Perceptual-Based Image Fusion for Hyperspectral Data

Terry A. Wilson, Steven K. Rogers, *Senior Member, IEEE*, and Matthew Kabrisky, *Life Senior Member, IEEE*

***Abstract*— Three hierarchical multiresolution image fusion techniques are implemented and tested using image data from the Airborne Visual/Infrared Imaging Spectrometer (AVIRIS) hyperspectral sensor. The methods presented focus on combining multiple images from the AVIRIS sensor into a smaller subset of images while maintaining the visual information necessary for human analysis. Two of the techniques are published algorithms that were originally designed to combine images from multiple sensors, but are shown to work well on multiple images from the same sensor. The third method presented was developed specifically to fuse hyperspectral images for visual analysis. This new method uses the spatial frequency response (contrast sensitivity) of the human visual system to determine which features in the input images need to be preserved in the composite image(s) thus ensuring the composite image maintains the visually relevant features from each input image. The image fusion algorithms are analyzed using test images with known image characteristics and image data from the AVIRIS hyperspectral sensor. After analyzing the signal-to-noise ratios and visual aesthetics of the fused images, contrast sensitivity based fusion is shown to provide excellent fusion results and, in every case, outperformed the other two methods.**

I. Introduction

DEVELOPMENT of new imaging sensors created a need for image processing techniques that can fuse images from different sensors or from multiple images produced by the same sensor [1]–[7]. For example, the Air Force uses information from multiple sensors to obtain a multispectral signature of potential targets. The advantage of multispectral data is that it provides better target detection and identification than a single wide-band sensor and it allows flexibility in choosing a particular narrow-spectral-band for individual types of targets [4], [8]. The disadvantage with multiple sensors is that it is difficult and sometimes impossible to fully register the different input sources. Usually there are differences in scale, rotation, and shift between the outputs of each sensor. To overcome some of the multisensor problems, research is being conducted with single sensors that simultaneously collect data in several bands. The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), which is described in the next section, is an example of a sensor that simultaneously records information in hundreds of spectral bands [8]–[11]. However, there is a price to pay for the fully registered hyperspectral data. Two particular problems will be addressed in this article.

The first problem is caused by the large amount of information generated by hyperspectral sensors. For example, a single pass over a target area with an AVIRIS sensor produces 224 separate images; one for each of the 224 spectral bands measured. This requires approximately 140 megabytes of disk storage. However, only a fraction of that data provides unique or usable visual information about a specific area of interest. The rest is redundant information along multiple bands or is noise due to the atmosphere. Therefore, it would be beneficial to extract the relevant information/features from a set of hyperspectral images to form a reduced set of images.

The second problem is caused by the need to have human photo analysts analyze hyperspectral sensor data. Since human operators cannot physically or mentally integrate information from multiple source images [5]–[7], a method to fuse the relevant information into a single image, or at least a smaller subset of images, would alleviate this problem.

The scope of this article is to investigate image fusion algorithms that produce images intended for human analysis. The idea is to extract a composite image from a set of input images that increases the visual information to a human observer, or at least preserves the visually relevant information from the source images. Three methods are presented, one developed by Burt and Kolczynski [4], another by Toet [5]–[7], and the third is a new fusion method based upon the human visual system.

The algorithms are evaluated in two different ways. First the algorithms' performances will be evaluated by fusing a set of test images. Second, the algorithms' ability to fuse different spectral bands from the AVIRIS hyperspectral sensor will be evaluated. It is assumed that the images to be fused will either come from a single sensor that produces fully registered images, or from multiple sensors that have been preregistered. It is also assumed that the data produced from the sensors may be treated as image data and that contrast sensitivity, which is discussed in Section II-B, is the key for determining the important features of an image [12]–[15]. Furthermore, it is assumed that the key to multi-image fusion is to fuse based upon pattern primitives (groups of pixels) and not at the pixel level alone [1]–[7].

II. Background

A. AVIRIS Hyperspectral Data

AVIRIS is an imaging spectrometer that simultaneously collects spectral information in the visible to infrared ranges.

Manuscript received January 19, 1995; revised September 17, 1996. This research was supported in part by the Advanced Tactical Reconnaissance System Program Office (SPO), Wright Patterson AFB, OH, and The Space Object Identification SPO, Maui, HI.

The authors are with the Department of Electrical and Computer Engineering, Air Force Institute of Technology, Wright Patterson AFB, Dayton, OH 45433 USA (e-mail: twilson@afit.af.mil).

Publisher Item Identifier S 0196-2892(97)03676-0.

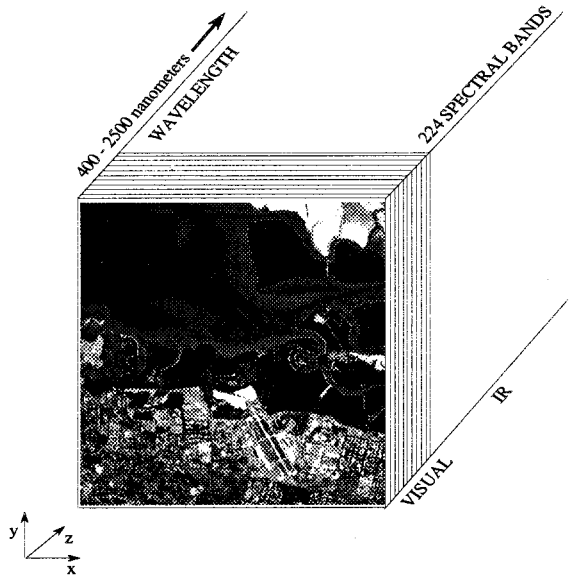

Fig. 1. Hyperspectral image cube representation taken from the AVIRIS sensor.

It records information in 224 spectral bands that range from 400 ηm to 2500 ηm in approximately 10 ηm increments with a pixel resolution (instantaneous field of view) of roughly 20 m at operational altitude. The AVIRIS sensor is a "whisk broom" type spectral sensor that collects 614 pixels (a swath width over 10 km wide) of data with each sweep. The number of lines of data that can be collected over an area is only limited by the amount of on-board data storage. The AVIRIS data presented here are all 512 lines. Thus, a single image represents visual information about an area of ground that is roughly 126 km^2 [8]–[11].

Creating images from the hyperspectral data is a two step process. First, a spectrometer measures the electromagnetic energy reflected or emitted from a surface. The intensity values produced by the spectrometer are stored as pixel values that represent a single intensity value associated with a physical location in the target area in the band being sampled. The pixel values are used to form an image cube, where the X and Y axis of the cube represent the pixel location, and the Z axis represents the spectral band or associated wavelength (Fig. 1). The hyperspectral data used in this article is radiometrically calibrated radiance data that was provided by the Jet Propulsion Laboratories (JPL) Pasadena, CA. The particular scene that was used is from an AVIRIS sensor that was flown over Moffett Field, CA.

