

EXTRACTING BUILT-UP AREAS FROM MULTITEMPORAL INTERFEROMETRIC SAR IMAGES

Leena Matikainen ^a, Juha Hyyppä ^a, Marcus Engdahl ^b

^a Department of Remote Sensing and Photogrammetry, Finnish Geodetic Institute,
P.O. Box 15, FIN-02431 Masala, Finland – Leena.Matikainen@fgi.fi, Juha.Hyyppa@fgi.fi

^b Laboratory of Space Technology, Helsinki University of Technology,
P.O. Box 3000, FIN-02015 HUT, Finland – mengdahl@avasun.hut.fi

KEY WORDS: Land Use, Urban, SAR, Segmentation, Classification, Mapping, Multitemporal, Interferometry

ABSTRACT:

Automatic extraction of built-up areas from a multitemporal interferometric ERS-1/2 Tandem SAR dataset was studied. The image data were segmented into homogeneous regions and the regions were classified using their mean intensity and coherence values and additional contextual information. According to comparison with a set of reference points, an overall classification accuracy of 97% was achieved when classifying the dataset into three classes: built-up area, forest and open area. Reference points in densely built-up urban areas were recognized as built-up with 100% accuracy. In small-house areas the percentage of reference points correctly classified as built-up ranged from 66% to 94%, depending on the channel combination and classification rules used. Use of texture or rules related to classes of neighbouring objects improved the accuracy. The classification process was highly automatic; training areas covered only about 0.12% of the study area and did not have any overlap with the reference points used in accuracy estimation. Built-up areas could be recognized clearly better than in some previous studies with interferometric ERS data. Possibility to classify built-up areas further into subclasses was investigated using digital map data. The results suggest that the built-up classes of Finnish 1:50 000 topographic maps (small-house areas, apartment house areas and industrial areas) are difficult to distinguish reliably from each other. On the other hand, a correlation was found between the percentage of an area covered with buildings and the mean intensity and coherence of the area in the imagery. This information was used to classify built-up areas into subclasses.

1. INTRODUCTION

Method development for automatic mapping and map updating using remotely sensed data is currently an important research topic. In addition to aerial photographs and optical satellite imagery, Synthetic Aperture Radar (SAR) data attract increasing attention. SAR images provide information complementary to that obtained from optical data and can be useful even if acquisition of optical images is difficult due to cloudiness or darkness. In the future, the importance of SAR data in mapping applications is expected to increase due to new systems providing data with high spatial resolution and with a variety of frequencies, polarization modes and incidence angles.

Previous studies have shown that interferometric SAR data including both backscatter intensity and coherence information have high potential for land-use mapping (Wegmüller and Werner, 1997; Dammert et al., 1999; Strozzi et al., 2000). For example, Strozzi et al. (2000) used different algorithms and test areas in Europe and their results suggested that land-use classification accuracies around 75% are possible with, in the best case, simultaneous forest and nonforest accuracies of around 80-85%. The accuracy of urban areas, however, was only around 30%. Problems in discriminating urban areas were also encountered by Dammert et al. (1999). In these previous studies, various numbers of European Remote Sensing Satellite (ERS) SAR images and features extracted from them have been used. Classification algorithms applied have been pixel-based methods.

The goal of the present study was to investigate how built-up areas could be extracted automatically from a multitemporal interferometric ERS-1/2 Tandem dataset and evaluate the applicability of the data to automatic updating of land-use

maps. Built-up areas in SAR images are typically clearly visible for the human eye, but due to their heterogeneous appearance, automatic extraction can be difficult. For example, a suburban area is typically comprised of houses, roads, yards and different types of vegetation. In an image with spatial resolution of 20-30 m, this area appears as a group of mixed pixels together with pure pixels from various land-cover classes. For map updating, detection of new built-up areas is one of the most important tasks.

An interferometric dataset created previously at the Helsinki University of Technology (HUT) within ESA AO3-277 project was used in the study. According to visual evaluation, the dataset is promising for land-use mapping applications. Forests, agricultural fields, urban areas and even main roads and railways are visible in the imagery. Due to the relatively low spatial resolution of the ERS data, automatic extraction of small or narrow objects, such as roads, is likely to be difficult, but for mapping land-use classes in a coarser scale the dataset could be useful. Engdahl and Hyyppä (2002) achieved an overall accuracy of 90% when classifying the dataset into six classes by using an unsupervised ISODATA classifier.

