AN ANALYSIS OF PYRMIDAL IMAGE FUSION TECHNIQUES

T. R. Meek
Utah State University
Logan, Utah

Abstract

This paper discusses the application of multiresolution image fusion techniques to synthetic aperture radar (SAR) and Landsat imagery. Results were acquired through the development and application of image fusion software to test images. The test images were fused using six image fusion techniques that are the combinations from three types of image decomposition algorithms (ratio of low pass [RoLP] pyramids, gradient pyramids, and morphological pyramids) and two types of fusion algorithms (selection and hybrid selection and averaging). Based upon test results, this study concludes that: small details in city areas make morphological pyramids ineffective, selection forms of fusion do not effectively combine the data, RoLP and gradient pyramids with hybrid fusion produce the best results, and optimum pyramid depth is dependent upon the size of detail in the images.

Introduction

This paper will give the reader a brief introduction to findings of my research at Utah State University. For a more comprehensive understanding the reader is referred to my thesis, Multiresolution Image Fusion of Thematic Mapper Imagery with Synthetic Aperture Imagery, at the Department of Electrical and Computer Engineering, Utah State University, Logan, UT[1].

The study used three image decomposition algorithms and two image fusion algorithms to form six techniques of multiresolution image fusion. The study was formed to demonstrate which technique had the best performance; this was determined by composite image quality and processor time required to produce the composite image.

The first objective of this study was to compare current image fusion techniques and diagnose their effectiveness in fusing SAR and Landsat imagery based upon the composite image results obtained through the fusion of real SAR and Landsat images and the time required to generate the composite image. This was accomplished by applying the fusion techniques to several test images, recording the time required for each technique to produce the composite image, and visually inspecting the image to observe composite image quality.

The second objective was to combine test images by applying the fusion techniques to real SAR and Landsat image data to obtain useful information from more than three remote sensor bands. This was accomplished by creating test images designed to show specific surface features and then fusing the test images to verify that the surface features have been successfully fused.

The third objective of this study was to determine the optimum pyramid depth for the fusion of remotely sensed imagery. This was determined by examining the same set of test images fused over a range of pyramid depths.

Image Structure and Multiresolution Pyramids

This study will use three to represent an image over a range of scales. The techniques used to represent the images over a range of scales are called pyramids; the three of interest are RoLP, gradient, and morphological pyramids.

A complete image description can be obtained by studying an image structure over a range of scales. When we zoom in on an image, we clearly see the substructure; however, we lose the clarity of the outlines. On the other hand, when we zoom out to look at the entire picture, the scene loses detail. It logically follows that relevant details of an image can be observed only within a certain range of spatial resolution. If we focus on the small details, we lose focus of the big picture; on the other hand, if we zoom out to see the whole picture, it is difficult to discern the small details.

A series of images progressively smaller in structural content can be created by repetitive application of a processing operator with a progressively increasing scale. This operator would eliminate details smaller than a certain size. This operator acts like a filter, just as sifting gravel through screens with different wire spacings filters gravel into different groups, dependent upon particle size. Repetitive application of this operator separates the image into scenes with different resolution of detail.

By reducing the sample frequency and increasing the filter size, a hierarchical relation is generated. Reducing the sample frequency is the same as subsampling the image. A pyramid is a sequence of images in which each image is a filtered and subsampled copy of its predecessor2 . The term "multiresolution pyramid" comes from the relationship where successive levels in a pyramid are reduced resolution copies of the input image[2-8].

The function that generates the next level of the pyramid could be called REDUCE since both the resolution and sample density are decreased. REDUCE would both filter and subsample the image. To create a pyramid starting with the source image as P_0,

$$\mathbf{P}_k = \mathrm{REDUCE}(\mathbf{P}_{k-1}) \quad \text{for } k = 1,2,...n, \tag{1}$$

where **n** is the number of levels in the pyramid.

