

ON SEGMENT BASED IMAGE FUSION

M. Ehlers, A. Greiwe and D. Tomowski

Institute for Geoinformatics and Remote Sensing - IGF
University of Osnabrueck, D-49074 Osnabrueck, Germany
mehlers@igf.uni-osnabrueck.de

KEY WORDS: Image fusion, segmentation, classification, multisensor analyses

ABSTRACT:

In this paper, we will investigate techniques for the combination of image fusion and segment based image analysis. We describe a color preserving iconic fusion with subsequent segmentation and classification, a ‘cookie cutter’ approach for the integration of high resolution RGB and low resolution hyperspectral image data and a decision based fusion technique, that combine high resolution panchromatic data with multispectral images. We will show that the combination of segment based image analysis and fusion techniques at iconic, feature and decision level can indeed improve the final analysis and can be seen as a first step to an automated processing line.

1. INTRODUCTION

The availability of remote sensing data that are needed for global, regional and local monitoring has greatly increased over the recent years. While the increase in spatial resolution for digital images has been hailed as a significant progress, methods for their automated analyses (i.e. land cover mapping, change analysis, GIS integration) are still in the process of being developed. Object (or segment) based preprocessing techniques seem to be an adequate methodology because inter-class variances can be minimized and the image interpretation techniques of the human eye be mimicked. However, the question of appropriate data fusion techniques within this context has hardly been addressed.

Over the last years, image fusion techniques have gained a renewed interest within the remote sensing community. The reason for this is that in most cases the new generation of remote sensors with very high spatial resolution records image datasets in two separate modes: the highest spatial resolution is obtained for panchromatic images whereas multispectral information is associated with lower spatial resolution. The ratios between panchromatic and multispectral imaging modes of one sensor vary between 1:2 and 1:5. For multisensor fusion, ratios can exceed 1:20 (e.g. Ikonos and SPOT merge). Consequently, for analyses that require both, high spatial and spectral information, fusion techniques have to be developed to extract ‘the best of both worlds’. The term ‘fusion’ exists in different forms in different scientific communities (see, for example, Edwards and Jeansouline, 2004).

Usually, the imaging community uses it to address the problem of sensor fusion, where images from different sensors (or different modes of one sensor) are combined. They can be classified into three levels: pixel level (iconic), feature level (symbolic) and knowledge or decision level (Pohl and Van Genderen, 1998).

Until now, of highest relevance for remote sensing data processing and analysis have been techniques for iconic image fusion for which many different methods have been developed and a rich theory exists. Unfortunately, for many fusion techniques we experience more or less significant color shifts which, in most cases, impede a subsequent automated analysis

(Ehlers and Klonus, 2004). Even with a fusion technique that preserves the original spectral characteristics, automated techniques do not produce the desired results because of the high resolution of the fused datasets.

For this purpose, feature based or decision based fusion techniques are employed that are usually based on empirical or heuristic rules. Because a general theory is lacking, fusion algorithms are usually developed for certain applications and datasets. To discuss the advantages and disadvantages on segment based image fusion techniques, we introduce three fusion methods (‘Ehlers fusion’, ‘cookie cutter’ approach and a decision based data fusion) in this paper. We will show, that the feature and decision based fusion approach is the most promising path of the two processing paradigms.

2. METHODOLOGY

2.1 Iconic fusion with segmentation and classification

The test dataset for the demonstration of the iconic fusion is a panchromatic SPOT image (recording date 16 March 2003) with 5 m pixel size and a multispectral Landsat ETM image (bands 1 - 5 and 7) with 30 m ground pixel size. A 1024 x 1024 subset of the SPOT image (Fig. 1) showing a region east of the City of Aachen (Germany) was registered to ground coordinates (German Gauß-Krüger system) and served as the master image. The Landsat image (Fig. 2) was registered to the SPOT image and resampled to 5 m pixel size using a nearest neighbour resampling algorithm.