The image, on the face of Fig. 1, is the picture from band 30. Band 30, which represents the spectral information in the 600 ηm to 680 ηm bandpass range, clearly shows the unique reflectance for the different types of materials: land, water, runway, etc.

B. Hierarchical Image Fusion Techniques

The two image fusion techniques currently proposed in the literature, which will be implemented in this article, are described here. The first is a method proposed by Toet [5]–[7]. Toet's method is a multiresolution pyramidal technique that uses the maximum contrast information in the Ratio of Low Pass (ROLP) pyramids to determine what features are salient (important) in the images to be fused [5]–[7]. The input images are converted into a multiresolution ratio pyramid in which the values in the pyramid represent what Toet calls the contrast details, where contrast is described in the next section. Then the ratio pyramids are compared point-by-point and the maximum value is retained for the composite pyramid. Once the composite ratio pyramid is formed the process of creating the ratio pyramid is reversed to recover the reconstructed composite.

Toet defends the decision to select the details for the composite image based upon maximum contrast by arguing that human vision is based upon contrast and that, by selecting details with maximum contrast, the resulting fused image will provide better details for the human analyst. Although this is sound reasoning based upon the desire to present the human analyst with the best visual image, it does not account for the fact that a noisy image is typically of higher contrast than an image that is not. Therefore, Toet's method would select the noisier parts of the images to be retained in the composite, which presents a potential loss of information about desired targets. Thus, a method for selection that is based upon the perceptual sensitivity of the human visual system and not just pure contrast is needed.

The second method is a technique developed by Burt and Kolczynski [3], [4]. Burt and Kolczynski's method is also a multiresolution pyramidal technique, but unlike Toet's method, Burt and Kolczynski have a more complex criteria to determine the salient features in the input images to be fused. Burt and Kolczynski use a match and saliency metric to determine which details are salient. The match value compares two regions in the input images and provides a number that describes if they are similar or match and the salience metric measures the energy in the oriented gradients of the input images. If the input image details are similar (within some threshold) then the details are averaged. If the input images are not similar enough (beyond the threshold) then the details from the image that are more salient are retained for the composite gradient pyramid. Once the fused gradient pyramid is formed, the reconstruction is performed to recover the composite image.

The fusion algorithm proposed by Burt and Kolczynski has several advantages over the method proposed by Toet. Since it averages similar input sources, instead of just picking some maximum value, it offers a potential for better noise reduction. It also allows the low contrast details to be preserved, if they are the salient features. The main disadvantage is that a template (weight matrix p) is needed to decide which features are salient. Since there are always problems with size, orientation, translation, etc., finding a template that will work well as a salient measure, will be very difficult if not impossible. One possible weight matrix p, proposed by Burt and Kolczynski, is a 3×3 matrix of ones; this allows the saliency to be based upon the local energy in the details. Although this provides a measure that has some useful applications, local energy in the details of some images may reflect a high value strictly due to noise in the image. Also,

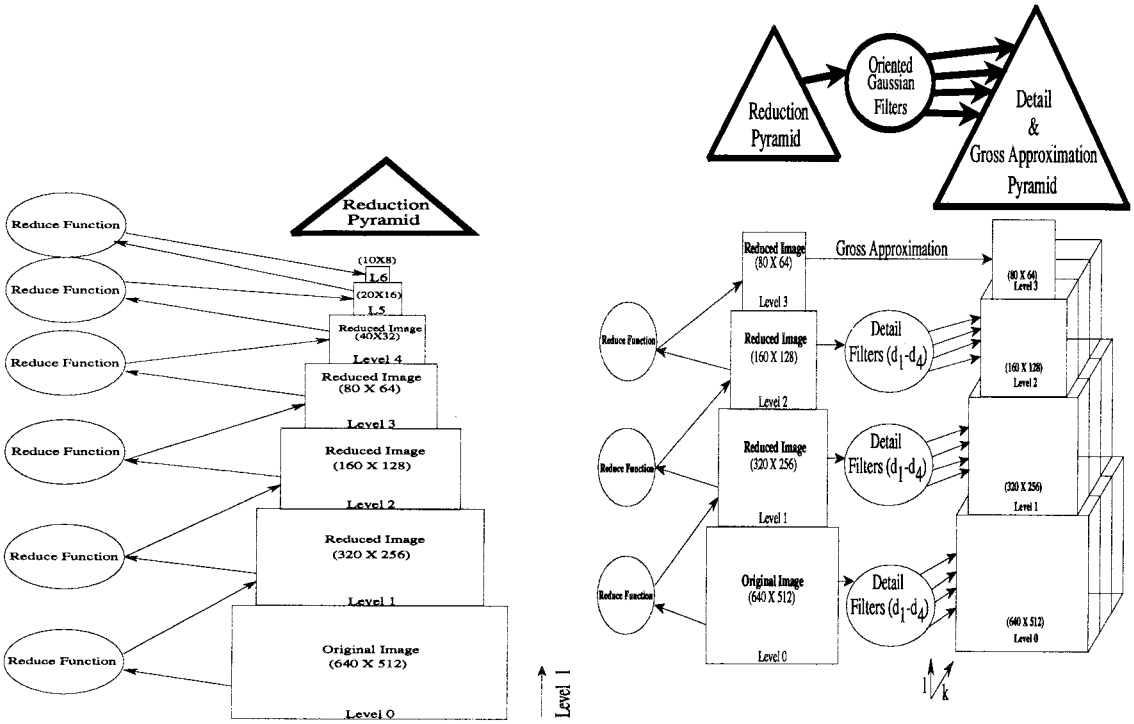

Fig. 2. Stage I of the three stage multiresolution pyramid algorithm.

energy in the images that may not fall within human perception may play a large part in the decision of how the images will be fused. Therefore, a method that uses the strengths of Burt and Kolczynski's fusion scheme, and still addresses the human visual system, is needed.

A third method for image fusion, which relies on the frequency response (contrast sensitivity) of the human visual system, is developed and presented in the next section. Mathematically, contrast sensitivity (which is a measure of how a person responds to contrast at threshold) is defined as the reciprocal of contrast [13], [16], where contrast is a measure of the difference in brightness across an image or scene.

The measure of how the human visual system responds to contrast, i.e., contrast sensitivity, is a function of spatial frequency [13], [16]. For example, it has been shown that the human visual system responds better, or is more contrast sensitive, to low spatial frequency components than it is to high spatial frequency components [16]. For instance, contrast sensitivity peaks for the frequency ranges from 2 to 10 cycles per degree (cpd) and then falls off sharply [13], [16].

