FUSION OF OPTICAL IMAGERY AND SAR/INSAR DATA FOR OBJECT EXTRACTION

Olaf HELLWICH, Manfred GÜNZL and Christian WIEDEMANN
Chair for Photogrammetry and Remote Sensing
Technische Universität München, D-80290 Munich, Germany
Phone: +49/89/289 22677, Fax: +49/89/280 9573
E-mail: Olaf.Hellwich@photo.verm.tu-muenchen.de
URL: http://www.photo.verm.tu-muenchen.de

Working Group III/6

KEY WORDS: Data fusion, Multi-spectral data, Radar, Multi-temporal, Object recognition

ABSTRACT

Optical imagery such as high-resolution panchromatic or multispectral data, and SAR/INSAR data show different information about the imaged objects, and have different advantages and disadvantages when used for object extraction or landuse classification. Multispectral optical image data is largely determined by the type of the material an object consists of. Panchromatic data which is often available with a higher resolution than multispectral data emphasizes geometric detail of the objects, e.g. the complex structure of anthropogenic objects such as road networks. In contrary to this, SAR data contain information about small-scale surface roughness and - to a lower degree - soil moisture. Height information derived by interferometric processing of SAR data contains large-scale surface roughness. These different types of information are referring to completely different object qualities and are, therefore, largely uncorrelated which helps to reduce ambiguities in the results of object extraction. The main advantage of passive optical imagery with respect to SAR data is the lack of the speckle effect leading to images with a far better extractability of linear as well as areal objects. A major advantage of SAR is its all-weather capability which allows the acquisition of time series of imagery with exact acquisition dates under any climatic conditions. In this paper, these complementary properties of SAR and optical image data are demonstrated and used to improve object extraction and landuse classification results.

1 INTRODUCTION

A concept for the fusion of optical image data with SAR/INSAR data for scene interpretation is presented. In Section 2 it is formulated as a semantic network which is implemented in form of a Bayesian network incorporating uncertainty information. The uncertainty information is provided by different algorithmic components of the developed approach using approximate probabilistic reasoning. In the Bayesian framework also contextual information is considered. Such information is e.g. that agricultural fields commonly have a rectangular shape. A shape parameter for the segmentation of grid data allowing to judge computationally very efficiently the deviation of the segment's shape from a rectangle is introduced in Section 3. In Section 4 multisensor fusion is demonstrated using two examples. In the first example, boundaries of areal objects and roads are extracted from optical data, and most of the landuse classes are found with the help of multispectral information. In the second example, a multitemporal SAR data set is combined with a multispectral optical image for the purpose of landuse classification.

2 BAYESIAN NETWORK FOR MULTISENSOR MULTITEMPORAL CONTEXTUAL CLASSIFICATION

In this section a Bayesian network for classifying multisensor multitemporal image data considering contextual criteria is introduced. Only the application of Bayesian networks is explained; for the theoretical background we refer to the literature. On the conceptual level the Bayesian network is based on a semantic network as previously described in (Hellwich and Wiedemann, 2000). The relations between real world objects and image data are modeled on three levels. The *real world level* contains the topographic objects such as forest, the *sensor level* the image data. The *geometry and material level* plays a mediating and connecting role between *sensor* and *real world level*. Its task is to take into account that the objects are often not directly the cause of the data contained in the images, but that certain material or geometric properties of the objects are more directly linked to the measured data.

Figure 1 shows a simple example of a Bayesian network for classification of image data. In this case the nodes or primitives of the network correspond with the pixels of the image. The network is used for landuse classification which is