2.1.1 Fusion technique

An overview flowchart of the Ehlers Fusion is presented in Fig. 3. The first step is to transform the low resolution multispectral image into an Intensity-Hue-Saturation (IHS) image working with three selected bands (RGB). Next, the panchromatic image P and the intensity component I are transformed into the spectral domain using a two-dimensional Fast Fourier Transform (FFT). The power spectrum of both images is used to design the appropriate low pass filter (LP) for the intensity component and high pass filter (HP) for the high resolution panchromatic image. Based on the ratio of pixel sizes between the high and low resolution images, cut-off frequencies for

these filters can be established (Ehlers et al. 1984). Filtering will be directly performed in the frequency domain as it involves only multiplications. An inverse FFT transforms both components back into the spatial domain. The low pass filtered intensity (I^{LP}) and the high pass filtered panchromatic band (P^{HP}) are added and matched to the original intensity histogram. At the end, an inverse IHS transform converts the fused image back into the RGB domain (Fig. 4). More information can be found in Ehlers and Klonus (2004) and Klonus (2005).

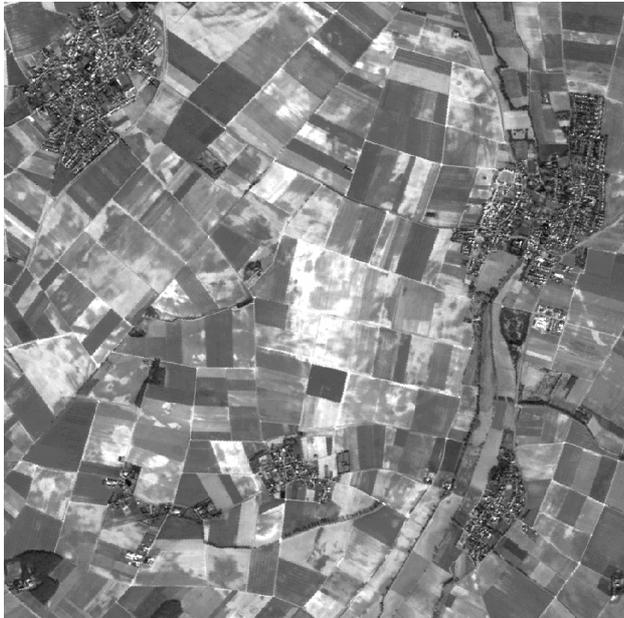


Figure 1. Panchromatic SPOT-5 image of 16 March 2003 (5 m pixel size)



Figure 2. Landsat ETM image (bands 1, 2, and 3) of 26 June 2001 registered to the SPOT-5 image and resampled to 5 m pixel size

The second step is to perform a segment based image analysis on the iconic fused image data. For the segmentation and classification we used the segmentation and minimum-distance algorithm in eCognition. As result (Fig. 5) we created a map with the classes "settlement" and "non-settlement" through a

supervised classification.

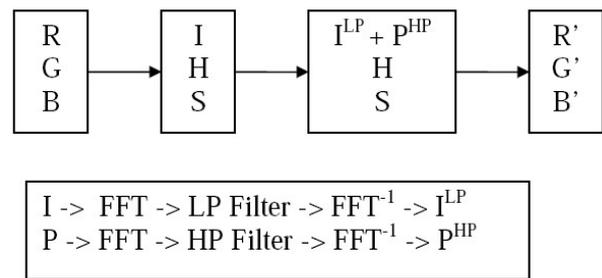


Figure 3. FFT based filter fusion using a standard IHS transform

2.1.2 Results



Figure 4. SPOT/ETM IHS image fusion after FFT based filtering

For a thorough quantitative evaluation, we compared the results of the FFT-based Ehlers fusion (Fig. 4) with the results for the same examination area for a number of standard fusion algorithms. So far, all tests concluded that the Ehlers fusion was optimal for the preservation of spectral characteristics which is required for any subsequent classification or interpretation. (Klonus & Ehlers 2006).

Additionally, we did some preliminary assessments of the results of the segment-based classification approach with the fused image data (Fig. 5). It is evident that the class "settlement" (red) contains agriculture areas (border regions of fields) as well as streets (elongated segments with a similar spectral signature as the class "settlement") outside the settlement regions. Based on manually digitized ground-truth data, we can calculate users' and producers' accuracy (Congalton and Green, 1993). As result, we achieved a users' accuracy of about 50 % and A producers' accuracy of about 88% for the class "settlement". One has to consider that urban classes are usually mixed classes and often show poor results compared to natural classes or water. Also, these are just the preliminary results of some initial tests which will have to be investigated further. However, even at this stage we can

conclude that fusion and subsequent segmentation does not seem to work well for urban settlement detection.