2. DATA

2.1 SAR imagery

A total of 14 ERS-1/2 complex image pairs acquired during the ERS Tandem mission in 1995-1996 were processed at HUT to create the dataset used in the study. The dataset covers Helsinki and its surroundings with a variety of built-up areas as well as forests and agricultural land. Total land area covered is about

1800 km². Processing of the data is described in the following sections. The final dataset included eight channels:

1. Temporal average of the backscattered intensity images
2. Temporal average of the Tandem coherence images
3. Average of two coherence images with long temporal baselines
4. The 1st principal component (PC) calculated from the Tandem coherence images
5. The 2nd PC calculated from the Tandem coherence images
6. The 1st PC calculated from the backscattered intensity images
7. The 2nd PC calculated from the backscattered intensity images
8. Textural feature calculated from the backscattered intensity images (GLCM Uniformity)

2.1.1 Interferometric processing and geocoding: The SAR data were interferometrically processed and geocoded using a commercial software package of Gamma Remote Sensing Research and Consulting AG. The Tandem data used in the study were acquired from two separate satellite tracks. In the interferometric processing a master image was chosen for both of the tracks and all images were co-registered with their respective master images. Interferograms and coherence images were created from the already co-registered interferometric pairs. Common band filtering (also called spectral-shift filtering) was applied in order to minimize the effects of the baseline geometry on coherence estimation, and multi-looking (5 azimuth looks) was performed in order to improve on the estimates of the interferometric phase and coherence. After multi-looking, the pixel size of the image data is approximately 20 m in both azimuth and slant range (at the ERS nominal look angle of 23°). The resulting 5-look interferograms were flattened using high-quality orbit information distributed by the DLR (Deutsches Zentrum für Luft- und Raumfahrt). The interferometric coherence was estimated with a square 5x5 pixels window using Gaussian weighing of the samples. In addition to the Tandem coherence images with a temporal baseline of 24 hours, two coherence images with longer temporal baselines were computed (35 and 245 days). Also, 5-look intensity images were generated and radiometrically calibrated for range spreading loss, antenna gain, normalized reference area and the calibration constant. After interferometric processing the image data were in two different geometries determined by the respective master images. An INSAR (interferometric SAR) digital elevation model (DEM) was used first in co-registering all image data into one geometry and then in orthorectification of the image data into map coordinates. Judging visually, the image data from different tracks coincide perfectly, which implies that sub-pixel accuracy was achieved.

2.1.2 Temporal filtering: The intensity and coherence images were filtered using a multitemporal filter described in Quegan and Yu (2001). Spatial filtering causes some loss of spatial resolution, which can be avoided in multitemporal filtering, at least in principle. According to visual evaluation, the temporally filtered images display markedly diminished noise with little or no reduction in spatial resolution. In fact, narrow linear structures are more clearly visible in the filtered images than in the unfiltered ones.

2.1.3 Principal Components Transformation: Due to the large number of SAR intensity and coherence images, it was necessary to reduce the dimensions of the image datasets prior

to classification. The well-known Principal Components Transformation (PCT) was used for this purpose. Since the direction of the first PC affects the directions of all the other PCs due to the orthogonality constraint, the data mean ought to be subtracted from the input data before applying the PCT. The so-called temporal average images were subtracted from the Tandem coherence and Tandem mean image datasets before applying the PCT. Temporal averaging refers to the simple procedure of computing the average of the co-registered time series of images. If water areas are included in the intensity and coherence datasets, the first PC will highlight the water bodies. This happens on intensity image time series because differing wind conditions make large differences in backscatter from water, and on coherence image time series because free water has extremely low coherence during summer and ice has often quite a high coherence. Since we were not interested in water, but land cover, the water areas were excluded by using a water mask that was created by thresholding backscatter intensity ratio images.

2.1.4 Texture: A textural feature, Grey Level Co-occurrence Matrix (GLCM) Uniformity, was calculated from the intensity images. This feature was first calculated separately for each of the 14 Tandem mean intensity images using a 9x9 pixels window and the results were then averaged.

