution wavelet coefficients of the XS image.

Yocky [8] described image merging and data fusion with the DWT for two spatially registered images wherein wavelet coefficients of the lower resolution image were simply replaced by those of the higher resolution image.

Li, et al., [9] described image fusion using the DWT. The algorithm for combining wavelet coefficients used an area based activity measure around a center pixel (maximum absolute value within a window) to determine which wavelet image to select from. The selection was also subject to a consistency rule. They reported that their fusion results surpassed those for a point (or pixel) based maximum selection rule using either Laplacian pyramid or wavelet transform decomposition.

Recently, Rockinger and Fechner [10] reported the application of a shift invariant discrete wavelet transform (SIDWT) to the fusion of image sequences. A point based, select-the-maximum-amplitude rule was used to combine WT coefficient data. Their results, which were evaluated by using a mutual information quality measure, reported that the SIDWT fusion method outperformed those using Laplacian pyramid, gradient pyramid, and shift variant DWT.

There is some similarity between the work reported in this paper and that of Rockinger and Fechner [10] because (1) an SIDWT is also used, and (2) their very effective method for illustrating shift dependency of the DWT fusion process is repeated here. In contrast, the multispectral image fusion method described herein used a different implementation of the SIDWT, and the performance was evaluated with simulated TM resolution test data.

The remainder of the paper includes (1) a brief overview of the DWT and implementation of an SIDWT, (2) a description of the fusion technique for three XS bands of TM and the SPOT P data, (3) examples that illustrate the shift variance of DWT fusion, and (4) a comparison between DWT and SIDWT fusion performance using simulated TM sensor data.

3. DISCRETE WAVELET TRANSFORM

The following overview of the DWT is summarized primarily from Rioul and Vetterli [11] and Strang and Nguyen [12]. A starting point for describing the DWT is the two-channel perfect reconstruction filter bank of figure 1a, where $x(n)$ is an input sequence, and $g(n)$ and $h(n)$ are, respectively, impulse responses of lowpass (LP) and highpass (HP) convolution filters; ($\downarrow 2$) denotes downsampling by 2 (every other sample discarded) and ($\uparrow 2$) is upsampling by 2 (zeroes inserted between samples). Thus, the analysis filter convolution and downsampling of figure 1a is the inner product of sequence $x(n)$ and time reversed $g(n)$, $h(n)$. That is,

$$y_0(k) = \sum_n x(n) g(-n + 2k) \qquad y_1(k) = \sum_n x(n) h(-n + 2k).$$

The reconstructed signal $\hat{x}(n)$ is obtained by upsampling $y_0(n)$ and $y_1(n)$, filtering by $g'(n)$, $h'(n)$ and adding. For certain filters having perfect reconstruction properties, then reconstructed $\hat{x}(n)$ equals $x(n)$ within a possible shift. For example, for even length, L , filters and where analysis filters $h(n)$ and $g(n)$ are related (as mirror filters) by $h(L - 1 - n) = (-1)^n g(n)$ and the synthesis filters $g'(n)$ and $h'(n)$ are time reversed copies of $g(n)$ and $h(n)$, then $x(n)$ is reconstructed from

$$x(n) = \sum_{k=-\infty}^{\infty} [y_0(k) g'(-n + 2k) + y_1(k) h'(-n + 2k)].$$

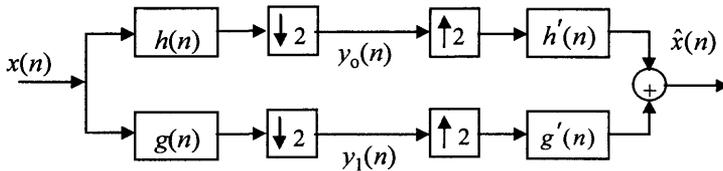


Figure 1a. Two channel filter bank

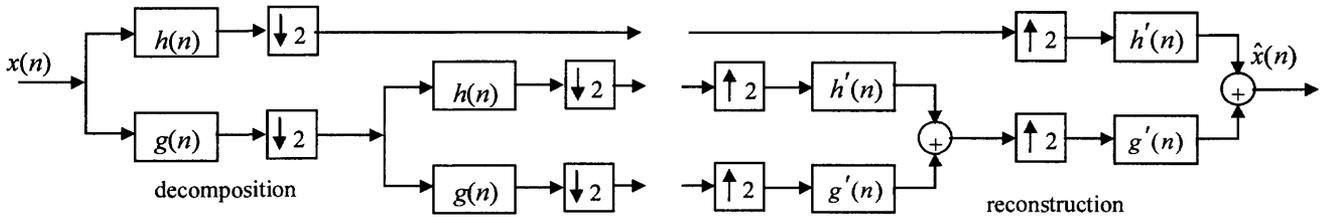


Figure 1b. Discrete wavelet transform

Figure 1. Two-channel filter bank and discrete wavelet transform.