Because standard human visual acuity tests only measure the high frequency response, contrast sensitivity is becoming a more acceptable measurement criteria. For instance, Stager and Hameluck showed that contrast sensitivity was better in determining a person's ability to perform air-to-ground search than standard visual acuity tests [17].

The contrast sensitivity response used here is from the 95 percentile curve [16]. An experimentally derived contrast sensitivity function (CSF) of the human visual system developed by Mannos and Sakrison [18] is also available. While some research shows that an individual's perception of contrast will be different for contrasts that are above threshold [12], [14], [15], the fusion method proposed here is based upon the threshold frequency response.

III. Hierarchical Image Fusion with Contrast Sensitivity Saliency

This section describes an image fusion technique based upon previously successful multiresolution decomposition and reconstruction methods [1], [4]–[7], [19], with an added ability to tailor the selection criteria (i.e., what is salient between images) to the contrast sensitivity of the photo analyst. Another method, which is similar to the approach here, but which is based upon a Daubachies wavelet decomposition/reconstruction is presented in [20]. The basic method employs a multiresolution algorithm that uses three stages: decomposition, fusion, and reconstruction.

A. Image Decomposition

The first step in image decomposition is constructing the reduction or Gaussian pyramid. The reduce function, shown in Fig. 2, is defined as a filter and down-sample. First the 5×5 Gaussian kernel w shown below, which is the same kernel used by Burt and Kolczynski [4], is used to filter an input image.

$$w = \dot{w} * \dot{w} = \frac{1}{256}\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \quad (1)$$

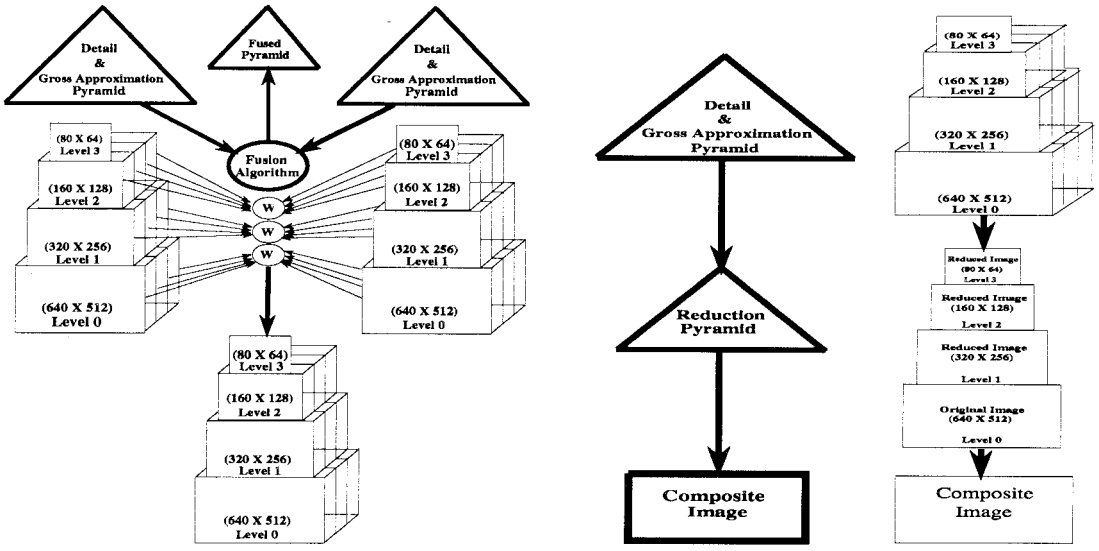

Fig. 3. Stages II and III of the three stage multiresolution pyramid algorithm.

$*$ is the convolution operator and

$$\dot{w} = \tfrac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \tag{2}$$

Next the filtered image is down-sampled by a factor of two by selecting every other point in the filtered image. The filtering operation is simply a convolution of the input with the Gaussian kernel w. The output of the reduce function is used as the input to the next stage of reduction. By successively filtering and down-sampling, an image pyramid that has the original image as the pyramid base with successive levels that are low-pass filtered and down-sampled versions of the level below is generated, see the **Reduction Pyramid** in Fig. 2.

The next step in image decomposition is performed by extracting the orientation gradient details and a gross approximation from the multiresolution pyramid. The orientation gradient details are extracted by using the filters, defined below, on each level (except the top) of the reduction (Gaussian) pyramid created in the previous step (Fig. 2). The top level is kept as the gross approximation. Burt calls this step creating the orientation gradient pyramid [4]. It is called the orientation gradient because the basis functions used for detail extraction are gradients of Gaussian patterns. The gradient filters d_1 through d_4, defined below, are used to extract information from the reduced (Gaussian) pyramid. The following equations define this stage of the decomposition:

$$D_{kl} = d_l * [G_k + \dot{w} * G_k] \tag{3}$$

$$d_1 = \begin{bmatrix} 1 & -1 \end{bmatrix}$$

$$d_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$$d_3 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$d_4 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4}$$

where $*$ is the convolution operator, D_{kl} are the details for level k and orientation l, G_k is the level k input from the reduced image pyramid, and d_1 through d_4 are the oriented gradient filters. The index k is the level of the resolution from the Gaussian pyramid and the orientation l is simply the index of the d_1 through d_4 filter used.

B. Detail Pyramid Fusion

Now that the oriented gradient pyramid has been formed for each input image, Stage II (detail fusion) is performed (Fig. 3). This is where the main difference lies between the contrast sensitivity method proposed here and Burt and Kolczynski's method [4]. Burt uses a match and saliency measure, which is based upon the weighted energy in the detail domain, to decide how the oriented gradient pyramids will be combined. The contrast sensitivity method presented here uses an idea similar to Burt and Kolczynski's match and saliency, but instead of using localized energy in the detail domain to compute the weighted averages, it uses a weighted energy in the perceptual domain; where the perceptual domain is based upon the frequency response (i.e., contrast sensitivity) of the human visual system.

When evaluating whether two images would be perceived as different (determining a match value), and deciding which details are more important (determining a saliency value), we need to base the evaluation on the perceptual model of the human observer. Thus, the criteria to decide which parts of the input images will be averaged to form the composite image and which ones will not is determined by the contrast sensitivity response of the human analyst.

The fusion stage is performed in the following manner:

1) Corresponding neighborhoods from the orientation gradient pyramids to be fused will be compared using the contrast sensitivity response.
2) If the corresponding neighborhoods differ by more than some threshold, the one that is more salient (i.e., has

more energy in the contrast sensitivity weighted frequency domain or "perceptual domain") will be retained in the fused orientation gradient pyramid.