2.2 Reference data

A set of 1313 reference points determined at HUT was used as reference data in estimating the accuracy of classification results. Aerial photographs, topographic maps and information from the National Forest Inventory of Finland were used to determine the reference points that represent seven land-cover classes (Engdahl and Hyypä, 2002).

Possibility to distinguish different types of built-up areas from each other was studied using digital map data from Topographic Database and 1:50 000 Map Database. These vector datasets were obtained from the National Land Survey of Finland. A 9.5 km x 9.5 km subarea around Otaniemi, Espoo, corresponding approximately to one topographic map sheet, was selected as a study area for this analysis. The area is mainly suburban but also includes some densely built-up areas, forest and fields.

In the 1:50 000 Map Database, most built-up areas are presented as polygons classified as small-house areas, apartment house areas and industrial areas. Some buildings are presented as individual buildings. For the analysis, built-up areas were converted into a raster map that shows small-house areas, apartment house areas and industrial areas with different labels. The pixel size of the map is 20 m and pixels correspond to those of the image data. The information of the map database in the subarea corresponds to situation in 1996 and is thus compatible with the image data acquired in 1995-1996.

In the Topographic Database, individual buildings are presented. Buildings in the 9.5 km x 9.5 km area were also converted into raster format. Pixel size of 2 m was now selected to also include small buildings in the raster map. Each pixel of the image data thus corresponds to 100 pixels in the building map. The Topographic Database has been updated in 2000, i.e. the map data are a few years newer than the imagery, which probably caused some distortions in the analysis results. Some buildings presented in the map did not exist when the images were acquired.

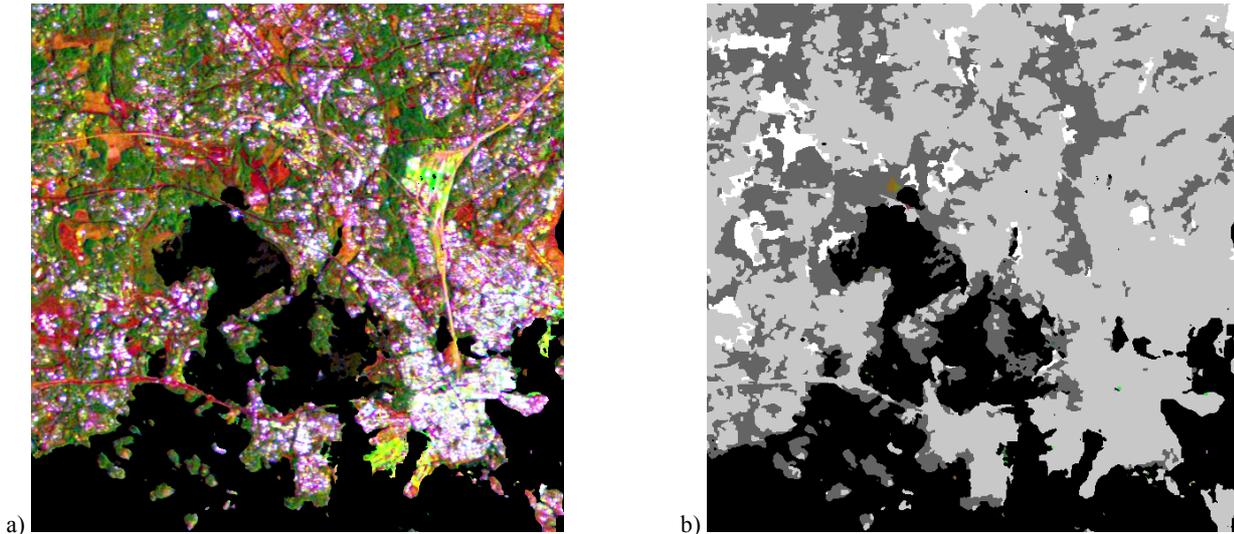


Figure 1. a) A 12 km x 12 km subarea of the imagery used in the study (Red: channel 2, Green: channel 1, Blue: channel 3). Data obtained from ESA AO3-277. b) Land-use classification result based on channels 2-6 and 8 (Light grey: built-up area, Dark grey: forest, White: open area, Black: water mask or missing data).