Figure 5. Result after a subsequent segmentation and classification (settlement in red color)

2.2 ‘Cookie cutter’ approach

For an analysis of a cookie cutter approach (i.e. segmentation of the high resolution images and subsequent analysis of these segments in the low resolution images) we used hyperspectral and high resolution ortho image datasets. Different datasets:

- Digital orthophoto data (LEO-Digital Camera Kodak DC14n)
- Digital elevation model (DEM) derived from cadastral data
- Digital surface model (DSM) derived from HRSC-A image data
- Hyperspectral image data (HYMAP)

More information about the sensors can be found in Bäumker et al. (1998), Neukum (1999), Cocks et al. (1998) and Greiwe (2006). The high spatial resolution data produced by LEO were digital airphotos. The photos were resampled to a spatial ground resolution of 0.25 m. An orthoimage was generated using softcopy photogrammetry software with a resulting horizontal accuracy of $s_{x,y} = 0.2$ m. Information about the surface elevation in the study area exists in two datasets:

- Digital elevation model (DEM, derived from cadastral data, grid size 12.5 m, vertical accuracy 0.5 m)
- Digital surface model (DSM, derived from HRSC-A image data, grid size 0.5 m, vertical accuracy 0.2 m)

The DSM was normalized (nDSM) by use of DEM data (Möller, 2003). The HyMap Sensor records 128 reflective bands covering the visible and near infrared range (VNIR) and the short wave infrared domain (SWIR) between 0.4 μm and 2.5 μm . With an operating altitude of 1,500 m and a scan frequency of 16 Hz data could be recorded with a GSD of 3 m across and 4 m along flight track.

2.2.1 Fusion technique

Our methodological approach for data fusion is characterized by an object-oriented segmentation of the geometric high resolution orthophotos and a spectral angle mapper (SAM) score generation of hyperspectral data. The method is based on a mutual support of both data types and a segment based endmember selection. The geometric location of the pixel in the hyperspectral dataset which represents an endmember of an urban surface type is determined by a segmentation of the high resolution image data. Pixels that are fully contained in a segment are candidates for the definition of reference spectra and are considered for the creation of a spectral library.

With the user-specific knowledge contained in spectral libraries, the hyperspectral data are classified by a SAM full pixel classification approach. The classification results are transformed to an 8-bit SAM score by a user-independent automated algorithm (see Greiwe et al., 2004 for more details). Due to the identical geometric registration of both image data the SAM scores provide additional feature values for the image segments of the high geometric resolution orthophoto. The end product of this approach is a map produced by the classified segments. The workflow of our approach is shown in figure 6:

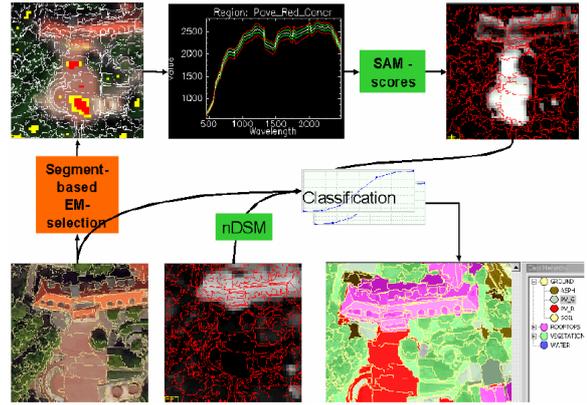


Figure 6. Segments of high resolution data (top left) are used for endmember selection in hyperspectral data (top right). Minimum distance classification and score image are fused using a linear membership function. Results are produced by a neural network classifier

For classification process a score for each pixel of the hyperspectral data has to be determined (SAM score). SAM values are calculated by the cosine of a spectral angle for each given reference spectra:

$$\cos \varphi = \frac{\sum_{i=0}^n e_i r_i}{\sqrt{\sum_{i=0}^n e_i^2} \sqrt{\sum_{i=0}^n r_i^2}} \quad (1)$$

where φ = spectral angle
 e = given image spectra
 r = reference spectra (endmember)
 n = number of classes

This procedure produces a class image and an additional layer

of SAM values, the "rule image" which contains n spectral angles ($\varphi_{1,\dots,n}$, see eq. 1) for each image pixel at n given endmembers. A value near zero in a rule image represents a good fit to a given endmember.

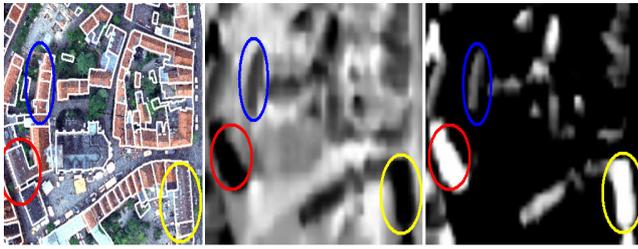


Figure 7. SAM score generation for urban class "dark red roof". High scores are indicated as white.

