Evaluation of Image Fusion Techniques for Large-Scale Mapping of Non-Green Vegetation

M. Bachmann and M. Habermeyer
DLR, Deutsches Fernerkundungsdatenzentrum, D-82230 Wessling
e-mail: martin.bachmann@dlr.de

ABSTRACT

In this paper, three different image fusion techniques are discussed and evaluated for the specific application of large-scale mapping of non-green vegetation. Additional synthetic scenes are used in the evaluation to exclude specific sources of errors and to cover a wide range of scene conditions.

Keywords: Image fusion, hyperspectral data, synthetic scenes, evaluation methodology, non-green vegetation.

1 INTRODUCTION

Non-green vegetation, including non-photosynthetically active, senescent and decayed plants and plant parts, is an important component of the Mediterranean ecosystem. Thus the mapping and monitoring of this particular vegetation component is necessary for monitoring land use changes as well as land degradation, forest fire management, and modelling the global carbon cycle. The utilization of hyperspectral remote sensing is appropriate for these purposes, since the application of multispectral sensors is associated with serious drawbacks when sparse canopies are being sensed. Regional and global coverage can be achieved using experimental and proposed spaceborne hyperspectral sensors (e.g. the proposed SAND mission [1]), but spatial resolution is limited. If small-scaled variations are to be mapped, it is necessary to fuse hyperspectral data with additional high spatial resolution images.

The aim of image fusion in this application is to improve the spatial resolution, which enables the detection of small-scaled objects and variations, and furthermore implying most of the spectral characteristics of the low spatial resolution image. Therefore, the following requirements are imposed on the image fusion techniques: i) Higher spatial resolution including textural properties should be introduced in the fused image ii) DN of fused image must be related to physical properties iii) Spectral information must be preserved to a very high degree iv) High number of bands must be fused.

2 METHODS AND CONCEPTS

For the evaluation the fusion techniques, it is suitable to exclude external sources of errors. Thus image fusion is performed on a set of simulated images, i.e. on spatially and spectrally degraded images simulating scenes acquired by different sensors. All errors resulting from geocoding, different view angles as well as temporal effects are therefore excluded from the evaluation and the results do not depend on other factors but the sensor properties and the fusion technique itself.

In order to simulate a ‘real-world’ sensor system with lower spatial resolution by using high spatial resolution images, the simulation of a larger ground sampling distance (GSD) alone is not a particularly good method since the neglect of certain sensor parameters has implications in all aspects of information extraction from the resulting imagery, from calculations of small target signatures to spatial-spectral signature mixing at each pixel. SCHOWENGERDT [3], pp. 83ff, gives examples for simulating images with lower spatial resolution by using high spatial resolution images and simulated point spread functions (PSFs). For the imagery evaluated in this study, the coarser spatial resolution is simulated using a gaussian PSF symmetrical in x and y size, without any shifts or errors.
in GSD. Thereby, the original pixel size of approximately 7m*7m was increased to the largest pixel size at which each fusion technique could be successfully applied, i.e. to approximately 35m*35m for the MMT and PCT fusion and to approximately 21m*21m for the SMA fusion. The low spectral resolution image is created by simulating the spectral sensor layout (number and position of bands, band FWHM) of the LANDSAT-TM 5 bands 1-5 and 7, similar to that used in airborne DAEDALUS ATM sensors. The usage of a panchromatic sharpening image did not lead to satisfying results for this application.

The database used in this study is based on ground-measured and image-derived spectra as well as a HyMap scene, all acquired during the DAISEX campaign 1999 in Barrax, Spain [2].

2.1 Synthetic Scenes

For a successful comparison of image processing techniques the spectral and spatial properties need to be known. In addition, to cover a variety of conditions, either many different images need to be evaluated, or one must have control over the spectral characteristics and the spatial arrangement of all image components. Consequently, simulated spectra or synthetic scenes are often used in remote sensing, e.g. [4] among others. Using the linear mixture model is a simple and straightforward approach for the generation of synthetic images. According to the linear spectral mixture model (e.g. [5] among many others), the reflectance of a pixel in band \( j \) can be calculated by

\[
\rho_j(x,y) = \sum_{e=1}^{k} \rho_{EM,e} \cdot f_{EM,e,j}(x,y) + \delta_j(x,y) \quad (j = 1, \ldots, n)
\]

where
- \( \rho_j \) denotes the reflectance value of the mixed spectrum in band \( j \) for pixel \( x,y \)
- \( k \) denotes the number of end-members
- \( f_{EM,e,j} \) denotes the fraction for end-member \( e \)
- \( \rho_{EM,e} \) denotes the reflectance for end-member \( e \)
- \( \delta_j \) is the residual error term
- \( n \) denotes the number of bands of the sensor