3) If the perceived difference is less than some threshold, the two neighborhoods will be summed according to a weighted average.

First a relative perceptual contrast D is computed for each layer of the detail pyramid.

$$D = \frac{S_A - S_B}{S_A + S_B} \tag{5}$$

where D is the percent difference and S_A, S_B are the saliencies computed for a neighborhood from level k and orientation l of the detail pyramids A and B, respectively. Saliency is computed as the amount of perceptual energy in a given set of details, as related to the frequency response of the human visual system [16]. Saliency is defined by

$$S = \sum_{m,n} C(m, n) X_I(k, l, m, n) \tag{6}$$

where C is the contrast sensitivity weight matrix representing the two-dimensional median response of the human visual system [16], [20], [21] and X_I is the magnitude of the energy normalized low frequency two-dimensional Fourier components from some neighborhood at level k and orientation l of the detail pyramid. The indices m, n are defined by the desired window size.

The method for extracting the saliency is to first energy normalize each level and orientation of the detail pyramids by dividing by the square root of the sum of the squares at each level and orientation. Next, extract the magnitude of the Fourier coefficients of an m, n sliding window over each input level and orientation. The size of the window and the step size may vary. A 40 $\times$ 40 window and a step size that varied with the resolution was used. To extract the Fourier coefficients, the 40 $\times$ 40 window is passed over the image, according to the step size, and the Fourier transform is computed for each. Then, the frequency components are energy normalized.

After the percent difference D is computed, it is compared to some threshold T. If the difference is greater than T, the input detail with the higher saliency value will be retained for the fused detail pyramid and the other detail will not. If the perceived difference is not greater than the threshold T, the input details will receive the following weights:

$$\begin{aligned} W_A &= \left(1 - \frac{S_B}{S_A} \times 0.5\right) \\ W_B &= 1 - W_A. \end{aligned} \tag{7}$$

The fused detail pyramid is created by summing the weighted details from image A and image B (Fig. 3). The weighted sum is defined as follows:

$$D_C = W_A D_A + W_B D_B \tag{8}$$

where D_C is the fused detail, W_A and W_B are the appropriate weight matrices (7), D_A and D_B are the details from the detail pyramids (3) from image A and B (7). The weighting is performed by a point-by-point multiplication of the weight matrix and the detail matrix (not by the typical matrix multiplication operation).

C. Fused Image Reconstruction

Once the gross approximation and detail pyramids have been fused into a single pyramid, stage III is performed. This is the reconstruction phase. Reconstruction is performed by combining the four layers of details into a single layer and then combining the gross approximation with the different levels of details (Fig. 3), to form the composite image. The order of reconstruction is:

1) Convert the oriented gradient pyramid into the second derivative pyramid, or what Burt and Kolczynski call the oriented Laplacian pyramid [4].
2) Convert the oriented Laplacian pyramid into the FSD (filter-subtract-decimate) Laplacian pyramid [2].
3) Convert the FSD Laplacian pyramid into the RE (reduce-expand) Laplacian pyramid.
4) Convert the RE Laplacian pyramid into the Gaussian pyramid.
5) Recover the composite image from the Gaussian.

Converting the oriented gradient pyramid into the oriented Laplacian pyramid is represented by

$$\vec{L}_{kl} = -\tfrac{1}{8}\, d_l * D_{kl} \tag{9}$$

where $\vec{L}_{kl}$ is the level k and orientation l of the oriented Laplacian pyramid, d_l is one of the detail filters described above for d_1 through d_4, and D_{kl} is the input level k and orientation l from the oriented gradient pyramid.

Converting the oriented Laplacian pyramid into the FSD Laplacian pyramid is represented by

$$L_k = \sum_{l=1}^{4} \vec{L}_{kl} \tag{10}$$

where L_k is the level k of the FSD Laplacian pyramid.

Converting the FSD Laplacian pyramid into the RE (reduce-expand) Laplacian pyramid is defined as

$$\tilde{L}_k \approx [\delta + w] * L_k \tag{11}$$

where $\tilde{L}_k$ is the level k of the reduce-expand (RE) Laplacian pyramid, w is the 5 $\times$ 5 Gaussian filter (1), and δ is defined as

$$\delta = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{12}$$

Converting the RE Laplacian pyramid into the Gaussian pyramid and then recovering the composite image from the Gaussian, is performed by repeatedly applying

$$\hat{G}_k = \tilde{L}_k + 4w * [G_{k+1}]_{\uparrow 2} \tag{13}$$

where $\hat{G}_k$ is the level of reconstruction and $[G_{k+1}]_{\uparrow 2}$ represents an up-sampling of the layer $k+1$ above to match the resolution of the current layer L_k of the RE Laplacian pyramid. If the formula has been applied enough times to successfully

TABLE I
SIGNAL-TO-NOISE RATIOS (SNR) FOR THE BAND 30 MODIFIED TEST IMAGES

Image Name	Correlated/Uncorellated	SNR of Affected Areas	SNR of Overall Image
band30_ 5dbcorr1	correlated	3.16:1	128.8:1
band30_ 5dbcorr2	correlated	3.16:1	95.5:1
band30_ 5dbcorr3	correlated	3.16:1	33.9

recover $\hat{G}_0$, it will be the reconstructed composite image. The up-sampling (**expand** function) is represented by

$$P_{l,k}(i,j) = 4 \sum_{m,n=-2}^{2} w(m,n) P_{l,k-1}\left(\frac{i+m}{2}, \frac{j+n}{2}\right) \tag{14}$$

where $P_{l,k}$ represents the expanded output level, $P_{l,k-1}$ represents the input level, and only integer indexes contribute to the sum. The expand function is accomplished by padding every other column and row with zeros and then convolving the zero padded image with the weight matrix w. In effect, an up-sampling by a factor of two is being accomplished.

D. Test Images

In order to evaluate the capabilities of the three fusion algorithms presented, multiple test images are generated with varying signal-to-noise ratios (SNR's) and types of backgrounds. The SNR's in decibels (dB) were computed as

$$\text{SNR} = 10 \log_{10}\left(\frac{Energy_{signal}}{Energy_{noise}}\right) \tag{15}$$

where $Energy_{signal}$ is the sum of the squares of the pixel values in the original image and $Energy_{noise}$ is the sum of the squares of the Random noise. This yields a SNR in dB. The corresponding SNR in terms of ... : 1 is computed from the dB value as

$$\text{SNR} = 10^{(\text{SNR in db}/10)}. \tag{16}$$

To create the test images, noise with a uniform distribution on the interval (0.0, 1.0), is scaled by a constant and is then added to various locations in the images. The image used was from band 30 from an AVIRIS data set acquired at Moffett Field, CA (see Fig. 1). The Moffett field image was chosen because it contains natural scenes (water, trees, fields) and urban structure (streets, runway, buildings). Band 30 was simply chosen arbitrarily.

