Performance of evaluation methods in image fusion

Sascha Klonus  
Institute for Geoinformatics and Remote Sensing  
University of Osnabrunck  
Germany  
sklonus@igf.uni-osnabrueck.de

Manfred Ehlers  
Institute for Geoinformatics and Remote Sensing  
University of Osnabrunck  
Germany  
mehlers@igf.uni-osnabrueck.de

Abstract - Many algorithms and software tools have been developed for fusing panchromatic and multispectral datasets in remote sensing. Also, a number of methods has been proposed and developed for the comparative evaluation of fusion results. To this date, however, no papers have been published that analyze effectiveness and quality of the evaluation techniques. In our study, methods that evaluate fusion quality are tested for different images and test sites. This analysis shows that in most cases the tested methods perform well, but are sometimes inconsistent with visual analysis results.

Keywords: Image fusion, SSIM, RMSE, Correlation Coefficients.

1 Introduction

Most of the earth observation satellites such as Spot, Ikonos, Quickbird, Formosat or Orbview and also some digital airborne sensors like DMC or UltraCam record image data in two different modes, a low-resolution multispectral and high-resolution panchromatic mode. A common feature for these sensors is the fact that the highest spatial resolution is recorded in their panchromatic mode whereas the multispectral recording mode produces images of reduced spatial resolution. The difference in spatial resolution between the panchromatic and the multispectral mode can be measured by the ratio of their respective ground sampling distances (GSD) and may vary between 1:2 and 1:5. This ratio can get worse if data from different satellites are used. For example, the resolution ratio between Ikonos (pan mode) and SPOT 5 (multispectral mode) is 1:10. The objective of iconic image fusion is to combine the panchromatic and the multispectral information to form a fused multispectral image that retains the spatial information from the high resolution panchromatic image and the spectral characteristics of the lower resolution multispectral image. Applications for integrated image datasets include environmental/agriculture assessment, urban mapping, and change detection.

Image fusion methods have mostly been developed for single-sensor, single-date fusion [1], [2], for example, Ikonos or Quickbird panchromatic images are fused with the equivalent Ikonos or Quickbird multispectral image.

Multisensoral or multitemporal fusion is seldom in use, or is only used with Landsat multispectral and Spot panchromatic data [3], [4]. Therefore most of the fusion methods show dependencies, if different sensors or data from different times are combined [5]. However with the advent of new sensors, which are either only panchromatic (Worldview 1) or multispectral (RapidEye), it becomes increasingly important to fuse multitemporal data from different sensors.

Therefore different fusion techniques are used in this study to test the performance of evaluation methods on these types of fusion.

2 Material and Methods

2.1 Datasets

For the single-sensor/single-date fusion, a panchromatic Formosat image, recorded on 25 August 2005 is fused with its equivalent multispectral image. The study area is located in France in the region Maussanne.

The next three images, the study area is located in Spain, around the village Santo Domingo de la Calzada. For the multisensor case, a panchromatic Ikonos image was used, recorded on 3 May 2005. This image was fused with a Spot 5 multispectral image from 10 April 2005.

For the multitemporal case, the same Ikonos image was used and fused with a Spot 4 scene recorded on 20 July 2005.

Additionally, as an alternative for the panchromatic input we used image data from the German RADAR satellite TerraSAR-X. In the spotlight mode, the satellite is able to record data with one meter ground resolution. The radar image was fused with the Spot 5 image of 10 April 2005. Radar/electro-optical fusion is a very demanding task, because two completely different sensors are used, an active and a passive sensor, with a completely different recording geometry. Fusion with radar, however, has the advantage that current information from the all-weather sensor can be combined with multispectral archive data. For example, after a catastrophe such as an earthquake or a tsunami, up-to-date radar can be fused with multispectral
data from a different time thus deliver enhanced products for visual inspection and overview.

All multispectral images are resampled before the fusion to the spatial resolution of the high resolution image using cubic convolution.