Pyramid reconstruction to recover an image from its pyramid will need an EXPAND function because each level differs in sample density. EXPAND is defined as follows:

$$\mathbf{P}_{k-1} = \mathrm{EXPAND}(\mathbf{P}_k) \quad \text{for } k = n-1, n-2, ...0, \tag{2}$$

where n is the number of levels in the pyramid. Specific details of the EXPAND and REDUCE operators are dependent upon which types of pyramids are used. Techniques used to generate pyramids can be classified into two types: (1) linear and (2) morphological. This study used two linear filters, ratio of low-pass (RoLP) and Gaussian, and one morphological filter. For more information on pyramid types and filters the reader is referred to my thesis.

Pyramid Fusion Techniques

This study used three pyramid techniques mentioned above to combine or fuse two source images into a single composite image. In order to do this, we must define a way to fuse two pyramids into a single pyramid.

Pyramids are simply a convenient way to represent an image over a range of spatial resolutions. By combining the images at each level of the pyramid, the composite image, formed by pyramid reconstruction, will have consistency over all resolutions.

When fusing two pyramids, each of the levels of the pyramids is fused into a composite level, resulting in a composite pyramid. Refer to Figure 1. Once the composite pyramid is formed, the fused image of the source images is generated, employing the pyramid reconstruction techniques associated with the technique used to generate the source pyramids. For example, if pyramids **A** and **B** were generated from two source images, the composite image resulting from the fusion of pyramids **A** and **B** would be reconstructed from the composite pyramid **C**. Each level of the composite pyramid is defined as

$$\mathbf{C}_k = \mathrm{FUSE}(\mathbf{A}_k, \mathbf{B}_k) \quad \text{for } k = n, n-1, n-2, ...0, \tag{42}$$

and n is the number of levels in the pyramid. FUSE is a function that converts the two images into the composite, using a fusion algorithm.

The FUSE function was implemented two different ways; one way used a selection approach, while the other used a selection and averaging approach. The selection approach selected the pixel of highest contrast and it went into the composite image. The selection and averaging approach, called hybrid averaging and selection in this paper, selected a pixel when correlation was low, however, when correlation was high the pixels from the two source images were averaged for the composite pixel value.

Test Approach

The three pyramid types for image decomposition mentioned earlier are employed in this study. The pyramids used in the fusion functions are the RoLP, gradient, and morphological pyramids. These three pyramid techniques can be combined with the two fusion techniques in six possible ways. The three combinations described in the current literature are: (1) using a RoLP pyramid for image decomposition with contrast fusion for image merging[2,7]; (2) using a gradient pyramid for image decomposition with hybrid averaging and selection fusion for image merging[3]; and (3) using a morphological pyramid for image decomposition with contrast fusion for image merging[4]. Three not discussed in current literature are: (1) using a RoLP pyramid for image decomposition with hybrid averaging and selection fusion for image merging; (2) using a gradient pyramid for image decomposition with contrast fusion for image merging; and (3) using a morphological pyramid for image decomposition with hybrid averaging and selection fusion for image merging.

In previous studies, composite Landsat and SAR images have been created by assigning specific bands, or ratios of bands, to specific pixel colors. Images that have pixels representing both SAR and Landsat data will be called "hybrid SAR and Landsat images" or simply "hybrid images." Hybrid images can be very useful, as presented in a thesis by David Oliver[9]; however, they limit the number of bands that can be viewed to three. By applying multiresolution fusion functions to source images, it is the purpose of this study to make it possible to effectively view images that contain information from more than three spectral bands. The remainder of this paper presents the results and conclusions of applying the six fusion functions to test images.

Fusing SAR and Landsat Images to Evaluate Fusion Functions

For this test imagery which contained mountainous and agricultural areas as well as urban areas was used. The results demonstrated that the gradient and RoLP pyramids with the hybrid fusion technique provide the composite images with the most well integrated fusion of source image features.