Therefore, it is necessary to create an abundance map containing the desired mixture proportion for every endmember (EM) in the desired spatial arrangement. Thereupon, the appropriate choice of EM spectra is more critical, since the spectra of vegetation components should correspond to the canopy reflectance and not to the reflectance of a single plant. An additional error, representing sensor noise, can be added. For a realistic scene, various spectra of one ground cover type can be used to simulate spectral variation. Furthermore, realistic atmospheric conditions can be added to the image by the inversion of an atmospheric correction. In this study, two different synthetic scenes based on the linear mixture model are used. To evaluate the separation of dry vegetation and bare soil under unfavourable conditions, a synthetic scene is generated using various levels of noise, since the lower signal-to-noise ratio (SNR) of spaceborne sensors tends to mask spectral absorption features as shown in Figure 1, complicating the identification of ground materials. The noise is introduced by adding realistic levels of sensor noise, which are amplified by the atmospheric correction and therefore result in a band-dependent noise distribution. The SNR test image consists of a mixture gradient between bare soil and dry barley at five different levels of sensor and atmospheric noise.

![Figure 1: Example of simulated spectra in order to test the influence of noise on fusion techniques](image)

The second synthetic scene is more complex, including both mixture and geometric test fields and still representing real-world features (see Figure 2). To represent the major land cover classes in drylands, the EM spectra used for creation are a typical bare soil in the Barrax area (SE Spain) and a dry barley spectrum, which has all characteristics of dry vegetation as well as an irrigated alfalfa spectrum, representing green vegetation. In addition to this real-world spectral content, the spatial pattern refers to natural conditions, too. The left quarters are designed to examine single-pixel separation midst a larger background. As real-world reference, these fields
simulate a decrease of vegetation, not in abundance but in a geometrical situation. Single trees surrounded by bare soil are an example. The large quarters to the right contain a horizontal gradient with green vegetation, dry vegetation and soil mixtures. The real-world equivalent would be a mostly dry vegetation canopy, which is slowly getting sparse, i.e. a slowly decreasing abundance of dry vegetation and an increasing soil abundance. In certain cases, this decrease in vegetation cover can be interpreted as different stages of degradation [6]. For the checkerboard fields, the same abundances are used for a 1pixel * 1pixel and for a 5pixel * 5pixel raster. Altogether four different variations in EM abundances are used to simulate both spectrally and spatially variation. The EM fractions are chosen in order to simulate spectrally homogeneous and heterogeneous mixtures of soil and dry vegetation. Such regular spatial distributions and mixture levels can be found in Mediterranean vineyards, fruit and olive tree cultures, where the plant spacing is often larger than the area covered by the plant.

Figure 2: Synthetic scene with spectral and spatial test patterns

From the mathematical point of view, checkerboards represent one extreme of local variation, i.e. a sudden and discrete change with high spatial frequency, while the gradients represent the other extreme, of a slow and continuous change. But still all of these patterns can be found in the Mediterranean environment, indicating the large spatial and spectral heterogeneity of these landscapes.

2.2 Evaluated Image Fusion Techniques

SMA Fusion

An image fusion technique based on spectral mixture analysis (SMA) was developed by GROSS & SCHOTT [7] and is evaluated in [4], [8]. In the following, this technique is referred to as ‘SMA-fusion’. The SMA-fusion technique sharpens low spatial resolution abundance maps by using a high spatial resolution image, which can be panchromatic or multispectral. First, the image with the lower spatial and higher spectral resolution is unmixed, using standard linear spectral unmixing techniques (see e.g. [5]). Thereafter the EMs are located to the high spatial resolution sharpening image, resulting in a set of EM fractions for every pixel in the high spatial resolution image, producing abundance maps of high spatial resolution. For a closer discussion, see [7], [8]. Using the SMA sharpening model, a critical limitation exists: a high similarity of EM spectra leads to confusion in the sharpening model, equal to the case when using similar EM spectra for linear spectral unmixing. For the sharpening model, ambiguity is increased since the EM reflectances are resampled to the spectral resolution of the high spatial resolution image. In other words, it must be ensured that the ground cover classes used for unmixing are also spectrally separable in the high spatial resolution image.