expressed by a *landuse* node on the *real world level*. On the *sensor level multispectral* data and interferometric synthetic aperture radar (*INSAR*) data nodes are shown. The connecting *geometry and material level* consists of *green vegetation*, *wood*, and *soil* nodes for the evaluation of multispectral data, and *small-scale roughness* and *large-scale roughness* for *INSAR* data. The states of the *landuse* node are landuse classes, e.g. forest, agricultural vegetation, and build-up areas. The states of the sensor nodes are for instance grey value vectors; those of the material nodes numerical degrees of presence. The direction of the arrows indicates that the objects to be extracted from the observations are introduced as root nodes and the observations as leaf nodes (cf. (Pearl, 1988, page 157ff.) versus (Kulschewski, 1999b, Kulschewski, 1999a, page 28f.)). This means that the observations are considered functions of the material properties of the object, and that the material itself is seen as a function of the object. In this way the direct sensor model is implemented through the *a priori* probability density function (pdf) of the root node and the conditional pdf's of the leaf nodes and all remaining nodes. Thus, the *landuse* node is described by the *a priori* pdf, i.e. by probability values for landuse variable ε such as $p(\varepsilon = \text{forest}) = 0.3$. The *multispectral* image data node is given a conditional pdf $p(\boldsymbol{x}_m|\boldsymbol{m}_g, \boldsymbol{m}_w, \boldsymbol{m}_s)$ where $\boldsymbol{x}_m$ is the grey value vector of the multispectral image data, $\boldsymbol{m}_g$ the degree of presence of green vegetation, $\boldsymbol{m}_w$ the degree of presence of wood, and $\boldsymbol{m}_s$ the degree of presence of soil.

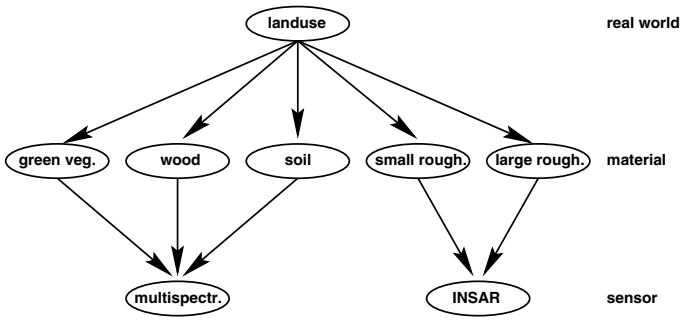


Figure 1: Bayesian network for multisensor landuse classification. (small rough.: small-scale roughness, large rough.: large-scale roughness)

For the classification of a pixel the leaf nodes are instantiated with the corresponding grey value vectors from the image data sets. Then the Bayesian network is evaluated, i.e. the probabilities are propagated according to the respective formulas, e.g. given by (Pearl, 1988, Koch, 2000, Kulschewski, 1999b), using the pdf's.

The model for pixel-based classification is extended to process multitemporal data allowing changes of the landuse node in time. Figure 2 shows the corresponding Bayesian network, in this case for one multispectral data set and multitemporal INSAR data sets acquired at later points in time. The arrows between the landuse nodes indicate the dependence of the state of the node at time t_i on the state of the node at time t_{i-1}. In the Bayesian network this dependence is implemented with the help of conditional pdf's $p(\varepsilon_i|\varepsilon_{i-1})$. In the extreme cases this function either does not allow any changes or it does not favor any particular state ε_i with respect to the previous state ε_{i-1}. In the first case, for the function values $p(\varepsilon_i|\varepsilon_{i-1}) = 0$ holds whenever $\varepsilon_i \neq \varepsilon_{i-1}$ which means that the landuse nodes could be united. In the second case, $p(\varepsilon_i|\varepsilon_{i-1}) = \text{const.}$, independent of the state of ε_{i-1} which means that all classes are equally probable, i.e. that the landuse nodes could be disconnected. In general the conditional pdf's of the landuse nodes give transition probabilities of landuse in time (Bruzzone and Serpico, 1997). For example, when in a certain area a change from forest to built-up land is to be expected this is expressed by the pdf. The conditional pdf's between landuse nodes developing in time can also include models describing dynamic developments such as growth and harvest models.