Correlated or uncorrelated noise was added to the test images in three different regions (target types), thus creating sets of three images to be fused. The correlated noise was generated by convolving the uncorrelated noise with a low-pass filter [22]. The level of correlation was computed as the full-width-half-max of the maximum amplitude of the autocorrelation of the low-pass filter. The full-width-half-max value used here was 4 pixels. Fig. 4 provides an example of three images containing correlated noise with a SNR of 3.16:1 (5 dB), within the area the noise was added. Table I provides a listing of the SNR's for the test images using band 30.

TABLE II
SIGNAL-TO-NOISE RATIOS (SNR) FOR THE FUSED TEST IMAGES CONTAINING CORRELATED NOISE

Fusion Method	Image Group	SNR of Overall Image After Fusion
Burt and Kolczynski	band30_ 5dbcorr	39.8:1
Toet	band30_ 5dbcorr	20.9:1
Contrast Sensitivity	band30_ 5dbcorr	147.9:1

IV. RESULTS

A. Hierarchical Image Fusion of AVIRIS Test Images

In this section, the results of fusing the test images described in the previous section are presented. The test images were fused using the algorithms by Burt and Kolczynski [4], Toet [5]–[7], and the contrast sensitivity algorithm.

One set of results are presented here for visual reference. Fig. 5 represents the results of fusing the three images shown in Fig. 4 that use correlated noise. The overall fusion results are presented in tabular form in Table II.

All of the image fusion that is reported in this article was implemented in the following manner. Burt and Kolczynski's fusion algorithm was implemented using the following parameters:

1) Six layers of decomposition (i.e., five levels of details) using Burt and Kolczynski's recommended w matrix and d_1 through d_4 filters [4].
2) A 3×3 p matrix of all ones.
3) An alpha of 0.9 [4].

Toet's fusion algorithm was implemented using six layers of decomposition and Toet's recommended weight matrix w [5]–[7].

The method presented here was implemented using the following parameters.

1) Six layers of decomposition (i.e., five levels of details) using Burt and Kolczynski's recommended w matrix [4].
2) A window size of 40×40.
3) A shift of 8.
4) A threshold of 0.20.

Fig. 5 and Table II show that the method presented here does a better job of de-emphasizing both uncorrelated or correlated noise in the input images than either Burt and Kolczynski's or Toet's methods. In most cases, the method presented in this article even increased the SNR of the fused results above that of any input image. However, even though the fused SNR was better than either Burt and Kolczynski's

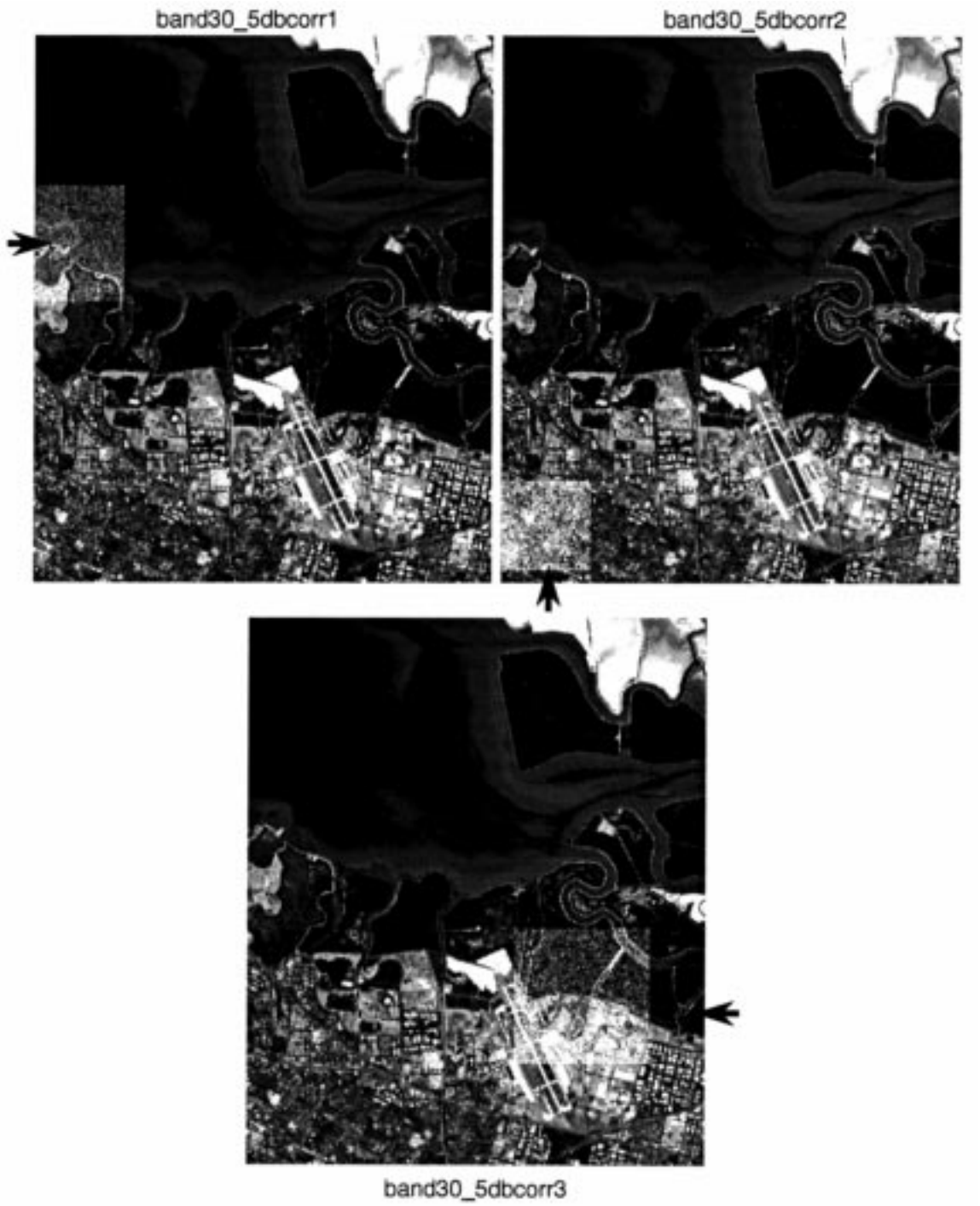

Fig. 4. Three AVIRIS test images of band 30 with correlated noise added at different locations in each test image.

or Toet's results, the fusion method developed here did not improve the SNR of the fused band 30 images over the SNR of the band 30 input images. Two possible explanations for the differences in the results are

1) The frequency of the added noise and the frequency content of the band 30 images combined to provide enough contrast sensitivity weighted energy to make that area more salient.
2) The reconstruction error was larger due to the higher energy content of the band 30 images, thus creating a larger difference between the fused composite and the ideal band 30 image.