3. LAND-USE CLASSIFICATION

At the first stage of the study, the objective was to distinguish built-up areas from other main land-use classes. The image dataset was segmented and classified using the object-oriented image analysis software eCognition of Definiens Imaging GmbH (2002). The segmentation method (Baatz and Schäpe, 2000) is based on bottom-up region merging. Objects corresponding to different hierarchical levels of the image can be created and used in further analysis. In classification, it is possible to use the supervised Nearest Neighbour (NN) method and/or membership functions defined on the basis of user's knowledge related to the classification task.

At first, segmentation of the data into relatively small and homogeneous regions was made (segmentation level 1) using image channels 1-3. The NN classification method was then used to classify the segments into three classes: built-up area, forest and open area (classification step 1). Water areas were excluded from the analysis by using the water mask (see Section 2.1.3) as a thematic layer in segmentation and classification. Initial training areas for the three classes were defined on the basis of maps and knowledge of the region and were imported into eCognition as a training area mask. These initial training areas covered 0.12% of the total land area in the imagery and included four built-up areas in a suburban area containing both densely built-up area and small-house area, two forest areas, two open areas on agricultural fields and one on a golf course. Training areas for built-up area covered mixed pixels of buildings and other land covers in addition to pure densely built-up area. The training areas did not have any overlap with the reference points used in accuracy estimation. In eCognition sample objects were automatically created on the basis of the training area mask. Finally, the sample objects were checked manually using digital map data and some corrections were made. In classification, mean values of the segments in various image channels were used. Accuracy of results (percentage of reference points correctly classified) for four channel combinations is shown in Table 1. The channel combinations are:

- Channels 1-3 that are well suited for visual interpretation (see Figure 1).
- Channels 2-6 with and without texture channel 8; these were the most useful channels according to preliminary analysis of training area statistics.
- Channels 1-8, i.e. all channels.

Classification result obtained when using channels 2-6 and 8 is presented in Figure 1. Confusion matrix and accuracy estimates for this classification are shown in Table 2. In Table 2, the reference points have been combined into three classes corresponding to those used in classification: open area contains points of fields, forest contains points of sparse and dense forest and built-up area contains points of small-house area, apartment house area, urban area and industrial area.

Recognizing built-up areas, especially small-house areas, as built-up using only the mean intensity and coherence values of the segments is difficult. In small-house areas a remarkable part of the land is covered with vegetation and contextual information related to the neighbourhood of the segments is needed to include these vegetated areas into the land-use class built-up area. The texture channel of the image dataset provided this type of contextual information. As another alternative, use of information on the classes of neighbouring segments was tested (classification step 2). Segmentation level 2 that was identical to level 1 was created to allow another classification to be performed for the same segments. Level 2 segments were first classified according to the level 1 result and segments classified as forest or open area were further divided into urban forest, other forest, urban open and other open using neighbourhood relationships (Definiens, 2000). Membership functions and a feature 'Relative border to built-up neighbour objects' were used. Urban open and urban forest were then grouped together with the built-up area. Rural forest and agricultural areas as well as larger green areas inside towns should still become classified as forest or open area. Classification step 2 proved to be useful in most classifications performed without the texture channel. Accuracy of the final result when using channels 1-3 or 2-6 and classification step 2 is shown in Table 1.

Table 1. Accuracy of land-use classification results compared with reference points.

Reference class	Correct classif. result	% of points in reference class correctly classified					
		Channels used					
		1-3 ₁₎	1-3 ₂₎	2-6 ₁₎	2-6 ₂₎	2-6 _{8 1)}	1-8 ₁₎
Field	Open	96	92	98	96	97	94
Sparse forest	Forest	91	91	96	96	94	96
Dense forest	Forest	98	98	98	98	98	99
Small-house area (residential)	Built-up	66	79	67	74	94	93
Apartment house area (residential)	Built-up	87	97	93	98	96	96
Urban area (city centre)	Built-up	100	100	100	100	100	100
Industrial area	Built-up	93	94	97	98	99	99
All classes		91	92	94	94	97	96

1) Accuracy after classification step 1

2) Accuracy after classification step 2

Table 2. Confusion matrix and accuracy estimates for land-use classification based on channels 2-6 and 8.