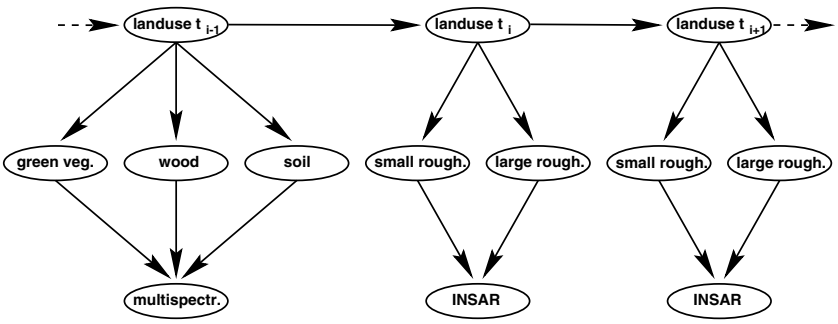


Figure 2: Bayesian network for multitemporal multisensor landuse classification.

The Bayesian network can be extended to incorporate spatial contextual information. For instance for the determination

of the state of the landuse node the state of neighbouring pixels can be considered. The direction of the arrows in Figure 3 indicate that the state of the central node, i.e. the node to be evaluated, depends on the state of the neighbouring nodes. A reasonable form of dependence could for instance be that a pixel has landuse forest with a high probability, if its neighbours are forest. As in reality also the neighbouring nodes depend on the state of the central node, i.e. as the dependence is also valid in the opposite direction, the notion of mutual dependence would result in an undirected graph including all pixels of the image. Then the Bayesian network for the classification of a single pixel would become a Markov random field $\boldsymbol{X}$ over the complete image whose joint pdf $p(\boldsymbol{X})$ could be evaluated using the well-known optimization methods for Markov random fields (Winkler, 1995, Koch and Schmidt, 1994, Hellwich, 1997). Figure 4 shows an example of a Markov random field defined on a grey value image using eight-neighbourhoods.

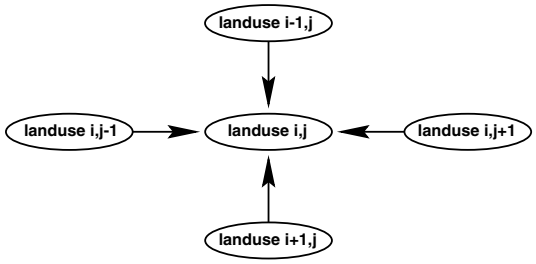


Figure 3: Bayesian network considering spatial contextual information. The indices indicate row and column of a raster image.

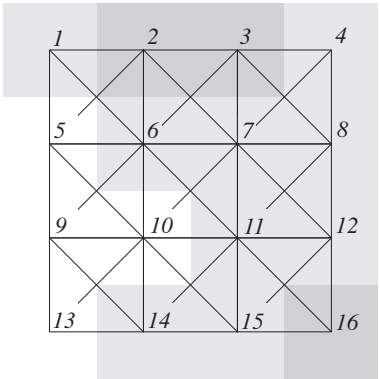


Figure 4: Markov random field defined on a grid (here: grey value image).

Furthermore, the Bayesian network can be extended to accommodate grouping and hierarchy of object parts or objects, and relations between different types of objects [1]. For instance, several pixels with a certain landuse could be considered as components of an agricultural field. Agricultural field objects could consist of surrounding edges and roads, and pixels enclosed by those edges.

In the following section a shape criterion for a group of raster elements is introduced. It is used to judge whether a group of pixels has a shape similar to a rectangle, usually the shape of an agricultural field, or not. If the group of pixels approximates a rectangle, this contributes to a high probability to be an agricultural field; otherwise this probability decreases.

3 SEGMENT SHAPE PARAMETER FOR GRID DATA

A criterion was developed which allows to judge the shape of a group of connected pixels of an image. While it is described in depth in (Günzl and Hellwich, 2000), here a short comment on its use in a region-growing algorithm is given.

The purpose of the criterion is to determine the compactness of an object out of grid data by compensation of the influence of the imaged grid. The compensation was found to be possible using a set of geometrical parameters. The concept lead to a shape parameter with several advantages:

1. The parameter is completely independent of the orientation of an object regarding the grid.

[1] For this purpose it is advantageous to make use of dynamic Bayesian networks (Kulschewski, 1999b, Murphy, 2000).