It is also important to note that just because two images differ in Euclidean distance, they may not be perceived as different. How well changes in local contrast can be perceived is a function of the contrast sensitivity [12], [13], [16].

Comparing Burt and Kolczynski's method to Toet's shows that Burt and Kolczynski's method provided better noise reduction, but it still weighted the noisy parts of the images as more salient than the non-noisy parts. It can also be seen from

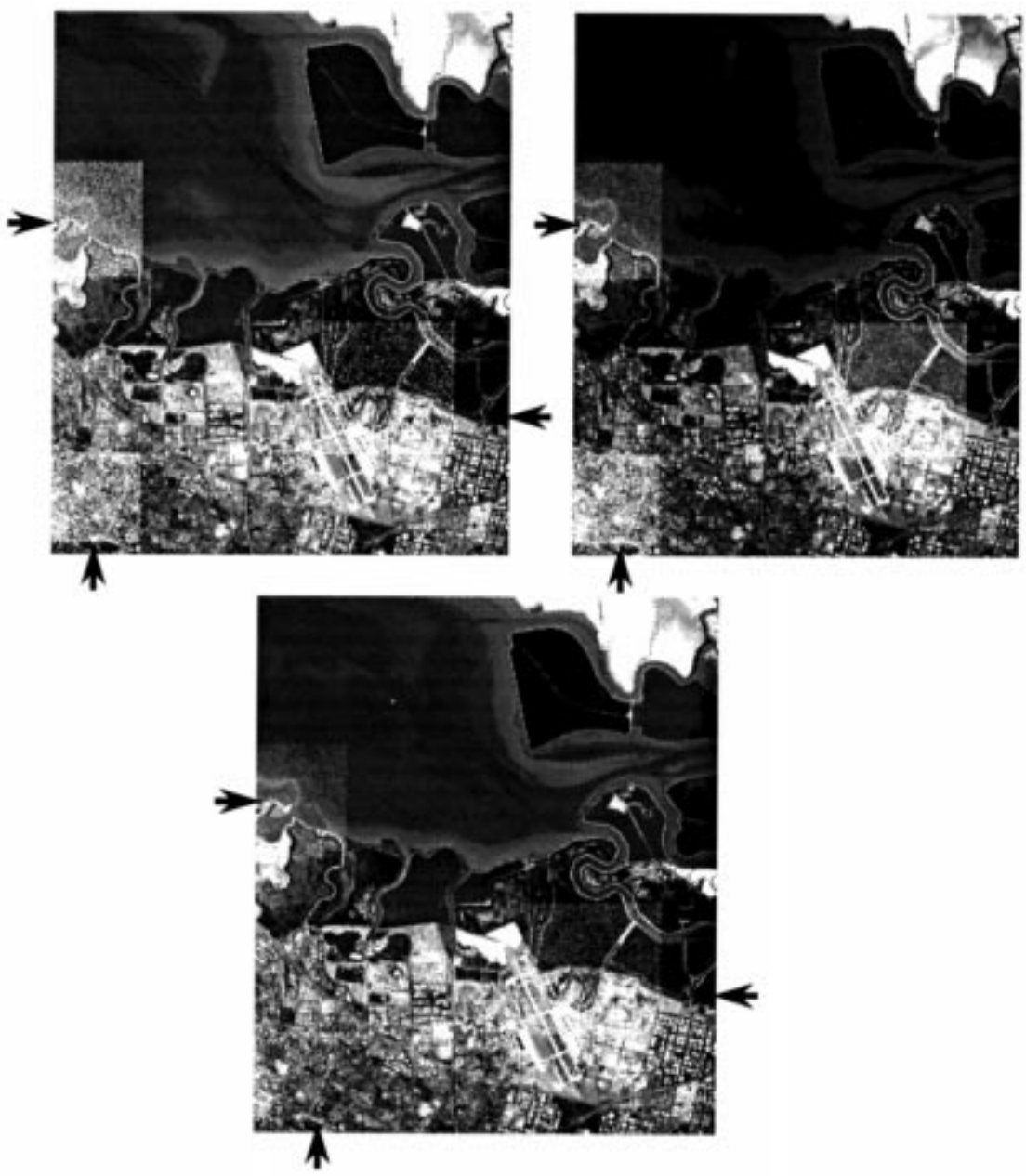

Fig. 5. Fusion results of test images using correlated noise added to band 30.

the fusion results that Toet's method selected the high-energy noisy parts of the input images to retain in the composite, even though the lower frequency parts of the input images contained the information desired. This loss of information in the input images, due to selecting the high energy noisy parts of the input images demonstrates the limitation of using pure contrast as a selection criteria. The contrast sensitivity method overcomes the previous methods' sensitivity to noise.

B. Hierarchical Image Fusion of IR and Visible Bands of AVIRIS Image Data

In this section, the results of fusing three bands of image data from the AVIRIS hyperspectral sensor are presented. The three bands include one image from the visible frequency range and two from the infrared. Specifically, they are band numbers 30, 60, and 90, which are centered at 677 ηm, 937 ηm, and 1225 ηm, respectively. Fig. 6 displays the three original (no added noise) AVIRIS input images.

It can be seen by comparing the fused images in Figs. 7–9 with the input images in Fig. 6 that all three methods do a good job of preserving visual information from each input image. However, comparing the three fused results to each other, Toet's fusion method does not have the same amount of detail that is present in the other two. For example, looking at the upper right hand corner of the fused images, it is easy to see that Toet's method causes the details to become washed out. In order to make the details in this corner more distinguishable, the overall intensity level of the image has to be reduced. The reduction in intensity then causes the other areas of the image to become less distinct. Also the runway and surrounding area