Classification result	Reference class			
	Open	Forest	Built-up	Sum
Open	567	5	3	575
Forest	3	182	14	199
Built-up	12	2	524	538
Unclassified	1	0	0	1
Sum	583	189	541	1313
Interpretation accuracy	97%	96%	97%	
Object accuracy	99%	91%	97%	
Mean accuracy	98%	94%	97%	
Overall accuracy				97%

4. CLASSIFICATION OF BUILT-UP AREAS INTO SUBCLASSES

Possibility to distinguish different types of built-up areas from each other was investigated using the 9.5 km x 9.5 km subarea and raster maps described in Section 2.2. Larger segments than those used in the land-use classification were created in eCognition by using image channels 1-3. The 1:50 000 map data were used as a thematic layer in segmentation, which assured that each segment belonged completely to one class in the map and comparing properties of segments from different classes was possible. The segmentation result together with various attributes of the segments (e.g. mean values and standard deviations in each channel, textural information) were exported from eCognition, and histograms of the attribute values were formed for different types of built-up areas. It was found out that the built-up classes of the 1:50 000 map cannot be distinguished well from each other with the tested attributes.

Histograms of segment mean values in image channel 5 (the 2nd PC calculated from the coherence images) are shown in Figure 2. This was one of the best results obtained.

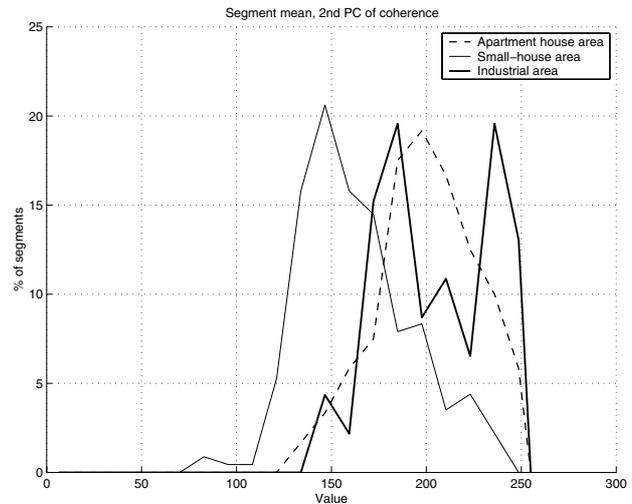


Figure 2. Histograms of segment mean values in image channel 5 in the built-up classes of the 1:50 000 topographic map.

Another segmentation was performed without map data, but otherwise with similar parameters as the previous one. For each segment, the built-up percentage, i.e. the percentage of the area covered with buildings, was then calculated using the building map derived from the Topographic database, and the correspondence between this built-up percentage and various attributes of the segments was studied. More promising results than above were now obtained. According to scatter plot figures, there was a clear dependence between the built-up percentage and intensity and coherence of the area in the image data. Figure 3 shows the relationship between segment brightness (calculated as the mean value of the segment mean value in channels 1-5) and the built-up percentage. Segments of the subarea were divided into three classes on the basis of the built-up percentage: < 10%, 10-20% and ≥ 20%. Histograms of segment brightness in these classes are shown in Figure 4.

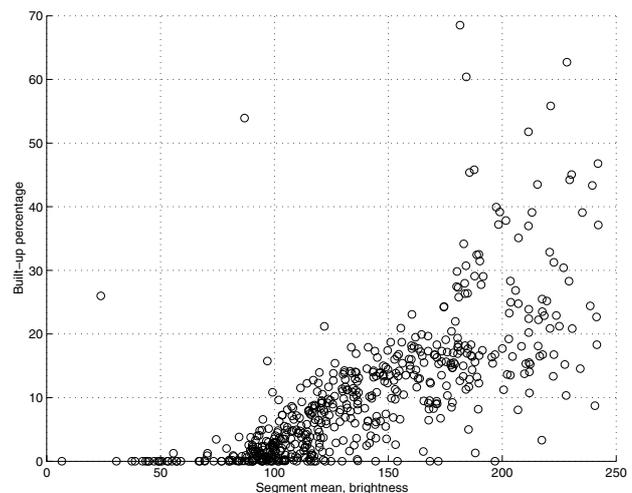


Figure 3. Scatter plot showing the relationship between segment brightness and the relative proportion of the segment covered with buildings.