2.2 Fusion methods

Eleven different fusion methods are used in this investigation:

A simple Wavelet Transform is used, which is implemented in the Erdas Imagine Software package. For image fusion, a wavelet transform is applied to the panchromatic image resulting in a four-component image: a low-resolution approximation component (LL) and three images of horizontal (HL), vertical (LH), and diagonal (HH) wavelet coefficients which contain information of local spatial detail. The low-resolution component is then replaced by a selected band of the multispectral image [6]. This process is repeated for each band until all bands are transformed. A reverse wavelet transform is applied to the fused components to create the fused multispectral image [7]. Generally, wavelet fused images produce good spectral preservation but poor spatial improvement

The AWL method [8] is one of the existing multiresolution wavelet-based image fusion techniques. It was originally designed for a three-band red-green-blue (RGB) multispectral image. In this method, the spectral signature is preserved because the high resolution panchromatic structure is integrated into the luminance L-band of the original low resolution multispectral image. Therefore, this method is only defined for three bands. It was extended to n bands by [9]. It maintains the spectral signature of an n-band image in the same way as AWL does with RGB images. This generalized method is called proportional AWL (AWLP). This method produces better results than standard wavelet algorithms, but the spatial improvement is in most cases still not acceptable.

The multiplicative method is derived from the four component technique of [10]. In this paper Crippen argued that of the four possible arithmetic methods only the multiplication is unlikely to distort the colors by transforming an intensity image into a panchromatic image. Therefore this algorithm is a simple multiplication of each multispectral band with the panchromatic image. The advantage of the algorithm is that it is straightforward and simple. By multiplying the same information into all bands, however, it creates spectral bands of a higher correlation which means that it does alter the spectral characteristics of the original image data.

The Brovey transformation was developed to avoid the disadvantages of the multiplicative method. It is a combination of arithmetic operations and normalizes the spectral bands before they are multiplied with the panchromatic image [11]. The spectral properties, however, are usually not well preserved

The color normalization (CN) spectral sharpening is an extension of the Brovey algorithm and groups the input image bands into spectral segments defined by the spectral range of the panchromatic image. The corresponding band segments are processed together in the following manner: Each input band is multiplied by the sharpening band and then normalized by dividing it by the sum of the input bands in the segment [12]. This method works well for data from one sensor, but if the spectral range of the panchromatic image does not match the spectral range of the multispectral images no spatial improvement is visible.

To fuse the images with the intensity-hue-saturation (IHS) fusion, three bands of a multispectral image are transformed from the RGB domain into the IHS color space. The panchromatic component is matched to the intensity of the IHS image and replaces the intensity component. We make use of the modified IHS fusion which was developed for a better fit of the fused multispectral bands to the original data [13]. After the matching, the panchromatic image replaces the intensity in the original IHS image and the fused image is transformed back into the RGB color space. This method works also well with data from one sensor, but for multitemporal or multisensoral fusion the results are in most cases not acceptable.

The Ehlers fusion is based on an IHS transform coupled with a Fourier domain filtering. This technique is extended to include more than 3 bands by using multiple IHS transforms until the number of bands is exhausted. A subsequent Fourier transform of the intensity component and the panchromatic image allows an adaptive filter design in the frequency domain. Using fast Fourier transform (FFT) techniques, the spatial components to be enhanced or suppressed can be directly accessed. The intensity spectrum is filtered with a low pass filter (LP) whereas the panchromatic spectrum is filtered with an inverse high pass filter (HP). After filtering, the images are transformed back into the spatial domain with an inverse FFT and added together to form a fused intensity component with the low-frequency information from the low resolution multispectral image and the high-frequency information from the high resolution image. This new intensity component and the original hue and saturation components of the multispectral image form a new IHS image. As the last step, an inverse IHS transformation produces a fused RGB image. These steps can be repeated with successive 3-band selections until all bands are fused with the panchromatic image (for a complete description of the method see [14]). The Ehlers fusion shows the best
spectral preservation but also the highest computation time.