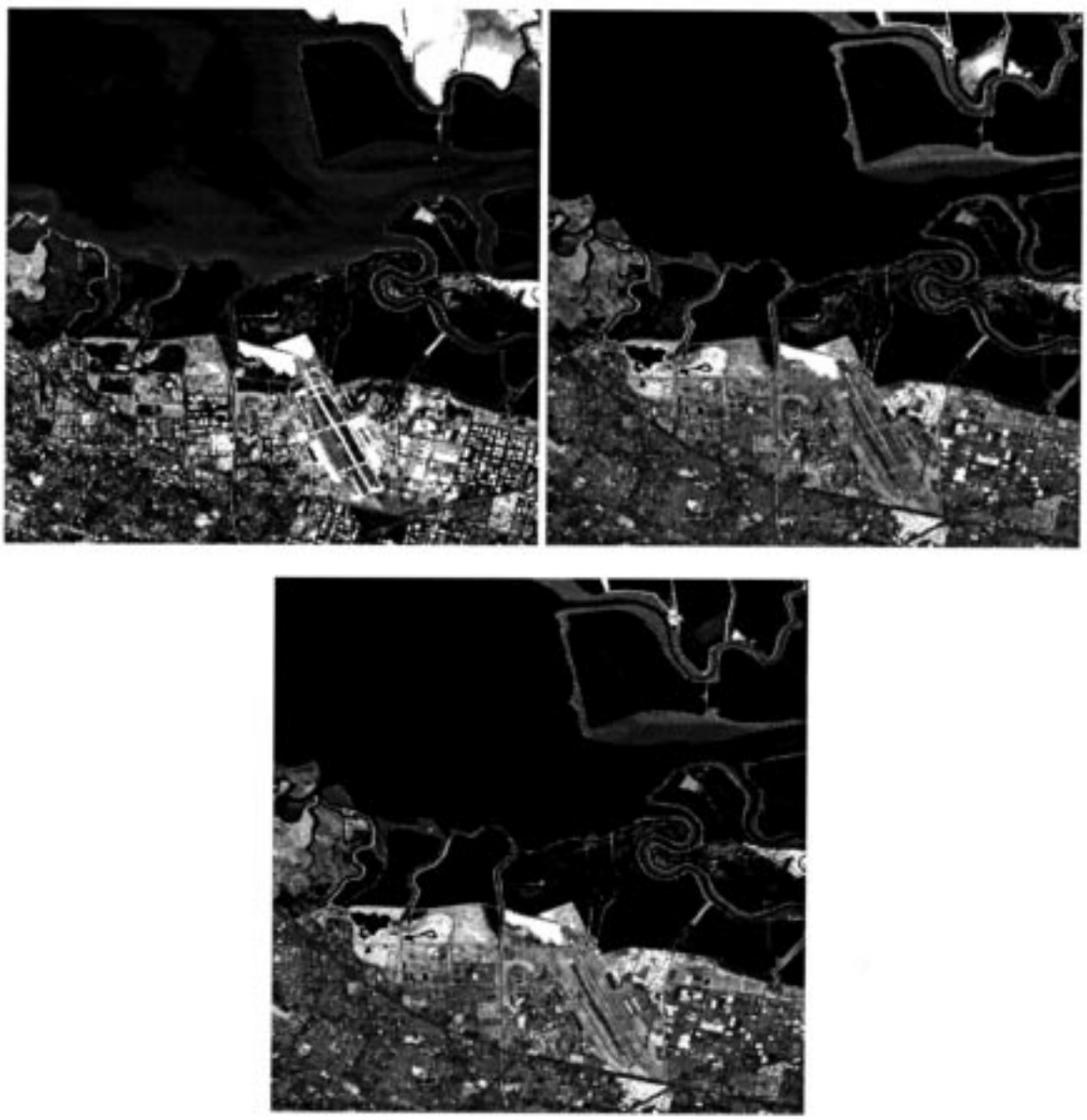

Fig. 6. AVIRIS images representing bands 30, 60, and 90.

is more clear in Burt and Kolczynski's method and the contrast sensitivity method than it is in Toet's fusion method.

Comparing the fusion results of the contrast sensitivity method with Burt and Kolczynski's shows that they both have similar characteristics. Each method appears to preserve features in the input images that are dominant. An example of this preservation is observed by looking at the road that is a dominant feature in bands 60 and 90 but not in band 30. The road is located in the lower third of the image and extends from the left edge of the image all the way to the right. It can clearly be seen as a continuous road in bands 60 and 90, but it is hard to distinguish in band 30. Looking at the fused results in Figs. 7 and 9 it can be seen that the road is preserved in the composite. Overall, the contrast sensitivity method provides better results, both aesthetically and numerically (SNR), to either Burt and Kolczynski's or Toet's methods.

An example of how fusion of the input bands provides better detail in the composite than in the individual input images alone is also shown in the fused images of all three methods. The airport that can be seen in the lower right quadrant of the input images has been combined in the composite image to provide more detail than was in all three of the input images separately.

C. Hierarchical Image Fusion of Multiple Bands of AVIRIS Image Data

The fusion method developed in this article can also successfully fuse many bands from the AVIRIS sensor without causing loss of information or loss of dynamic range. The input images are from bands 30 through 40 from the AVIRIS hyperspectral sensor. The combined band-pass range of the images is from 670 ηm to 750 ηm. Fig. 10 shows the results of fusion.

Fig. 7. Burt and Kolczynski fusion results of AVIRIS bands 30, 60, and 90.

Fig. 9. Contrast sensitivity fusion results of AVIRIS bands 30, 60, and 90.

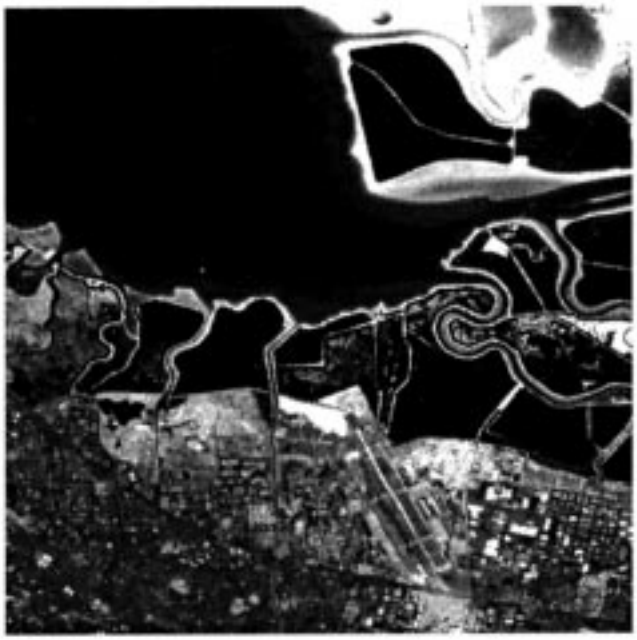

Fig. 8. Toet fusion results of AVIRIS bands 30, 60, and 90.

Fig. 10. Contrast sensitivity fusion results of AVIRIS bands 30 through 40.

V. Conclusions

The fusion of various combinations of AVIRIS data, presented in the previous section, provides some examples of how the fusion algorithms perform. The number of bands and which bands are chosen will be a function of the desired information to be obtained.