2. The geometrical parameters of a segment derived out of merged segments can be determined from the parameters of the original segments. Thus, applied to region growing by merging, the computational effort is independent of the shape, size and girth of the regions and therefore constant. This enables shape-controlled region growing with the same order of computational effort as without shape control.
3. The same set of parameters also enables the compensation of the lateral ratio of rectangular objects. This can be carried out independently of the orientation compensation.

To be independent of user interaction the region-growing-by-merging algorithm starts with all pixels as competing seed points. Initially, every pixel builds a region with four boarders to its neighbouring pixels. With every step the one boarder that is minimizing a given criterion is removed.

This quality criterion of a region is defined on the basis of radiometric and shape characteristics as well as computational requirements. A region should be homogeneous and have a rectangular, not a fractal, shape. The merging criterion therefore must measure boarder smoothness. One easy to calculate parameter is the enclosed area divided by the boarder length. The problem is that, if applied on grid data, this gives preference to squares with boarders parallel to the grid. In contrary to this, for grid data given the number of pixels P within the region, the number of edges E and corners C forming the boarder the new parameter describes the shape of the area by:

$$r_{PEC} = \frac{2E^2 + 16 - C^2}{32P} \tag{1}$$

It can easily be shown that every square with parallel or diagonal boarders derives to $r_{PEC} = 1$ (cf. Fig. 5). Other shapes get higher values. For any other orientation of the square (Günzl and Hellwich, 2000) show that deviations can easily be compensated for by a simple extension to equation 1. The same is valid for the ratio of the side lengths of rectangles such that any rectangle with arbitrary orientation with respect to the grid is provided with the optimum parameter value.

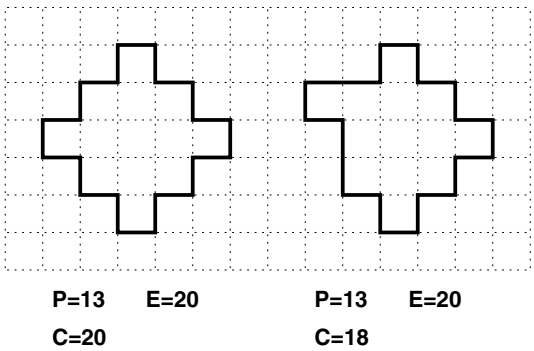


Figure 5: Segment shape parameter for grid data. In case of the left strictly squared region $r_{PEC} = 1$, whereas in case of the equally large right region which is not squared $r_{PEC} \approx 1.2$.

To implement the criterion of equation 1 it is necessary to continuously keep track of the boarder length and the number of corners within each boarder. Therefore, the boarders were implemented as cyclic doubly-linked lists. Every boarder is part of both lists of two neighbouring regions. The algorithm keeps track of the number of edges E, the number of corners C and the angle with which it is connected to the following boarder sections within both chains. The boarders are organized within a binary heap data structure (Cormen et al., 1990, page 140) enabling a merging to be performed with an effort of order $O(m \ln b)$ where m is the mean number of neighbours a region has, and b is the number of boarders surrounding and separating from each other the original pixels of the image.

Summarizing, the parameter describes the deviation of the shape of a segment from a rectangle. It is "1" whenever the segment is rectangular – independent of the location and orientation of the segment with respect to the grid, as well as of the ratio of the side lengths of the rectangle. When the segment increasingly deviates from the rectangular shape, the parameter increases continuously. As a result this computationally efficient shape parameter is very well suited to segment areas of agricultural landuse usually consisting of rectangles.

4 RESULTS

In this section, two examples of scene interpretation based on multisensor fusion are shown. The theoretical framework described in the previous sections has not yet been fully applied for the processing of these examples. Nevertheless, the preliminary results already show promising results.