Fig. 1b shows the filter bank tree for the fast DWT, and inverse DWT. Filters $g'(n)$ and $h'(n)$ are time reversed copies of $g(n)$ and $h(n)$; however, the specific requirements are as follows. In fig. 1b the two channel analysis (decomposition) filter bank process is iterated on the LP (and downsampled) output. If $g(n)$ and $h(n)$ are halfband LP and HP filters, each iteration halves the width of the lowband. The current highband spectrum corresponds to the difference between the current and previous lowband spectra. In addition, the time resolution is halved because of the subsampling. The relationship between wavelets and filters $g(n)$, $h(n)$ is [12]

$$\phi(t) = \sum \sqrt{2} g(n) \phi(2t - n) \quad \text{and} \quad w(t) = \sum \sqrt{2} h(n) \phi(2t - n),$$

where $\phi(t)$ and $w(t)$ are, respectively, the continuous-time scaling function and wavelet. The scaling function has the same general shape as the LP filters' impulse response $g(n)$. Development of the DWT needs only a discrete LP filter $g(n)$ that meets certain conditions [11]. As described in [12] [13], Daubechies constructed a family of orthogonal filters that generate orthogonal wavelets having compact support. The filters used in this work were for the Haar wavelet, and the 4 coefficient (Db₂) and 8 coefficient (Db₄) Daubechies wavelets.

3.1 Shift invariant discrete wavelet transform (SIDWT)

A major drawback of the DWT for multiscale image and signal analysis is the fact that transform is not shift invariant and consequently, the wavelet coefficients can change (in fact, dramatically) for minor shifts of the input signal [14]. In contrast, however, a translation of the input that is a multiple of all subsampling factors produces a simple translation of the transform coefficients. Because of this shift variance, pixel-level image fusion results based on the DWT will depend on the scene shift [10].

Shensa [15] described a method for removing the shift-variant effects of the DWT: compute the transform for all shifts of the input sequence. However, an alternative implementation is to replace the decimation (i.e., downsampling by 2) occurring at each level (scale) of decomposition by a split into odd and even indexed sequences, each of which becomes a starting point for the next level of decomposition [15]. The result is a nonorthogonal, overcomplete multiresolution decomposition that is shift invariant (SI.) Lang, et. al., [16] in their description of an SIDWT algorithm and the work of Shensa [15] note that when the downsampling operator, which retains even indexed samples, is applied to a sequence that has been shifted by one, odd indexed samples are retained. This property leads to an algorithm for an SIDWT decomposition.

In this report, implementation of the SIDWT was based on the two channel, 1D DWT filter bank tree decomposition (fig. 1b) applied to unshifted and shifted (by one) input sequences. This produces a split into even/odd indexed sequences as required by the alternative implementation for SIDWT [15]. For subsequent scales of decomposition, this process is repeated on both the odd and even sequences of the lowpass filter results. At each scale of the reconstruction, the odd and even sequence decompositions are each reconstructed separately (as shown by the fast DWT reconstruction of figure 1b) and the results averaged, after realignment to correct for the shift.

Application of the 1D DWT to rows and columns of an image gives the separable 2D WDT [11]. The shift invariance of this scheme is demonstrated later in this paper using the methodology of [10].

4. MULTISPECTRAL IMAGE FUSION PROCESS

The multispectral fusion process, with SIDWT, shown in figure 2 is for a three spectral band image and coregistered higher resolution P image. For color image interpretation, the fusion of XS (as an RGB false-color composite) and P should produce an enhanced image with minimal distortion of the original color. Visually perceived color properties of image features can often be better described in terms of intensity, hue, and saturation instead of red, green, and blue [2], [17]. Intensity (I), and the closely related value (V), hue (H), and saturation (S) refer, respectively, to the subject sensations of brightness, color, and color purity [2]. As shown in figure 3, the RGB image is first transformed to HSV; then V is enhanced by pixel-level fusion.

The RGB-HSV transform, which is based on a hexcone model, converts RGB values into a color space more suitable for color image processing. The intensity component, or roughly V (as V is the maximum of R, G, and B), can be contrast stretched, with no change in H and S of the resulting image [2].