When human observers have to analyze many images from a hyperspectral sensor they can quickly become inundated with too much information. Also, many times it is helpful to combine information from several spectral bands to get a more complete picture of a scene. This is where the fusion algorithms can help. They can be used to combine the desired set of bands in order to present the photo analyst with a reduced set of images to analyze. This decreases the data for the analyst and increases the information in a given image. An added benefit of the contrast sensitivity based fusion method is that it can be tuned to the contrast sensitivity of a particular photo analyst. This ensures that details detectable by the analyst in the source images will be maintained in the composite.

An important point needs to be made about the distinction between the fusion methods presented here and the principle

components (PC) techniques that are often used to reduce image data sets. The method presented here relies on a distance metric in the perceptual space and not in the Euclidean space as does PC. In other words, contrast sensitivity based fusion is more concerned with how a human observer would perceive the differences in two images and uses this to decide on whether to average image details or to pick details from one and discard the other. It is not trying to minimize some mean squared error in Euclidean space as the PC does. The resultant fused images are designed to reduce the number of images that a human analyst has to analyze. They are not intended for image compression.

Overall, the results of the previous section show that contrast sensitivity based fusion is an excellent means of combining multiple images from the AVIRIS hyperspectral sensor for human analysis. Contrast sensitivity based fusion results in improved signal-to-noise ratios over a wide range of noise levels and backgrounds, and the results are independent of the type (correlated versus uncorrelated) and strength of the added noise.

REFERENCES

[1] A. Akerman, "Pyramidal techniques for multisensor fusion," in *Proc. SPIE Int. Soc. Opt. Eng.*, 1993, vol. 1828, pp. 124–131.

[2] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, pp. 532–540, 1983.

[3] ———, "Merging images through pattern decomposition," *SPIE Appl. Digital Image Proc.*, vol. 575, no. 3, pp. 173–181, 1985.

[4] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *1993 IEEE 4th Int. Conf. Comp. Vision*, vol. 4, pp. 173–182, 1993.

[5] A. Toet, L. J. van Ruyven, and J. M. Valeton, "Merging thermal and visual images by a contrast pyramid," *Opt. Eng.*, vol. 28, no. 7, pp. 789–792, 1989.

[6] A. Toet, "Hierarchical image fusion," *Machine Vision Appl.*, pp. 1–11, 1990.

[7] ———, "Multi-scale contrast enhancement with application to image fusion," *Opt. Eng.*, vol. 31, pp. 1026–1031, May 1992.

[8] J. N. Rinker, "Hyperspectral imagery: What is it?—What can it do?," U.S. Army Eng. Topographic Labs., DTIC AD-A231 164, pp. 1–25, 1990.

[9] W. Porter and H. T. Enmark, "A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS)," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 834, pp. 24–31, 1987.

[10] G. Vane, M. Chrisp, H. Enmark, S. Macenka, and J. Solomon, "Airborne visible/infrared imaging spectrometer (AVIRIS): An advanced tool for earth remote sensing," in *Proc. 1984 IEEE Int. Geosci. Remote Sensing Symp.*, vol. 2, pp. 751–757.

[11] G. Vane, R. Green, T. Chrien, H. Enmark, E. Hansen, and W. Porter, "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 44, pp. 127–143, May 1993.

[12] M. W. Cannon, "A study of stimulus range effects in free modulus magnitude estimation of contrast," *Vision Res.*, vol. 24, no. 9, pp. 1049–1055, 1984.

[13] E. Peli, "Contrast in complex images," *Opt. Soc. Amer.*, vol. 7, pp. 2032–2040, Oct. 1990.

[14] E. Peli, R. B. Goldstein, G. M. Young, and L. E. Arend, "Contrast sensitivity functions for analysis and simulation of visual perception," *Opt. Soc. Amer.*, vol. 3, pp. 126–129, Feb. 1990.

[15] G. S. Rubin and R. A. Schuchard, "Does contrast sensitivity predict face recognition performance in low-vision observers," *Opt. Soc. Amer.*, vol. 3, pp. 130–133, Feb. 1990.

[16] M. S. Sanders and E. J. McCormick, *Human Factors in Engineering and Design*, 7th ed. New York: McGraw-Hill, 1993, pp. 91–102.

[17] P. Stager and D. Hameluck, "Contrast sensitivity and visual detection in search and rescue," *Defence Civil Inst. Environ. Medicine*, Toronto, Ont., Canada, June 1986.

[18] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 525–536, July 1974.

[19] *Multiresolution Image Processing and Analysis*, A. Rosenfeld, Ed. New York: Springer-Verlag, 1984.

[20] T. Wilson, S. Rogers, and L. Myers, "Perceptual-based hyperspectral image fusion using multiresolution analysis," *Opt. Eng.*, vol. 34, pp. 3154–3164, Nov. 1995.

[21] T. Wilson, "Perceptual based image fusion with applications to hyperspectral image data," M.S. thesis, Air Force Inst. Technol., Wright Patterson AFB, Dayton, OH, 1994.

[22] K. H. Fielding, "Spatio-temporal pattern recognition using hidden Markov models," Ph.D. dissertation, Air Force Inst. Technol., Wright Patterson AFB, Dayton, OH, pp. 57–59, 1994.

Terry A. Wilson was born in Chicago Heights, IL, on January 5, 1962. He received the B.S. degree in electrical engineering from the University of Florida, Gainesville, in 1990 and the M.S. degree in electrical engineering from the Air Force Institute of Technology, Wright Patterson AFB, Dayton, OH, in 1994.

He is a Captain in the United States Air Force at Wright Patterson AFB.

Steven K. Rogers (S'76–M'78–SM'95) is a Professor in the Department of Electrical and Computer Engineering at the Air Force Institute of Technology, Wright-Patterson Air Force Base, Dayton, OH. He is presently conducting an extensive research program in neural networks. The research program addresses the problems inherent in making smart computers. He has published more than 200 papers in the areas of neural networks, pattern recognition and optical information processing and the textbook *Introduction to Biological and Artificial Neural Networks for Pattern Recognition*.

Matthew Kabrisky (A'53–M'58–SM'86–LS'95) received the Ph.D. degree in electrical engineering from the University of Illinois, Normal, in 1964, and the B.S.E.E. and M.S.E.E. degrees from the Polytechnic Institute of New York, Brooklyn, in 1951 and 1952, respectively.

He has been a Member of the Faculty of the Air Force Institute of Technology, Dayton, OH, since February 1961. He is currently a Professor Emeritus. He and his students have published many reports and papers on the subjects of target and speech recognition, robot guidance, motion sickness, and instrumentation and modeling of the human central nervous system. He has also published a book on mathematical modeling of the human visual system.