In the first case a high-resolution (1m pixel size) panchromatic image (Fig. 6 a)) is combined with a multispectral image (Fig. 6 b), 4m pixel size) and a digital surface elevation model (DEM) with 1 m grid size. The imaged area is located in Upper Bavaria and consists of agricultural fields, patches of forest, and rural residential areas. The road network was extracted automatically from the high-resolution image starting in open areas detected by landuse classification. The multispectral data was analyzed using unsupervised clustering and classified into 15 landuse classes with an algorithm including a Markov random field model supporting the detection of continuous areas belonging to a single class. Ambiguities between non-vegetated areas and buildings were solved with the help of a high-objects class extracted from the DEM (Fig. 6 c)). Edges from the high-resolution image were combined with the road network in order to extract agricultural field units. Inside of an agricultural field a majority filter was applied to the classification of the multispectral data selecting the mode of the classes occuring in a field as the landuse of the complete field. The resulting classification (Fig. 6 d)) was compared with a limited set of ground truth data. Of the eight agricultural fields with known crop one summer wheat field was misclassified as winter wheat. The classification procedure is described in detail in (Hellwich and Wiedemann, 2000). In addition to the usefulness of multisensor fusion, this example demonstrates that also the combined extraction of different object types, here agricultural fields and roads, improves the accuracy and reliability of the resulting interpretation of the scene.