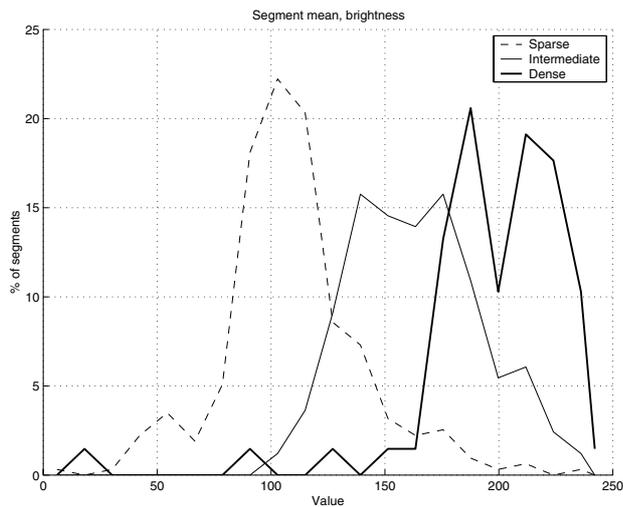


Figure 4. Histograms of segment brightness in three building-density classes.

Classification of all built-up areas in the imagery further into ‘building-density classes’ was tested using the eCognition software and threshold values determined from the histograms of Figure 4. A new segmentation level was created between the small segments used in land-use classification and large segments determined on the basis of the classification result (Section 3). The size of the segments on the new level was comparable to the size of the segments used in the analyses described above, but a somewhat different homogeneity criterion was used in segmentation. Built-up areas of the classification result based on channels 2-6 and 8 (Figure 1) were divided into three subclasses using the brightness values of the segments and the threshold values. The final result is shown in Figure 5.

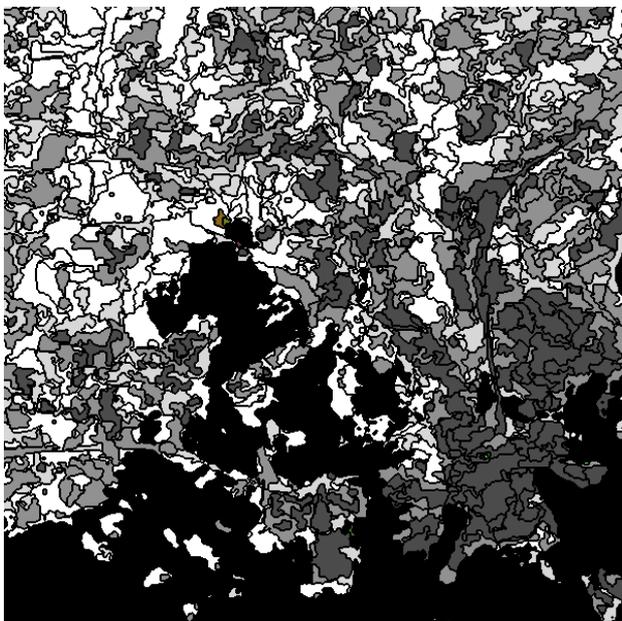


Figure 5. Classification of built-up areas into three building-density classes using brightness of segments. Built-up areas are shown in grey (light grey: sparsely built-up, dark grey: densely built-up). Forests and open areas are shown in white.

5. DISCUSSION

Results in Table 1 show that a high accuracy was achieved when classifying the dataset into three classes: built-up area, forest and open area. All channel combinations tested in the study gave an overall classification accuracy better than 90%. The best result was obtained using channels 2-6 and 8, i.e. all channels containing coherence information, the first PC calculated from the intensity images and texture. In this result, the overall accuracy compared with the reference points was 97%. Table 2 shows that the accuracy was high for each of the three classes; the mean accuracies for built-up area, forest and open area were 97%, 94% and 98%, respectively.