As shown in Fig. 7, dark red roof tiles in the orthophotos (left) receive a low value in the corresponding rule image after SAM classification (center). The SAM score image (right) is a positive 8-bit gray scale image of the rule image. The transformation is done by the following equation:

$$s_{c,i} = 255 \cdot \frac{\varphi_c^{\max} - \varphi_i}{\varphi_c^{\max} - \varphi_c^{\min}} \quad \text{with } \varphi_c^{\min} \leq \varphi_i \leq \varphi_c^{\max} \quad (2)$$

- where
- $s_{c,i}$ = SAM-score for Pixel i with respect to class c
 - φ_i = spectral angle of investigated pixel
 - φ_c^{\max} = spectral angle which leads to a score of zero for class c (upper border)
 - φ_c^{\min} = best fitting spectral angle for class c which leads to a score of 255.
 - n = number of classes

Further information can be found in Greiwe (2006).

Information about the average height of a segment and the RGB values from the orthophotos can be used as additional feature information. Like a DSM, the 8-bit SAM score layer is also stored in a gray scale image and averaged by overlay operation in a GIS. As a result, for each given class, an average SAM score is available (Fig. 8).

The creation of image objects (segments) and the final classification is performed within the software eCognition. This software provides a neural network classifier which allows the user to create feature specific membership functions considering the SAM scores.



Figure 8. Sam scores for a segment

2.2.2 Results

Nineteen different classes were defined with a differentiation in material classes in order to prove the methodology. It has to be noted that many classes were undistinguishable in the RGB image. For example, red roof tops were divided into "red roof concrete" and "red roof clay". Three different classification scenarios were defined to investigate the performance of the presented approach: A minimum distance classification applied on the RGB feature space of the ortho image, an additional combination with the segment's elevation and at last the implementation of SAM scores into the classification process. For each of the classification scenarios the overall accuracy was estimated (see figure 9). The relative low overall accuracy of the RGB scenario could be explained by the strong similarities of the defined classes in the visible domain.

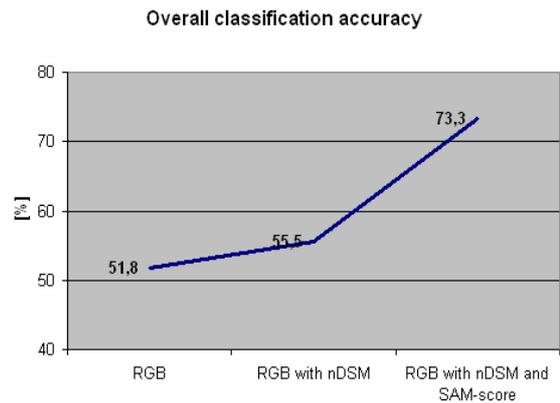


Figure 9. Increasing overall classification accuracy

The cookie cutter approach improved the classification accuracy by nearly 20%. Using this technique, the benefits of an integration of hyperspectral image data into the classification process (e.g. the differentiation of surfaces with similar features but different materials) could be realized.

2.3 Decision based data fusion

As a basis for the decision based fusion process, we selected a high and medium spatial resolution satellites data to develop, implement and test a method for the automated detection of settlement areas.



Figure 10. Panchromatic SPOT-5 image (5 m pixel size)

The high resolution satellite datasets were comprised of panchromatic images from SPOT-5 (Fig. 10) with 5 m GSD and KOMPSAT-1 with 6.6 m GSD (Fig. 11). Medium resolution multispectral data were obtained from Landsat ETM and Aster datasets with 30 m and 15 m resolution, respectively. Our method was applied to two randomly selected test areas (25 km² each), using panchromatic and multispectral satellite data. For the first area, data from SPOT (recording date 16 March 2003) and Landsat (recording date 26 June 2001) were used, and for the second, KOMPSAT-1 (recording date 20 May 2004) and Aster data (recording date 3 August 2003).



Figure 11. Panchromatic KOMPSAT-1 image (6.6 m pixel size)

The aim was to produce a binary mask with the classes "settlement" and "non-settlement". Settlement is understood as a sum of real estates, traffic surfaces, commercial areas, sport and recreation facilities as well as parks and graveyards (Apel and Henckel, 1995).