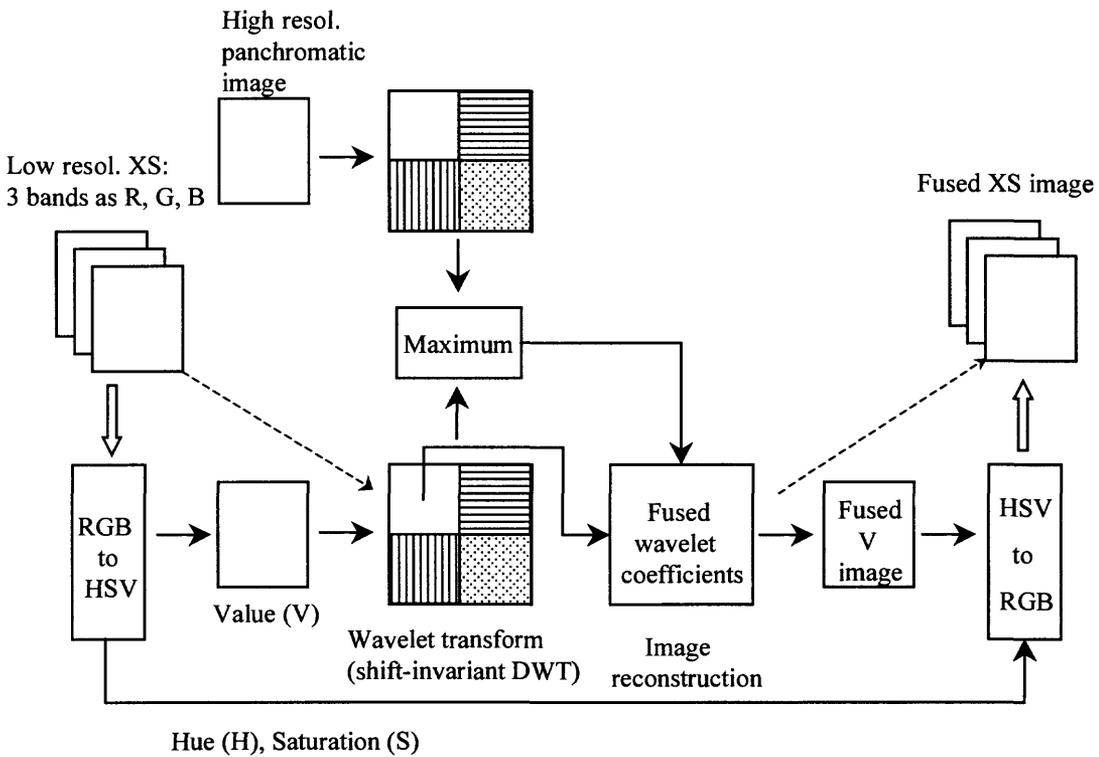


Figure 2. Multispectral (XS) image fusion process for shift invariant-discrete wavelet transform.

XS data are transformed to hue, saturation, and value. A “maximum” selection rule [3] is used to combine transform coefficients of the coregistered XS and higher resolution panchromatic image. The \dashrightarrow shows single band data used for a non-HSV fusion process that is described later.

The three spectral band TM data are first transformed from RGB to HSV. The fusion is then done for the V and higher resolution P image, after multiresolution decomposition using an SIDWT. Because this image pair (P, V) fusion process uses an SIDWT decomposition and pixel-based “maximum” selection rule for combining transform coefficients, it is similar to the fusion process of [10]. However, as previously noted, this report used a different implementation of the SIDWT and addresses multispectral images.

5. RESULTS

The shift dependency of the two fusion schemes (DWT and SIDWT) is illustrated in the error plots of figure 3. This analysis closely follows the informative example in [10]. For each fusion scheme, the RMSE is between the fusion result for the unshifted image pair and the fusion result, (after realignment) of a shifted (along rows) image pair. This test was for a single band of TM data (band 2) coregistered to a georeferenced SPOT P image (the HSV transform was omitted so as to simplify the test.) The P image, at 10-m ground sample interval (GSI), was produced by registering (cubic resampling) the SPOT “system corrected” (level 1A) data to a 1-m GSI color-infrared (CIR) digital orthophotoquad image [18]. Two wavelet filters, Haar and 8-coefficient Db_4 , and two fusion methods were used. The plots for SIDWT are coincident at RMSE equal to zero, and demonstrate the shift invariance of this process. With DWT fusion, the smoother wavelet has lower error. For three scales of decomposition, the period of the plots is $8 = 2^3$ because a translation of the input that is a multiple of all sub-sampling factors produces a simple translation of the transform coefficients [10].