The visual results of the fusion techniques using morphological pyramids showed that morphological pyramids do not work well for remotely sensed imagery (of this spatial resolution). The low quality of the composite images that used morphological filters is due to the small resolution of the visual elements; for example, roads are only one or two pixels wide. In this case, using a 2x2 or a 3x3 structuring element for the morphological filters does not allow reconstruction of such small detail. The image detail cannot be reconstructed with a resolution higher than that of the structuring element. Hence, the composite image does not contain the important, small details from the source images because the details are smaller than the structuring element. A smaller structuring element is not feasible because using a structuring element on the order of one pixel does not filter the image at all. Morphological filters work well for high-resolution images where image substructures are large in comparison with the structuring element. However, with these images the entire city is blurred.

Results obtained using the selection technique for image fusion are unacceptable in this study because they select details from only one image or another. In the composite images, cities and mountains end up being represented by the SAR information, and agricultural areas are represented by the Landsat images. This is because Landsat images have the highest return values from agricultural areas, and SAR images have the highest return values in cities and mountains. Since the areas of saturated return correspond to the highest pixel values and the contrast selection technique is based upon ratios of pixel intensity, selection fusion leads to composite images that have the SAR information in city areas and Landsat information in the agricultural areas.

Fusing Hybrid SAR/Landsat Images

In the previous section, SAR and Landsat images were fused to obtain information about how the fusion functions work for remotely sensed imagery. Fusion applications, however, are more likely to use hybrid images. Hybrid images have the advantage of using multiple sensors to obtain single source images. The benefit of having the different bands of TM sensors and SAR sensors is that each band shows a particular feature of the surface. When these features are understood, source images can be formed to show specific information about a given surface area. In this section, the fusion techniques are applied to a sets of source images used to show the applicability of fusion of hybrid SAR and Landsat images. A list of the figures used in this section is given in Table IV.

For the first set of composite images, a source image designed to show urban, suburban, and agricultural areas was combined with another source image designed to show health of vegetation. For the first image, we want to use a band that reflects vegetation and a band that reflects anthropological structures; this will demonstrate the distinct difference between urban and agricultural areas. Landsat band 2 is in the visible spectrum and returns a peak value for vegetation. Landsat band 4 is in the near infrared spectrum and shows healthy vegetation and land/water interfaces. Both SAR bands C and L reflect well from artificial structures and would work well for this image; however, band L has a higher return from artificial structures than band C, and band C reflects from vegetation. For the first image, we use Landsat bands 2 and 4 and SAR band L. The second source image is designed to show health of vegetation, so both Landsat bands 2 and 4 are again used. The third band of the second source image is one that has a low return in vegetated areas; Landsat band 3 is in the visible-light spectrum that corresponds to chlorophyll absorption.

The first image emphasizes land use categories and land/water boundaries, and the second shows health of vegetation and land/water boundaries. After fusing the source images using RoLP and gradient pyramids with hybrid fusion it was concluded that the fusion of the source images was successful because the information represented by Landsat bands 2 and 4, which show health of vegetation and water/land boundaries, remains virtually unchanged, whereas the composite images make viewing the vegetation in the city easier without changing the ease of viewing the land use information. This is because the pixels represented by the color red are now the data represented by the fusion between the SAR L band and the Landsat band 3. This does not increase the amount of green present in the image; it only makes it easier for a human analyst to observe because of the way we perceive contrast. The reason for this is described by Weber's Law[10]. By decreasing the amount of red in the local area, it decreases the amount of contrast between the green and red pixels; hence, the vegetation (green) is more easily observed by a human analyst because the red is reduced. If we were to simply reduce the intensity of the red pixels, the contrast between green and red in the city would be still be easily observed; however, in areas where the red pixels are the only source, the contrast would also be reduced. The use of fusion allows varying the amount of change in contrast for a given area based on the correlation between the two source images.