2.3.1 Fusion technique

Contrary to the iconic image fusion techniques as described above, the images we used were rectified to ground coordinates but otherwise left in their original format. Parameters such as texture and shape were extracted from the high resolution panchromatic data, vegetation information from the multispectral images (Fig. 12).

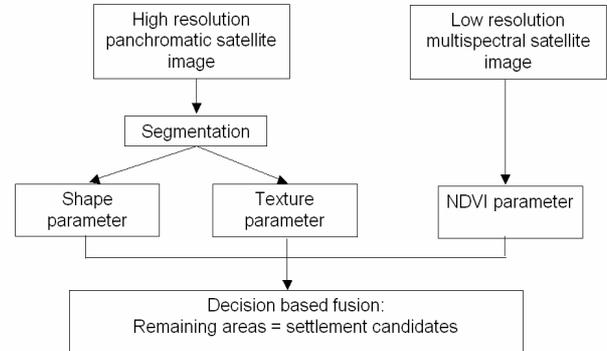


Figure 12. Decision based data fusion process.

Using an adaptive threshold procedure, the information from the image datasets were fused and formed a binary mask for the areas "settlements candidates" and "definitely no settlements". This process was repeated at a hierarchy of differently sized segments with a set of different threshold parameters at each level. The hierarchical network (Fig. 13) of segments consists of three levels.

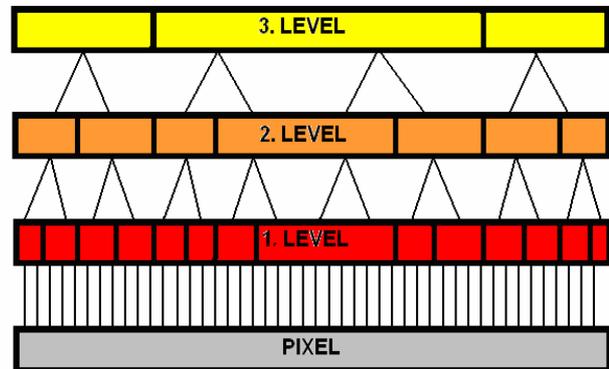


Figure 13. Hierarchical network of segments for the decision based fusion

The size of the segments decreases from level 3 (coarse) to level 1 (fine). The segmentation in eCognition was applied solely to the panchromatic data. The classification algorithm starts at the third level. For each segment of the newly generated class "settlement", texture and form parameters as well as an average NDVI were calculated. The "gray level co-occurrence" (GLC) matrices (Haralick et al., 1973) that examine the spectral as well as the spatial distribution of gray values in the image form the basis for the texture calculation.

A GLC matrix describes the likelihood of the transition of the gray value *i* to the gray value *j* of two neighboring pixels. For the differentiation of "settlement" and "not-settlement" we used the inverse distance moment (IDM) derivative from the GLC matrix. With the application of the IDM, it is possible to

distinguish between heterogeneous and partially homogeneous non-settlement areas (Steinnocher, 1997).

$$IDM = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (3)$$

where N = row or column number
i, j = grey value combination in row i and column j in the GLC matrix
P_{i,j} = appearance probability of a grey value pair

The next step of the method starts at the second segmentation level, in which the threshold values for the classification characteristics (texture, form and NDVI) are increased. Additionally, the classification characteristics are only calculated for the settlement areas (so-called filial segments) that are part of a non-excluding area at the third level (Ehlers et al., 2005). At the first segmentation level, the classification rules are again applied but with highest restriction parameters. Finally, the settlement segments were merged and cleaned by automated filter procedures to eliminate small remaining agriculture segments and to include urban parks and lakes in the settlement areas. Result is a binary mask (called endlevel). More details on this algorithm can be found in Ehlers et al. (2005) and Tomowski et al. (2006).

2.3.2 Results

Despite the differences between the use datasets, the results were very similar (see Fig 14 and Fig. 15). Contiguous settlement areas (conurbation areas) were detected with a high accuracy. For both test areas the borders between “settlement” (red) and “non settlement” (no color) are represented with a low level of generalization (borders are not smooth).