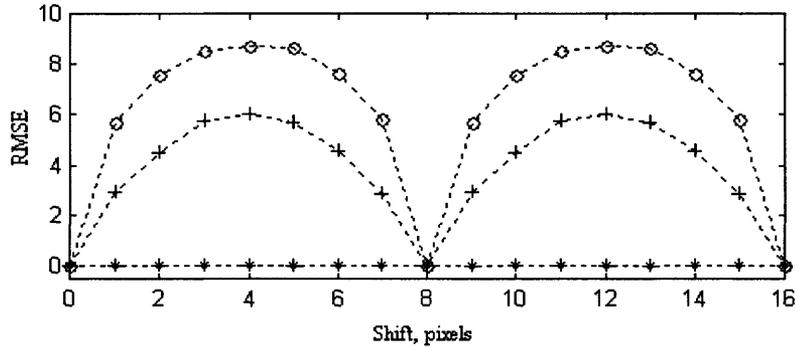


Figure 3. WT image fusion error versus image shift, in pixels.

The four plots shown are DWT fusion with the Haar (o) and Db_4 (+) wavelet and shift-invariant DWT fusion for Haar (.) and Db_4 (*) wavelet. The RMSE is between: (a) the fusion image for a nonshifted image pair, and (b) realigned result after fusion of the right-shifted image pair. Plots for the shift-invariant examples are coincident at zero error. There were three scales of wavelet decomposition.

Fig. 4 compares imagery for SI and SV fusion of the P, TM-2 data for the Haar wavelet. Artifacts are clearly visible in the contrast-stretched difference image.

5.1. Results for Simulated Data

The SIDWT and DWT results were compared using simulated high and low spatial resolution test imagery that was derived from a 1-m GSI, CIR digital orthophotoquad image. The spectral bands of the CIR data (as false color RGB) are visible green, red, and near-infrared [19]. A 10-m RGB reference image was produced by 10-by 10-pixel averaging and 10:1 downsampling; the coregistered P image was simply the average of the two visible bands. To simulate 30-m GSI imagery having the TM sensors spatial resolution, 1-m CIR data were convolved with the sensors' point spread function, separable and modeled at 1-m sample distance, and then 30:1 downsampled [2], [20].

Because of the TM sensors' scanning optics, image data are acquired in both forward and reverse scan directions, 16 lines (rows) at a time [1]. Thus, this test imagery simulates only a west to east scanned scene [22], and the consequences are believed minimal with respect to the objectives here - a relative comparison of SIDWT and DWT fusion methods. The 30-m and 10-m images were coregistered by an affine transformation with cubic interpolation and then fused. Results are given in table 1 as RMSE between fused and 10-m reference for three wavelets. The SIDWT based method produced the lower error, and this error decreases with the smoother wavelets. In part, the larger error for the near-IR (R) band is due to the lower correlation, as visible between P and R band images.

Wavelet Type	RMSE: SIDWT Fusion			RMSE: DWT Fusion		
	R	G	B	R	G	B
Haar	10.13	6.72	6.69	11.45	8.13	7.94
Db_2	9.61	5.91	5.88	10.54	7.06	6.89
Db_4	9.50	5.75	5.71	10.16	6.60	6.49

Table 1. Comparison of SIDWT and DWT Fusion Results for Simulated Data.

This table compares the RMSE between fused and true images for both SIDWT and DWT based methods; and three wavelet types. The lowest error (best fidelity) resulted from SIDWT fusion with the smoother wavelet (Db_4)

Results including the simulated TM data at 30-m IFOV are shown in fig. 5 for one (near-IR) band of the C-IR RGB image. The RMSE between the R, G, B bands of the reference and co-registered simulated TM resolution image was, respectively, 19.76, 19.55, and 18.38.

6. Conclusion

A shift invariant implementation of the discrete wavelet transform was described and applied to multispectral image fusion. The process makes use of a hue, saturation, and value color space transform to minimize color changes, and a reported point-wise “maximum” selection rule to combine transform coefficients. A comparison of the performance of SIDWT and DWT based fusion with simulated TM 30-m spatial resolution test data showed that the SIDWT technique produced imagery with fewer artifacts and lower RMSE between fused and the full-resolution reference. Image examples with TM and SPOT 10-m panchromatic illustrate the reduction in artifacts due to the SIDWT-based fusion.