Composite Image Appearance and Quality

The composite image appearance depends upon several factors. In this study, Landsat and SAR images were converted into bitmaps for fusion. This preserved the pixel data. In order for the REDUCE and EXPAND functions to work properly, the input image size needs to be equal to a power of two, plus one. For example, the two sizes used in this study were 257x257 and 513x513. This works because $257 = 2^8 + 1$ and $513 = 2^9 + 1$. If a source image is passed to the function that generates a pyramid, the image needs to have x and y sizes that are a power of two, plus one. If the dimensions of the source image do not meet this criterion, the image needs to be resized. When resizing occurs, pixel values must be interpolated; and the data used are no longer exact. Therefore, the data in the composite image are not exact.

Another consideration in the composite image appearance is the storage format used. If an image is stored as a JPEG file, the true pixel values are not saved; quantization is necessary for the compression of the file. Once again, because the data used in the source images are not exact, the data contained in the composite image are not exact. It is necessary for the user to decide what accuracy of pixel values is necessary for image analysis and take the necessary precautions when fusing the images.

When using multiresolution image fusion, the number of levels in the pyramids contributes to the quality of the composite image. For example, when the source images are decomposed into pyramids six levels deep, large image features will fuse better than if pyramids only two levels deep were used. The depth of a pyramid is an important parameter. If the depth is too deep, processing time is wasted. On the other hand, if the pyramid is not deep enough, the larger image subfeatures will not blend well. If it is known that the remote sensing imagery for a desired study has surface/subsurface features of a very small scale, the pyramid depth does not need to be as deep; a depth of two would work fine. In relation to the overall size of the source images used here, the features in the city are very small, while the features in the mountains are quite a bit larger. It is important to note that the pyramid depth must accommodate the fusion of the largest subfeature in the source images; this is what will determine the necessary pyramid depth.

Comparing the images with each other, it can be seen that, after a pyramid depth of two, the added pyramid levels do not visually add much detail to the composite image. The most noticeable change, upon visual inspection, is in the color of the mountains. By subtracting the images it is noticed that the difference from level to level is, indeed, in the mountainous regions of the image. The difference in the mountains was expected because the image features in the mountains are much larger than the image features in the city. By inspecting the resultant images, it can be observed that the detail in the composite images for the city areas did not change noticeably for any of the pyramid depths used after a depth of two. For any set of images, the optimum pyramid depth depends upon the size of the details considered for analysis. The larger the details in the source images, the deeper the pyramid depth needs to be for satisfactory fusion. If the pyramids were skipped all together (a pyramid of depth = 1), the composite image has a higher level of detail missing as opposed to using a depth of two.

Findings

The findings from this study are verified by the fused images generated from SAR and TM data of the same terrestrial scenes. It has been shown that there is a more effective way to view composite SAR and Landsat images than by simply viewing three bands at one time. The cost of the more effective composite image is computation time. The following Table is a summary of processor time for the various fusion techniques that were studied.

The fastest fusion techniques use morphological filters; however, as the composite images indicate, morphological filters do not work well

COMPUTATION TIMES FOR FUSION TECHNIQUES

Technique	Time (Seconds)
Morphological-Selection1	16.2
RoLP-Selection	130.1
Morphological-Hybrid	260.2
RoLP-Hybrid	631.2
Gradient-Selection	1142.6
Gradient-Hybrid	1823.4

Computational time on a Pentium 266 MHz processor with 64 MB of RAM for 513x513 pixel images. For an Image size of 257x257 all computational times should be quartered.

for these types of remotely sensed data because of the small-scale of image subfeatures in comparison with the structuring element.

COMPUTATION TIMES FOR FUSION TECHNIQUES

Technique	Time (Seconds)
Morphological-Selection1	16.2
RoLP-Selection	130.1
Morphological-Hybrid	260.2
RoLP-Hybrid	631.2
Gradient-Selection	1142.6
Gradient-Hybrid	1823.4

Computational time on a Pentium 266 MHz processor with 64 MB of RAM for 513x513 pixel images. For an Image size of 257x257 all computational times should be quartered.