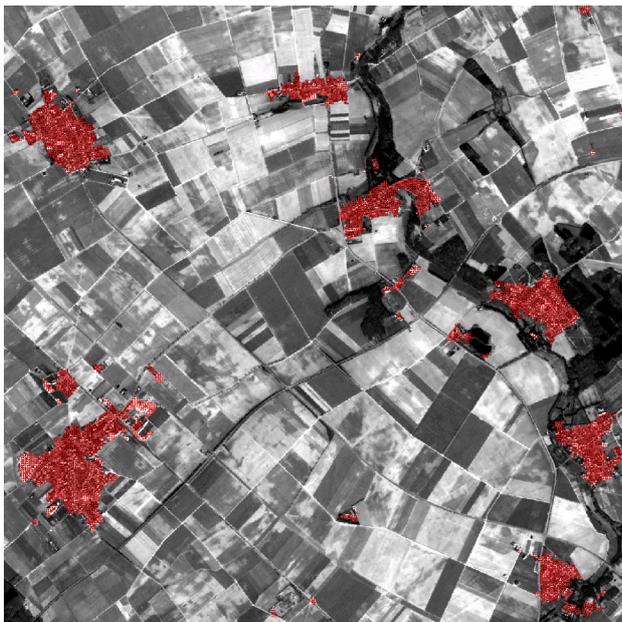


Figure 14. Binary Mask for the SPOT-5 test dataset image (settlement in red color)

It is evident that only a few vegetated areas such as playgrounds or parks are missing and small houses or farms outside the

kernel settlements are not completely included.



Figure 15. Binary Mask for the KOMPSAT-1 image (settlement in red color)

To analyze the final accuracy, settlement areas were manually digitized and compared to the results of the hierarchical processing at each level (Tab. 1)

Hierarchical Level	SPOT-5 / Landsat ETM	KOMPSAT / Aster
3	13.57%	45.28%
2	69.98%	84.18%
1	86.90%	95.03%
Final	96.34%	97.26%

Table 1. User accuracy for the detection of settlement areas

For both combinations, results are almost identical and exceed 95% user accuracy at the final level. Kappa values are 0.8427 and 0.8968 for the first and the second test area, respectively (Cohen, 1960). On an evaluation scale as proposed by Ortiz et al. (1997), which ranges from "very bad" to "excellent", the results for both test areas can be regarded as "excellent".

3. CONCLUSIONS

All presented fusion techniques use the benefits of a combination of high spatial and high spectral resolution remote sensing.

The iconic Ehlers fusion has shown much better results than all other standard and newly proposed iconic fusion techniques (Klonus and Ehlers, 2006). As result we can summarize, that the Ehlers fusion integrates color and spatial features from multispectral and panchromatic images, by minimizing color distortion, so that the fused image has almost identical spectral characteristics as the original image. However, the benefits of the introduced iconic image fusion do not automatically lead to enhanced classification results for the introduced segment oriented classification approach.

The presented feature based ‘cookie cutter’ approach uses the benefits of a fusion of high spatial (orthophoto) and high spectral resolution image data (hyperspectral) at the feature level. It

could be proved that a segment based endmember selection results in a suitable spectral library (Greiwe 2006). With the automated SAM score generation, additional feature values for the image segments could be generated. As result, the additional inclusion of hyperspectral image data into a classification process of high spatial resolution image data shows significant improvements (Fig. 9) and allows material differentiation of urban surfaces.

With the introduced decision based fusion technique we created an efficient and accurate semiautomatic procedure for the detection of settlement areas. Through the segment and hierarchical classification approach it was possible to improve the classification results at each classification step. Furthermore, this procedure works equally well with different multisensor satellite data without altering the procedure or the employed parameters step (Tomowski et al., 2006).

In the comparison to the pixel-based based classification procedures, (like the maximum likelihood method) it is evident that the introduced feature ('cookie cutter') and decision based fusion techniques are significant improvements for the design of future automated result driven processing lines. Through the adoption of object oriented (better: segment oriented) image processing methods and data fusion techniques it is possible to avoid inter-class variances (like the salt-and-pepper effect (Meinel et al., 2001) and to enhance the classification accuracies. In our opinion, a feature and/or decision based fusion seems to be the most promising technique for the improvements of classification accuracy.