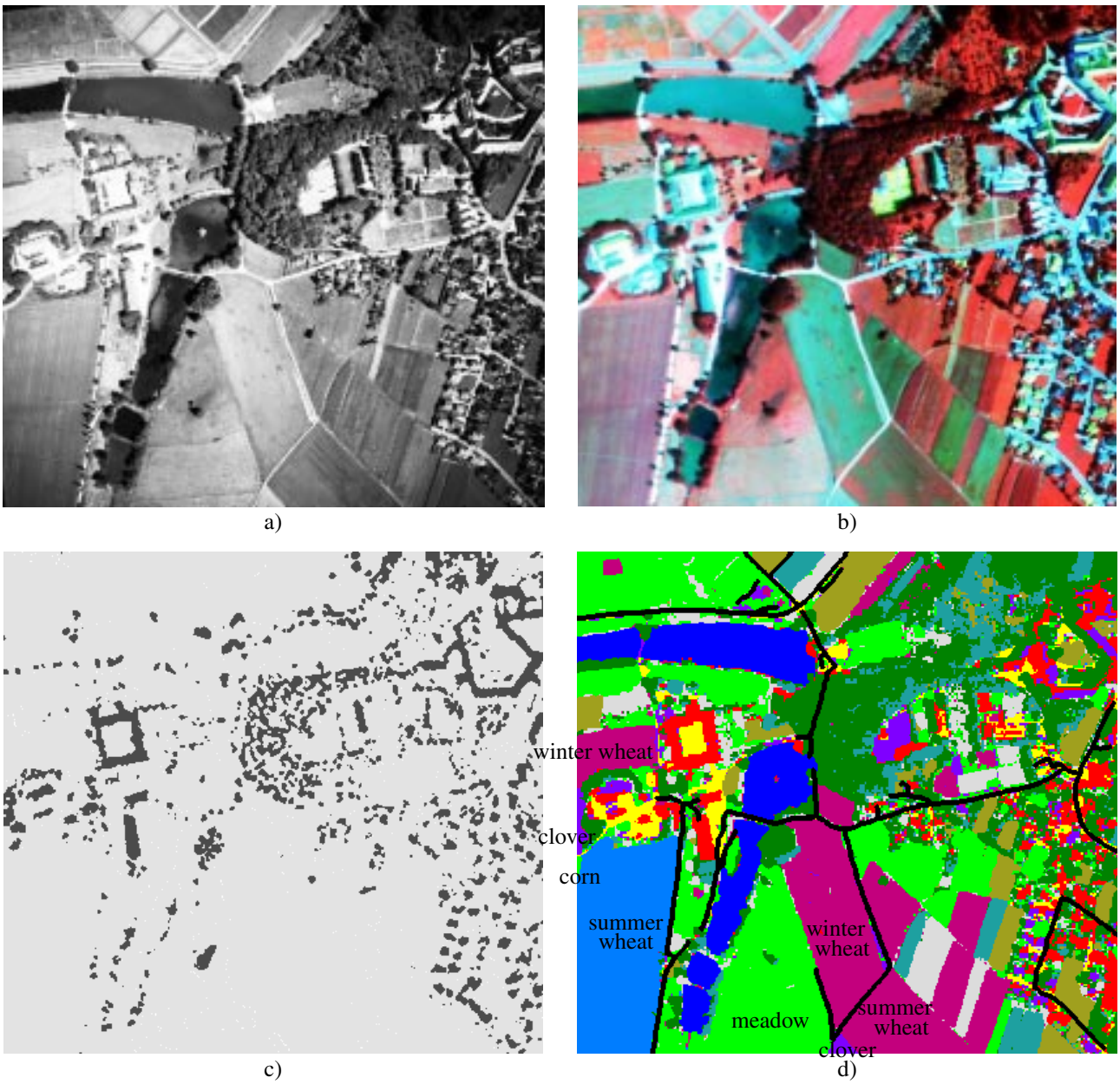
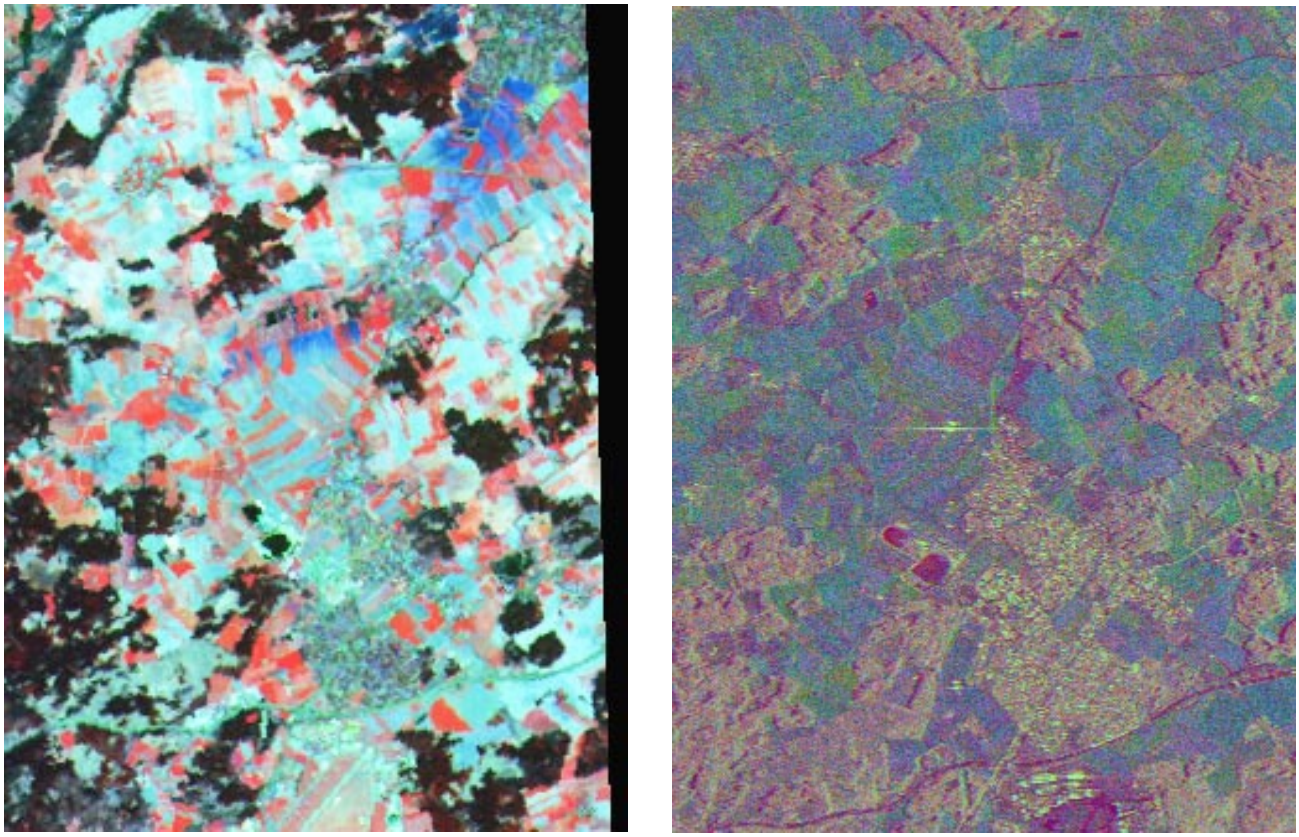


Figure 6: a) High-res. panchromatic image, b) IR color image, c) high objects (dark) extracted from DEM, d) landuse classification (with imprinted ground truth information).