Fusion techniques using RoLP pyramids are the second fastest. The composite images formed from RoLP techniques are beneficial in that they present the data from both source images in the composite scene better than do the simple hybrid images used in previous studies. Compared with techniques that use gradient pyramids, however, the composite images from RoLP pyramids tend to appear blurry. This appearance of being blurry, however, is an artifact of the reduced contrast in the composite image. The quality remains approximately constant between the gradient and RoLP composite images (given equal pyramid depths).

Gradient pyramids yielded composite images in approximately equal quality to, or slightly better than, the RoLP pyramids (gradient composites do not appear as blurry); however, the computation time for gradient pyramids is about four times that of the RoLP pyramids. The gradient pyramid is actually a combination of four pyramids. When exact analysis of imagery is necessary, the gradient pyramid may be the best choice; however, for most applications, the RoLP pyramid with hybrid fusion will work just as well.

It was found in this study that the selection approach to image fusion is undesirable. Because the selection technique chooses only one band or another to represent, it does not fuse them. Although the selection technique is almost twice as fast as the hybrid selection and averaging technique, the hybrid approach yields a composite image with better detail. A result of this study, therefore, is the suggestion that, for general fusion applications, a RoLP pyramid with hybrid selection and averaging fusion should be used. If small details are a concern and time is not a constraint, the gradient pyramid with hybrid fusion may yield a slightly better result.

This study has demonstrated that the optimum pyramid depth depends upon the largest important subfeature in the image. If the features of concern are small in detail, like the cities in the case of satellite data, a pyramid depth of two will produce the same result as a pyramid depth of six. On the other hand, if the features in the image are large, like mountains, a much deeper pyramid is necessary.

A simple rule to follow would be to use a pyramid depth of two if the image features desired for fusion are on the order of one to two pixels. For most applications, a pyramid depth of three would work fine. If the image features are very large, on the order of hundreds of pixels, a pyramid depth of six would be appropriate. Rarely would a pyramid depth of greater than five or six be needed. Based upon the results from figures 16 and 17, a pyramid depth of at least two should be used in all fusion applications because the amount of detail added when changing from a pyramid of depth one to a pyramid of depth two is sufficiently large.

Literature Cited

1 T. Meek, "*Multiresolution Image Fusion of Thematic Mapper Imagery with Synthetic Aperture Imagery.*" M.S. thesis, Utah State University, Logan, UT. 1999.

2 A. Toet, "Hierarchical image fusion," *Machine Vision and Applications*, vol. 3, no. 1, pp. 1-11, 1990.
3 P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," *1993 IEEE 4th International Conf. on Computer Vision*, vol. 4, pp. 173-182, Feb. 1993.
4 S. Marshall, G. K. Matsopoulos, and J.N.H Brunt, "Multiresolution morphological fusion of MR and CT images of the human brain," in *Mathematical Morphology and Its Applications to Image Processing*, J. Serra & P. Soille, Eds., Dordrecht, Netherlands: Kluwer Academic Publishers, 1994, pp. 317-324.
5 P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. on Comm.*, vol. 31, no. 4, pp. 532-540, 1983.
6 A. Toet., "A morphological pyramid image decomposition," *Pattern Recognition Letters*, vol. 9, no. 4, pp. 255-261, 1989.
7 A. Toet, L. J. van Ruyven, and J. M. Valeton, "Merging thermal and visual images by a contrast pyramid," *Optical Engineering*, vol. 28, no. 7, pp. 789-792, 1989.
8 A. Toet, "Multiscale contrast enhancement with applications to image fusion," *Optical Engineering*, vol. 31, no. 5, pp. 1026-1031, 1992.
9 D. Oliver, "*Image processing for integration of Landsat Thematic Mapper imagery with SIR-C synthetic aperture radar imagery.*" M.S. thesis, Utah State University, Logan, UT. 1996.
10 A. K. Jain, *Fundementals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.