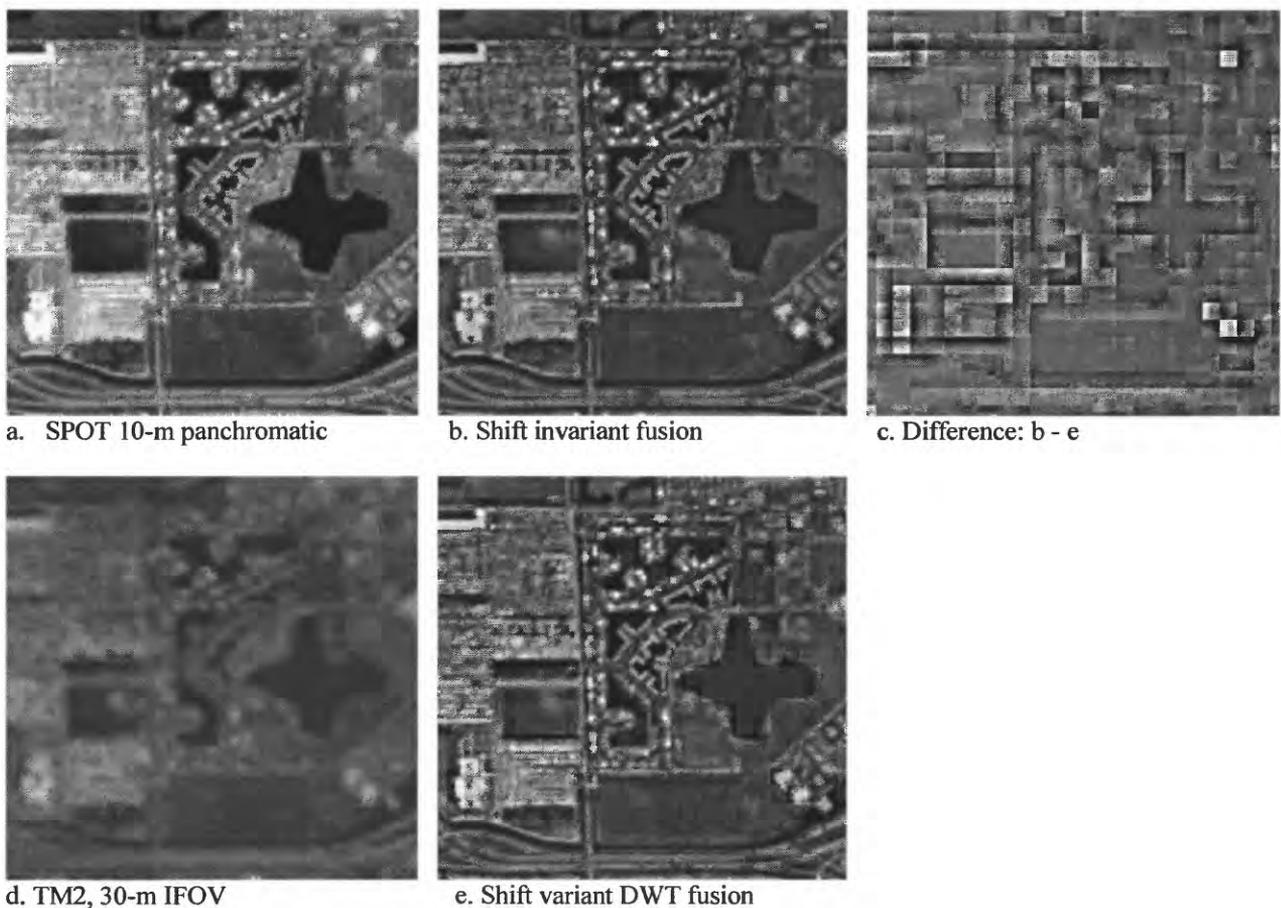


Figure 4. Comparison of Shift Invariant-DWT and DWT based fusion.

For the single band test process of figure 2, using Haar wavelet, b shows the result for shift invariant DWT and e represents the DWT fusion of coregistered images a and d. a is SPOT 10-m panchromatic data (level 1A), after registration to an exact 10-m base image. d is TM 30-m GIFOV visible band 2 data. c is the difference between b and e (with contrast stretch) and shows artifacts related to shift variant DWT fusion process; compare with e. b, d, and e have identical contrast stretch.



a. Panchromatic, 10-m



b. SIDWT Fusion of a, e.



c. Difference: true - fusion result b.



d. Simulated 30-m TM (near-IR band)



e. d coregistered to a.