MMT Fusion

A different image fusion technique based on linear spectral unmixing, developed by ZHUKOV et al, is described in [9], [10], [11], see also [8], [12]. One important feature of the multisensor multiresolution image fusion technique (MMT) is that it does not depend on correlation between the input images, thus allowing the fusion of disparate spectral ranges, e.g. the fusion of VIS imagery with NIR imagery. The MMT is designed to preserve all the radiometric information of the high spectral resolution image while improving the spatial resolution. In this approach, the high spatial resolution image needs to be classified prior to fusion, while the high spectral resolution image is not modified. According to the nomenclature of ZHUKOV, the high spatial resolution image, which is used for classification, is called ‘classifying image’ (CI), while the high spectral resolution image is titled ‘measuring image’ (MI).

The MMT technique is based on unmixing of the MI and afterwards reconstruction of a sharpened high spectral resolution image, using the spatial information of the CI. At first, the high spatial resolution image, being either panchromatic or multispectral, needs to be classified. This can be done by unsupervised or supervised classification
techniques. In accordance with [11] & [12], the number of classes is considered a critical parameter in the MMT fusion, since too few classes reduce details in the fused image and too many classes may result in an unstable solution, therefore introducing artifacts. In this study, an unsupervised ISODATA classifier with 50 to 100 classes is applied. Thereafter, the average spectral signature taken from the MI is calculated for each class in the CI and is assigned to this class, resulting in a high spatial and high spectral resolution image. The disadvantage of this method is that the MI spectral signatures are averaged over the total area of each class. To overcome this limitation, a moving window approach is added to reduce the within-class averaging to the window size, but it also results in a slight or medium low-pass filter effect. In the following, this moving window based algorithm is called MMT-MW, while the algorithm averaging over the whole image is named MMT-WI.

PCT Fusion

The principal component transformation based fusion technique (PCT) is applied to evaluate a technique that does not fulfill the demands set up in §1, since every component substitution technique for data fusion causes a certain loss in information. Furthermore, since the PCT technique is based entirely on image statistics, the physical relationship between the image DNs and the object sensed is lost ([13] among others).

Based on the standard PC transformation, it is assumed that the first PC, containing the highest variance of all bands, is closely related to the overall intensity of the remotely sensed image and should therefore be similar for all images of this scene, i.e. showing similarity to the first PC (or the panchromatic band) of an image acquired by another sensor. This assumption is not necessarily true since the overall intensity is always coupled to the spectral information of a scene, but sometimes to a small degree, thus permitting PCT fusion. The PCT fusion technique assumes that this similarity is given, and that therefore the first PCs of high and low spatial resolution images only differ in the spatial information content. When substituting the first PC of the low spatial resolution by that of the high spatial resolution, information is increased. To avoid spectral distortions, the DN range of the new first PC must be adjusted to that of the old one by matching the histograms. Afterwards, the inverse PC-transformation is applied, resulting in an image with more spatial information and approximately the same spectral information of the original low spatial resolution image. Artifacts or contrast reversals occur in cases where the similarity assumption is not true.

2.3 Evaluation Methodology

A simple method for comparing the performance of different image fusion techniques is the assessment of the visual quality. Highly important is the usage of the same display look-up-table and histogram stretch to ensure comparability. Criteria are the sharpness of edges, the visibility of small features and tonal variations. The main disadvantage is that this method depends on the observer, therefore it can not be considered objective. Another basic method is the comparison of image statistics, namely the mean, standard deviation, minimum and maximum DN value as well as the shape and DN distribution of the histogram. This comparison is a fairly good estimator of image quality and is easy to handle (e.g. [14]). Since a ‘perfect reference’ image exists in the methodology of this study, the comparison of image statistics is meaningful because thereby the degree of similarity between reference and fused image is represented. In addition, the correlation coefficient between the fused and the reference image can be calculated as a statistical measure of fusion performance (e.g. [13]). Next, the absolute sum of errors between the reference and the fused image is a suitable measurement, too. Also the image statistics of the difference image are used in the evaluation. Especially the standard deviation is a suitable measure for the quality of image fusion [10]. In addition, [13] use the difference of the image mean (i.e. meanRef – meanFused), but this doesn't seem to be a suitable measure since the mean and the median of an image do not depend on the scale, whereas the dynamic range (i.e. the difference between minimum and maximum value) and the standard deviation will increase when the spatial resolution is improved [15]. An additional criterion is the preservation of the spectral image content, which can be evaluated by comparison of reflectance spectra before and after fusion, or between a reference and the fused image. In addition, spectral similarity measures like spectral angle mapper can be applied. Since the objective of this study is the evaluation of image fusion techniques for the mapping of non-green vegetation, a comparison of the unmixed fused images is furthermore necessary, using the statistical measures presented above. All of these measures are in accordance with the general concept for the evaluation of image fusion results proposed by [16].
3 RESULTS AND CONCLUSION