For the second example, 23 ERS-1/2 intensity data sets were acquired during the growing seasons of 1996, 1997, and 1998. Figures 8 a) to c) show a section of three ERS-2 images taken in May, June and July 1996. Furthermore, a Landsat TM data set of April 1996 was available (Fig. 7 a) and Fig. 8 d)). First, the ERS data was subjected to a principle component analysis (Fig. 7 b)). Using the first three principle components the data was clustered into eight classes. The classification results (Fig. 8 e)) show a high degree of fragmentation due the speckle effect. Then, agricultural field segments were extracted from the Landsat data. As in the first example, inside of the field segments a majority filter was applied to the landuse classification results. The results of this step are shown in Figure 8 f). A comparison with the original ERS classification (Fig. 8 e)) reveals that segment boundaries resulting from the ERS data should have been added to the segment boundaries stemming from the Landsat TM data before applying the majority filter.



a) b)

Figure 7: a) Landsat TM image (red: band 4, green: band 3, blue: band 7), b) first principal components of multitemporal (23 dates) ERS-1/2 SAR data.

5 CONCLUSIONS

An approach to scene interpretation using multisensor fusion was introduced. It is suggested to extract geometric details from high-resolution panchromatic optical imagery, material properties from multispectral optical imagery, details of object development in time from SAR data, and to solve remaining ambiguities with the help of three dimensional data. This work will be continued to further integrate the various methods of object extraction and classification into a Bayesian network framework using uncertainty information provided by the individual methods.

ACKNOWLEDGMENTS

The authors thank E. Aigner, F. Kurz, and R. Ludwig for providing some of the data sets.

REFERENCES

Bruzzone, L. and Serpico, S. B., 1997. An Iterative Technique for the Detection of Land-Cover Transitions in Multitemporal Remote-Sensing Images. IEEE Transactions on Geoscience and Remote Sensing 35(4), pp. 858–867.