The principal component (PC) transform is a statistical technique that transforms a multivariate dataset of correlated variables into a dataset of uncorrelated linear combinations of the original variables. For images, it creates an uncorrelated feature space that can be used for further analysis instead of the original multispectral feature space. The PC is applied to the multispectral bands. The panchromatic image is histogram matched to the first principal component (sometimes to the second). It then replaces the selected component and an inverse PC transform takes the fused dataset back into the original multispectral feature space [4]. The advantage of the PC fusion is that the number of bands is not restricted (such as for the original IHS or Brovey fusions). It is, however, a statistical procedure which means that it is sensitive to the area to be sharpened. The fusion results may vary depending on the selected image subsets [17].

The Gram Schmidt fusion simulates a panchromatic band from the lower spatial resolution spectral bands. In general, this is achieved by averaging the multispectral bands. As the next step, a Gram Schmidt transformation is performed for the simulated panchromatic band and the multispectral bands with the simulated panchromatic band employed as the first band. Then the high spatial resolution panchromatic band replaces the first Gram Schmidt band. Finally, an inverse Gram Schmidt transform is applied to create the pansharpened multispectral bands [15]. This method usually produces good results for fusion images from one sensor, but it is also a statistical procedure like the PC, so that the fusion results may vary depending on the selected datasets.

For the high pass filtering (HPF) fusion, first the ratio between the spatial resolution of the panchromatic and the multispectral image is calculated. A high pass convolution filter kernel is created and used to filter the high-resolution input data with the size of the kernel based on the ratio. The HPF image is added to each multispectral band. Before the summation, the HPF image is weighted relative to the global standard deviation of the multispectral bands with the weight factors again calculated from the ratio. As a final step, a linear stretch is applied to the new multispectral image to match the mean and standard deviation values of the original input multispectral image [16]. It shows acceptable results also for multisensorial and multitemporal data. Sometimes the edges are emphasized too much.

To apply the University of New Brunswick (UNB) fusion algorithm [18], a histogram standardization is calculated for the multispectral and panchromatic bands of the input images. The multispectral bands in the spectral range of the panchromatic image are selected and a regression analysis is calculated using a least square algorithm. The results are used as weights for the multispectral bands. Via multiplication with the corresponding bands and a following addition, a new synthesized image is produced. To create the fused image, each standardized multispectral image is multiplied with the standardized panchromatic image and divided by the synthesized image. This method was designed for single-sensor, single-date images and does not produce acceptable results for multisensor and/or multitemporal fusion. It is used as the standard method for Quickbird pansharpening.

### 2.3 Evaluation methods

The evaluation procedures are based on the verification of the preservation of spectral characteristics and the improvement of the spatial resolution. First, the fused images are visually compared. The visual appearance may be subjective and depends on the human interpreter, but the power of the visual cognition as a final backdrop cannot be underestimated. Second, a number of statistical evaluation methods are used to measure the color preservation. These methods have to be objective, reproducible, and of quantitative nature. We make use of the following techniques:

The correlation coefficient (CC) between the original multispectral bands and the equivalent fused bands. This value ranges from -1 to 1. The best correspondence between fused and original image data shows the highest correlation values.

For a per-pixel deviation (PD) (see [19], pp. 147-160), it is necessary to degrade the fused image to the spatial resolution of the original image. This image is then subtracted from the original image on a per-pixel basis. As final step, we calculate the average deviation per pixel measured as digital number (DN) which is based on a 8-bit or 16-bit range. Here, zero is the best value.

The RMS error as proposed by [19] is computed as the difference of the standard deviation and the mean of the fused and the original image. The best possible value is again zero.

The Structure Similarity Index (SSIM) was proposed by [20]. The SSIM is a method that combines a comparison of luminance, contrast and structure and is applied locally in an 8 x 8 square window. This window is moved pixel-by-pixel over the entire image. At each step, the local statistics and the SSIM index are calculated within the window. The values vary between 0 and 1. Values close to 1 show the highest correspondence with the original images.
The objective is to find the fused image with the optimal combination of spectral characteristics preservation and spatial improvement. Consequently as a third step two different quantitative methods are chosen to quantitatively measure the quality of the spatial improvement.