4. REFERENCES

- Apel, D. and Henckel, D. 1995. Flaechen sparen, Verkehr reduzieren – Möglichkeiten zur Steuerung der Siedlungs- und Verkehrsentwicklung. In: *Difu-Beitraege zur Stadtentwicklung*, 16, Deutsches Institut für Urbanistik (Berlin), pp. 29-40.
- Bäumker, M, Brechtken, R., Heimes, F.-J. & Richet, R., 1998. Hochgenaue Stabilisierung einer Sensorplattform mit Daten eines (D)GPS-gestützten Inertialsystems. *Zeitschrift für Photogrammetrie und Fernerkundung ZPF*, pp. 15-22.
- Cocks, T., Jansen, R., Stewart, A., Wilson, I. & Shields, T., 1998. The Hymap airborne hyperspectral sensor: system, calibration and performance. In: *Proceedings of 1st EARSeL Workshop on Imaging Spectroscopy*. Zürich, Switzerland, pp. 37-43.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educ and Psychol Meas*, 20, pp. 37-46
- Congalton, R. & Green, K., 1993. A Practical Look at the Sources of Confusion in Error matrix Generation. *Photogrammetric Engineering & Remote Sensing*, 59(5), pp. 641-644.
- Edwards, G. & Jeansoulin, R., 2004. Data fusion - from a logic perspective with a view to implementation. *International Journal of Geographical Information Science*, 18(4), pp. 303-307.
- Ehlers, M., 1984. *Digitale Bildverarbeitung*. Schriftenreihe des Institutes für Photogrammetrie und Ingenieurvermessungen, 9, Universität Hannover.
- Ehlers, M. & Klonus, S., 2004. Erhalt der spektralen Charakteristika bei der Bildfusion durch FFT basierte Filterung. *Photogrammetrie-Fernerkundung-Geoinformation*, 6, pp. 495-506.
- Ehlers, M., Michel, U., Bohmann, G. & Tomowski, D., 2005. Entscheidungsbasierte Datenfusion von multisensoralen Fernerkundungsdaten zur Erkennung von Siedlungsgebieten. In: *GEO-GOVERNMENT - Wirtschaftliche Innovation durch Geodaten - Vorträge 25. Wissenschaftlich-Technische Jahrestagung der DGPF*, 14, pp. 209-216.
- Greiwe, A., M. Bochow & M. Ehlers, 2004. Segmentbasierte Fusion geometrisch hochauf-gelöster und hyperspektraler Daten zur Verbesserung der Klassifikationsgüte am Beispiel einer urbanen Szene, *Photogrammetrie-Fernerkundung-Geoinformation (PFG) 6/2004*, pp. 485-494.
- Greiwe A., 2006. Detektion von Referenzspektren in multisensoralen Bilddaten. Dissertation at the University of Osnabrueck. http://elib.uni-osnabrueck.de/publications/diss/E-Diss556_the-sis.pdf (accessed 31. May 2006)
- Klonus, S., 2005: *Untersuchungen zu Datenfusionsverfahren in der Fernerkundung*. MSc Thesis, University of Vechta (unpublished).
- Klonus, S. & Ehlers, M., 2006. Spectral Characteristics Preserving Image Fusion with the Ehlers Fusion Algorithm (in preparation)
- Meinel, G., Neubert, M. and Reder, J., 2001. Pixelorientierte versus segmentorientierte Klassifikation von IKONOS-Satellitenbilddaten - ein Methodenvergleich. *Photogrammetrie Fernerkundung Geoinformation*, 3, pp. 157-170.
- Möller, M., 2003. *Urbanes Umweltmonitoring mit digitalen Flugzeugscannerdaten*. Wichmann Verlag, Heidelberg.
- Neukum, G., 1999. The Airborne HRSC-A: Performance Results and Application Potential. In: *Photogrammetric Week 1999*. Wichmann Verlag, edited by Fritsch, D. and Spiller, R., pp. 83-88.
- Ortiz, M., Formaggio, R. and Epiphanyo, J., 1997. Classification of croplands through integration of remote sensing, GIS and historical database. *International Journal of Remote Sensing*, 18, pp. 95-105.
- Pohl, C. & van Genderen, J., 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing*, 19, pp. 823-854.
- Steinnocher, K., 1997. Texturanalyse zur Detektion von Siedlungsgebieten in hochauflösendenpanchromatischen Satellitenbilddaten. In: *Salzburger Geographische Materialien (AGIT IX, 2.-4. Juli)*, 26, pp. 143-152.
- Tomowski, D., Ehlers, M., Michel, U. & Bohmann, G., 2006: Objektorientierte Klassifikation von Siedlungsflächen durch multisensorale Fernerkundungsdaten. In: *GI-reports@igf*, 3, Universität Osnabrück (in print).