Figure 5. Fusion of simulated TM and higher resolution panchromatic image.

b is one spectral band (near-IR) of the result of SIDWT fusion (fig. 2) of image a and the simulated TM resolution, three band image (one band shown in d). c is the difference, with contrast stretch to improve visibility, between b and a true, 10-m near-IR image (not shown).

ACKNOWLEDGMENT

I thank R.A. Schowengerdt for TM sensor PSF data and guidance in its application.

7. REFERENCES

1. J. A. Richards, "Remote Sensing Digital Image Analysis – Second Ed., Springer-Verlag, Berlin, pp. 75-88, 1992.
2. R. A. Schowengerdt, *Remote Sensing, Models and Methods for Image Processing*, Academic Press, San Diego, 1997.
3. P. J. Burt, "The Pyramid as a Structure for Efficient Computation," ch. 2, of *Multirate Image Processing and Analysis*, pp. 6-35, A. Rosenfeld, Ed., Springer-Verlag, New York, 1984.
4. A. Toet, L. J. vanRuyven, and J. M. Valetton, "Merging thermal and visible images by a contrast pyramid," *Optical Engineering*, v. 28, n. 7, pp. 789-792, 1989.
5. P. J. Burt, R. J. Kolczynski, "Enhanced image capture through fusion," Fourth Intl. Conference on Computer Vision, IEEE Computer Society, pp. 173-182, 1993.
6. A. E. Iverson, J. R. Lersch, "Adaptive image sharpening using multiresolution representations," in *Algorithms for Multispectral and Hyperspectral Imagery*, A. E. Iverson, Ed., Proc. SPIE 2231, pp. 72-83, 1994.
7. R. Aloisi, Y. Grabit, "Multispectral image resolution enhancement to improve efficiency of spectral analysis algorithms," in *Algorithms for Multispectral and Hyperspectral Imagery II*, A. E. Iverson, Ed., Proc. SPIE 2758, pp. 160-167, 1996.
8. D. A. Yocky, "Image merging and data fusion by means of the two-dimensional wavelet transform," *J. Opt. Soc. Am.*, v. 12, n. 9, pp. 1834-1845, 1995.
9. H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Models and Image Processing*, v. 57, n. 3, pp. 235-245, 1995.
10. O. Rockinger, T. Fechner, "Pixel-level image fusion: the case of image sequences," in *Signal Processing, Sensor Fusion and Target Recognition*, I. Kadar, Ed., Proc. SPIE 3374, pp. 378-388, 1998.
11. O. Rioul, M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, pp. 14-38, Oct. 1991.
12. G. Strang, T. Nguyen, *Wavelets and Filter Banks*, Wellesly-Cambridge Press, MA, 1996.
13. K. R. Castleman, "Wavelet Transforms," ch. 14 of *Digital Image Processing*, Prentice Hall, NJ, pp. 303-350, 1996.
14. E. P. Simoncelli, W. T. Freeman, A. E. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Information Theory*, v. 38, n. 2, pp. 578-607, 1992.
15. M. J. Shensa, "The discrete wavelet transform: wedding the A' Troun and Mallat algorithms," *IEEE Trans. Signal Processing*, v. 40, no. 10, pp. 2464-2482, 1992.
16. M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. Wells, "Nonlinear processing of a shift invariant DWT for noise reduction," in *Wavelet Applications II*, H. Szu, Ed., Proc. SPIE 2491, pp. 640-651, 1995.
17. A. R. Smith, "Color gamut transform pairs," Proc. SIGGRAPH '78, in *Computer Graphics*, v. 12, no. 3, pp. 12-19, 1978.
18. "Digital Orthophotos," USGS Fact Sheet FS-129-96, Jan. 1997.
19. J. R. Jensen, Author-Ed., "Urban/Suburban Land Use Analysis," ch. 30 of *Manual of Remote Sensing – Second Ed.*, v. 2, R. Colwell, Amer. Soc. Photogrammetry, Falls Church, VA, 1983.
20. Private communication; R. A. Schowengerdt, Oct. 1998.
21. C. Schuler, "Thematic mapper protoflight model line spread function," Proc. Seventeenth Int. Symposium on Remote Sensing of Environment, v. 3, pp. 1265-1275, 1983.