High pass correlation (HCC): Correlation between the original panchromatic band and the fused bands after high pass filtering. This algorithm was proposed by [21]. The high pass filter is applied to the panchromatic image and each band of the fused image. Then the correlation coefficients between the high pass filtered bands and the high pass filtered panchromatic image are calculated.

Edge detection (ED) in the panchromatic image and the fused multispectral bands: For this, we selected a Sobel filter [17] and performed a visual analysis of the correspondence of edges detected in the panchromatic and the fused multispectral images. This was done independently for each band. The value is given in percent and varies between 0 and 100. 100% means that all the edges in the panchromatic image were detected in the fused image.

3 Results

3.1 Fusion

Eleven fusion techniques were applied to the four datasets of input images. In total, 44 fused images were produced and analyzed with the six different evaluation procedures in this study. The results for the different fusion methods are presented in table 1. The figures present the average of all bands from all 11 fusion methods used for the 4 datasets. SD stands for single-sensor/single-date, MS for multisensor, MT for multitemporal and R for radar fusion).

Tab.1: Evaluation Results for single-sensor/single-date (SD), multisensor (MS), multitemporal (MT) and radar fusion (R).

<table>
<thead>
<tr>
<th></th>
<th>SD</th>
<th>MS</th>
<th>MT</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM</td>
<td>0.834</td>
<td>0.746</td>
<td>0.551</td>
<td>0.485</td>
</tr>
<tr>
<td>CC</td>
<td>0.906</td>
<td>0.855</td>
<td>0.686</td>
<td>0.447</td>
</tr>
<tr>
<td>PD</td>
<td>12.003</td>
<td>13.458</td>
<td>29.051</td>
<td>33.928</td>
</tr>
<tr>
<td>RMSE</td>
<td>10.304</td>
<td>15.030</td>
<td>28.161</td>
<td>37.125</td>
</tr>
<tr>
<td>HCC</td>
<td>0.892</td>
<td>0.717</td>
<td>0.677</td>
<td>0.384</td>
</tr>
<tr>
<td>ED</td>
<td>94.981</td>
<td>92.887</td>
<td>93.144</td>
<td>91.307</td>
</tr>
</tbody>
</table>

The spectral evaluation methods, SSIM, CC, PD and RMSE show the problems with multisensor and multitemporal fusion. The best results are achieved for the single-sensor case, for which most of the methods are created. If only two sensors are used with a recording date close to each other, than the fusion results are still acceptable for most of the methods. But if the difference in the recording date of the two sensors is more than two months, most of the methods produce results with severe color distortions. This gets worse, if a radar sensor is used as an input for the fusion. Only the Ehlers and the simple Wavelet fusion produced results which showed good color preservation: Correlation coefficients and SSIM were around 0.9 and PD and RMSE below 3.

The edge detection results verify that most methods could inject the high resolution into the multispectral image for all tested fusion methods. Only the high pass correlation coefficients are to some degree influenced by the different fusion types. The reason for this should be investigated, because visual inspection did not show any obvious structural differences. Only the low correlation for the radar fusion could be explained by the special characteristics of a radar image. Here, only the Brovey fusion produces good results with a correlation coefficient above 0.9.

3.2 Evaluation

As stated before, in most of the cases the evaluation methods produce results that are consistent. Problems, however, occur especially for SSIM, PD and RMSE, if the values of the fused image are not in the same range as the original images. This happens often especially for the multiplicative and the Brovey fusions. The images show a good visual appearance but poor statistics. Only the correlation coefficients coincide with the visual appearance. This is in accordance with the findings reported in [22].

Another example for a failure of the RMSE evaluation method is given in figures 1 to 3. Displayed are the multispectral Spot 5 image (fig. 1), the same image fused.
with the Gram Schmidt (fig. 2) and the modified IHS (fig. 3) methods. It is quite obvious that the modified IHS image preserves the colors significantly better than the Gram Schmidt method. With the Gram Schmidt fusion, fields in the lower left and the upper right edge show a red color in the color infrared band combination thus changing bare ground to a vegetated field. The colors in the modified IHS seem to be better fitting. But the RMSE for the Gram Schmidt fused image is 1.8869 compared to 2.3603 for the modified IHS fusion suggesting a better color preservation for the Gram Schmidt method.