For the given scenes, the SMA technique is not successful when sharpening images degraded by more than a factor of 3*3, while the MMT and PCT fusion can be successfully applied when spatial resolution is degraded by a 5*5 PSF. Next, in addition to the different input images, SMA fusion results in sharpened abundance maps and thus can not be directly compared with the techniques described below and is therefore evaluated separately. For a comprehensive discussion of results, please refer to [8].

From the evaluation of the synthetic scenes, the following points can be concluded: the MMT and PCT techniques primary depend on the spatial arrangement of objects in a scene; the spectral properties of these objects are less important. This is shown by the fusion performance of the checkerboard patterns: the fusion results vary between the 1 pixel * 1 pixel and 5 pixel * 5 pixel patterns, whereas the different EM abundances used for the generation do not influence the fusion performance. Only the SMA fusion is affected by the different spectral properties, since it depends on the spectral separation. Noise does not effect the fusion itself, but the noise is also introduced into the fused image, as it is shown when fusing the synthetic scene with different noise levels. Only the MMT-MW algorithm differs, since the lowpass effect of this algorithm reduces the noise but results in a lesser sharpening.

Figure 3: Fusion results - subset of HyMap-scene Barrax
Left: spatially degraded by simulated PSF (GSD ~25m)  
Right: restored to ~5m GSD by MMT-WI Fusion

Figure 4: Fusion results - magnified subsets of HyMap-scene Barrax
R: HyMap Band 25, G: HyMap Band 40, B: HyMap Band 105

Reference  MMT Moving Window Algorithm  PCT Fusion
Without Fusion  MMT Whole Image Algorithm
The evaluation of the **PCT** fusion technique is complicated since the findings for the given scenes are ambiguous. On one hand, the statistical and visual results are generally good, indicating a successful fusion. But on the other hand, the introduction of unpredictable errors results in fused images neither reliable nor usable for further image processing. In addition, since the physical relationship between the image DNs and object sensed is lost, no reasonable analysis based on physical values can succeed. In comparison with other studies (e.g. [13]), it is obvious that the good PCT fusion results in this study are due to fortunate image statistics. The application of the PCT fusion technique is suggested for visualization purpose only.

For the applications presented in §1 of this study, the discrimination of spectrally similar objects (i.e. non-green vegetation and bare soil) is essential. Since the **SMA** technique depends on a good spectrally separation of classes, the utilization results in only marginal improvements for the scenes used in this study. The visual impression of the fused images is coarse, clearly showing a blocky structure and artifacts. Though the image statistics of the unfused image are closer to reference, the reduced sum of error shows a slight improvement. To summarize, the benefits of the SMA fusion for the mapping of non-green vegetation are small, and the utilization of the SMA image fusion technique is dissuaded. For different applications and / or other environments where objects of interest are better spectrally separable, the SMA fusion produces significantly better results.

In contrast, **MMT** fusion technique significantly improves the mapping of non-green vegetation. As mentioned before, the MMT technique depends on the classification of the CI image; therefore, an adequate pre-processing is essential. Furthermore, an adaptation of the MMT parameters to the specific input images is recommended. In contrary to [9], the MMT-WI algorithm performed significantly better than the MMT-MW algorithm for the given scenes in this study. It is supposed that this is due to the usage of a higher class number, thus minimizing the loss in spectral information. Both the visual sharpening impression and the statistical measurements indicate a high improvement when using the WI algorithm, and the change in spectral information is tolerable. The MW algorithm results in a lesser sharpening and a slighter change in spectral information. For the application of this study, the MMT-WI does fulfill all requirements and the application of this fusion technique is suggested. The MMT-MW technique is suitable for applications demanding an even higher spectral fidelity or when sensors with low SNR are to be fused.

**ACKNOWLEDGMENTS**

The authors would like to thank Dr. Boris Zhukov of DLR Berlin-Adlershof for providing the MMT program package.

**REFERENCES**