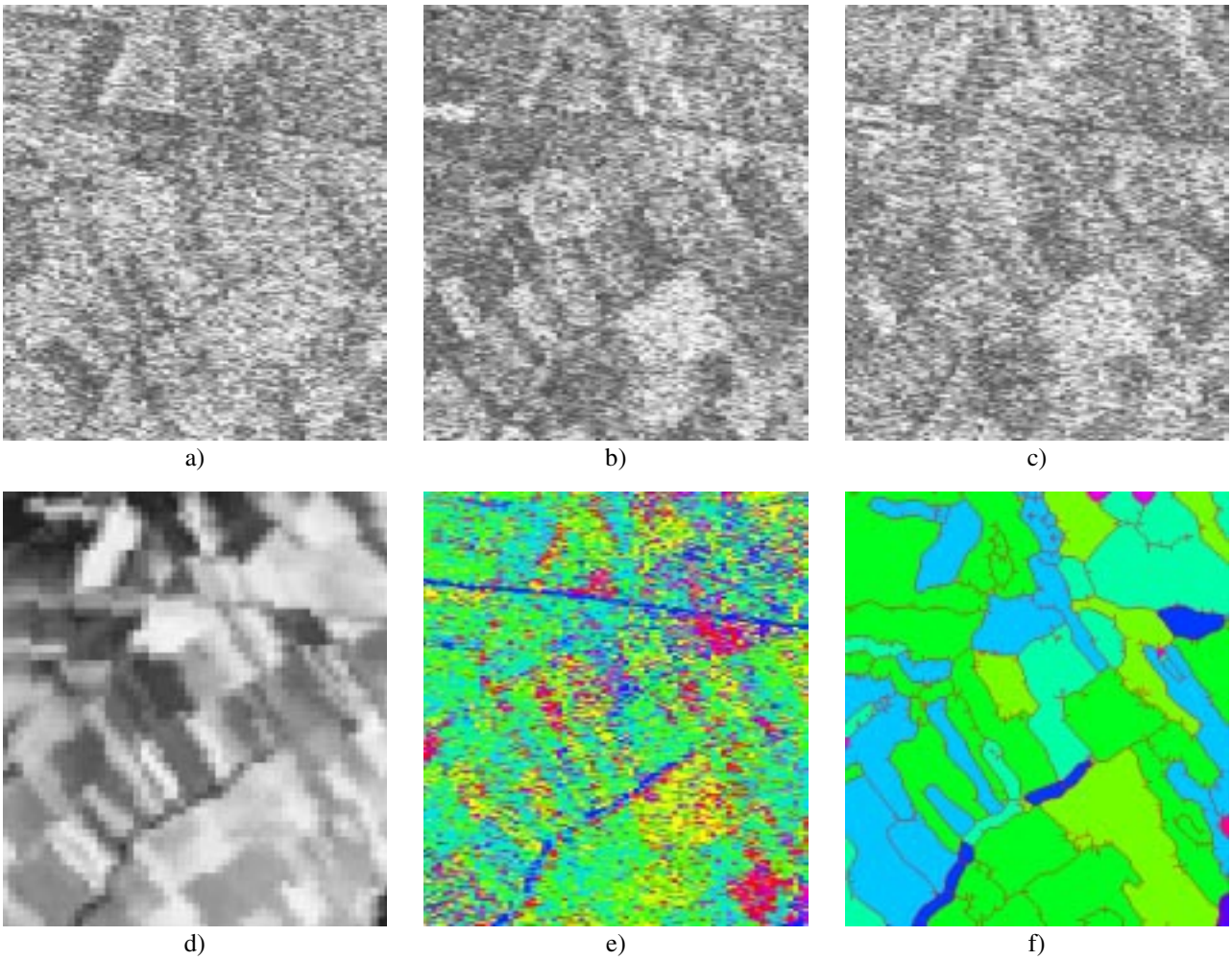


Figure 8: a) to c) ERS-2 SAR intensity data of 96/05/11, 96/06/15 and 96/07/20, d) Landsat TM data (band 4) of 96/04/21, e) classification of ERS data, f) classification of ERS data based on agricultural field segments from TM data.

Cormen, T. H., Leiserson, C. E. and Rivest, R. L., 1990. Introduction to Algorithms. McGraw-Hill, New York.

Günzl, M. and Hellwich, O., 2000. A New Segment Shape Parameter for Grid Data and its Application to Landuse Segmentation. In: ISPRS Congress Amsterdam. in preparation, pre-version available.

Hellwich, O., 1997. Linienextraktion aus SAR-Daten mit einem Markoff-Zufallsfeld-Modell. Reihe C, Vol. 487, Deutsche Geodätische Kommission, München.

Hellwich, O. and Wiedemann, C., 2000. Object Extraction from High-Resolution Multisensor Image Data. In: Third International Conference Fusion of Earth Data, Sophia Antipolis, pp. 105–115.

Koch, K.-R., 2000. Einführung in die Bayes-Statistik. Springer-Verlag, Berlin.

Koch, K.-R. and Schmidt, M., 1994. Deterministische und stochastische Signale. Dümmler, Bonn.

Kulschewski, K., 1999a. Modellierung von Unsicherheiten in dynamischen Bayes-Netzen zur qualitativen Gebäudeerkennung. Shaker Verlag, Aachen.

Kulschewski, K., 1999b. Recognition of Buildings Using a Dynamic Bayesian Network. In: W. Förstner, C.-E. Liedtke and J. Bückner (eds), Semantic Modeling for the Acquisition of Topographic Information from Images and Maps SMATI 99, pp. 121–132.

Murphy, K. P., 2000. Bayes Net Toolbox for Matlab 5. http://www.cs.berkeley.edu/~murphyk/bayes/bnt.html.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, San Mateo, CA.

Winkler, G., 1995. Image Analysis, Random Fields and Dynamic Monte Carlo Methods. Applications of Mathematics, Vol. 27, Springer-Verlag, Berlin.