All other evaluation methods such as SSIM, CC, and PD show better values for the modified IHS image, which would suggest that the RMSE is an inconsistent procedure to measure the quality of spectral characteristics preservation.

However, also the correlation method produces sometimes values that do not seem to be consistent. As an example, the next three images show the results of the single-sensor/single-date fusion using the multiplicative and Gram Schmidt fusion methods (fig. 4 – fig. 6). The image fused with multiplicative seems darker and a little bit blurred (fig. 6) when compared to the original multispectral image (fig. 4). The Gram Schmidt image, on the other hand, shows good color preservation. But the correlation coefficients are lower for the Gram Schmidt fusion (table 2). The other evaluation methods, however, such as SSIM, PD and RMES show better values for the Gram Schmidt fusion.
Figure 6. Multiplicative fused Formosat image displayed in the band combination: red, green, blue.

Tab. 2: Correlation coefficients for the Formosat scene

<table>
<thead>
<tr>
<th></th>
<th>Multiplicative</th>
<th>Gram Schmidt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 1</td>
<td>0.9496</td>
<td>0.8751</td>
</tr>
<tr>
<td>Band 2</td>
<td>0.9473</td>
<td>0.8698</td>
</tr>
<tr>
<td>Band 3</td>
<td>0.9291</td>
<td>0.8795</td>
</tr>
</tbody>
</table>

In most analyses, emphasis has been placed on the spectral evaluation. It is, however, also mandatory to investigate the performance of the pansharpening algorithms as far as the spatial improvement is concerned. Otherwise, the original image with no spatial improvement would produce the best results. An example for this is presented in figures 7 to 9. These images show some results for the multitemporal fusion.

Figure 7. Multispectral Spot 4 image, recorded on 20 July 2005 displayed in the band combination NIR, red, green.

Figure 8. Multispectral Spot 4 image fused with the Ikonos panchromatic image using the Ehlers method, in the band combination NIR, red, green.

Both images preserve the colors very well. This is also confirmed by the statistical evaluation (tab. 3). But if the spatial improvement is considered, only the Ehlers fusion could inject the high spatial resolution structures into the multispectral image. This is in accordance with the HCC values in table 3. The value for the Ehlers fusion is almost 50 per cent higher.

Figure 9. Multispectral Spot 4 image fused with the Ikonos panchromatic image using the simple wavelet method displayed in the band combination NIR, red, green.

Tab. 3: Quantitative evaluation results for multitemporal fusion

<table>
<thead>
<tr>
<th></th>
<th>SSIM</th>
<th>CC</th>
<th>PD</th>
<th>RMSE</th>
<th>HCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ehlers</td>
<td>0.996</td>
<td>0.999</td>
<td>0.717</td>
<td>0.309</td>
<td>0.925</td>
</tr>
<tr>
<td>Wavelet</td>
<td>0.986</td>
<td>0.996</td>
<td>1.429</td>
<td>0.101</td>
<td>0.637</td>
</tr>
</tbody>
</table>

4 Conclusions

This study proves not only the importance of evaluation methods that should be consistent and the necessity of a
combined method for a quantitative assessment of spatial improvement and spectral preservation. The idea of image fusion is to pansharpen multispectral information, which is not the case if the spatial structures in the fused images are only slightly improved when compared to the original. Then the fused image looks very similar to the original one and produces excellent results in the statistical evaluations for color preservation. As a matter of fact, the best results would be produced if no pansharpening is performed. It is also evident, that quantitative evaluation methods sometimes produce results that cannot be sustained by visual inspection. The reason for this needs to be further investigated. At this time, we can only conclude that a visual analysis has to be a significant part of all quality evaluation procedures and different evaluation methods should be used to avoid that outliers could falsify the final results.

**Acknowledgments**

The TerraSAR-X project is supported by the German BMWI through the DLR with the contract number 50EE0704.

**References**


[23]