Reference points in urban areas were recognized as built-up with 100% accuracy in each classification (Table 1). In small-house areas and apartment house areas the percentage of reference points correctly classified as built-up was 66-94% and 87-98%, respectively. The texture channel 8 appeared to be useful in recognizing built-up areas, especially small-house areas, as built-up. The best results in small-house areas were obtained when channel 8 was used. It is worth noting that the mean value of the texture feature for each segment was used in classification, but due to the pixel-based calculation of the texture channel with a relatively large window size of 9x9 pixels, the mean value for a given segment was based on data both from the segment and its neighbourhood. Use of the texture channel thus brought some contextual information related to the neighbourhood of the segments into classification. When the texture channel was not used, the interpretation accuracy of built-up areas could be clearly improved by using information on the classes of neighbouring segments. For example, when using channels 1-3, the second classification step with the neighbourhood information improved the percentage of small-house areas and apartment-house areas recognized as built-up from 66% to 79% and from 87% to 97%, respectively. On the other hand, this classification step decreased the percentage of fields classified as open area; that is, some fields were incorrectly classified as built-up.

The classification results, especially for built-up areas, were clearly better than those obtained in some previous studies with interferometric ERS-1/2 Tandem data (e.g. Strozzini et al. 2000), although it must be noted that direct comparison of classifications with different class definitions and different number of classes is impossible. The main explanation for the high accuracy achieved in the present study was probably the good quality of the interferometric dataset. Engdahl and Hyypä (2002) achieved an overall accuracy of 90% when applying an unsupervised pixel-based classification to the same dataset. In their classification, two built-up classes (mixed urban and dense urban) were included and interpretation accuracies for these classes were 80% and 91%, respectively. If the classes were combined in the final result, the accuracy of built-up would increase near to that obtained in the present study. It can be expected, however, that the object-oriented approach applied in the present study has some benefits compared with pixel-based classification. Classification of segments is easier than classification of single pixels, especially in a built-up area with both man-made structures and vegetated surfaces. Object-oriented classification also gives homogeneous regions as a result without any postprocessing.

Despite the high accuracy achieved, visual comparison of the results with digital map data revealed that some built-up areas were missing in the results. These areas are not clearly visible in

the imagery, probably mainly due to a dense forest cover. On the other hand, small areas in forest, especially rocky or hilly areas, were sometimes misclassified as built-up. Some preliminary numerical comparisons between the classification results and the 1:50 000 map data were also made. A raster map showing fields, meadows and open areas in addition to the different classes of built-up areas was created for the 9.5 km x 9.5 km subarea (see Section 2.2) and used in the comparisons. The percentages of small-house areas, apartment house areas and industrial areas correctly classified as built-up when using channels 2-6 and 8 were 78%, 96% and 99%, respectively. The percentages of fields, meadows and open areas classified as open were 93%, 54% and 38%, respectively. Boundary pixels of the classes in the map were excluded from the analysis. The accuracies seem to be somewhat lower than those obtained using the reference points. Further investigations on the subject are needed, but it can be expected that problems arise e.g. from generalization of information in the map and definition of classes such as open area that may include different types of areas. Some areas that appear similar in the imagery can also belong to different classes in the map; for example, one grass-covered area can be a park classified as open area in the map but another similar area can be a garden assigned to small-house area.

When studying the possibility to classify built-up areas into subclasses, the classes of the 1:50 000 maps appeared to be difficult to distinguish from each other as shown in Figure 2. On the other hand, results presented in Figure 3 suggest that intensity and coherence values of an area in the imagery are clearly related to the building density of the area and could thus be used to recognize areas with different building densities. According to visual evaluation, the classification result obtained when using this information was promising (Figure 5). Accuracy of the result, however, has not yet been tested.

6. CONCLUSIONS

Good results were obtained when multitemporal interferometric ERS data were used to distinguish built-up areas from forest and open areas using an object-oriented classification approach. The dataset also appeared to be promising for classifying built-up areas into subclasses according to building density. Further development of methods for automatic land-use mapping and map updating thus seems to be possible. It can be expected that such methods have high potential in the future, especially when interferometric SAR data with high spatial resolution becomes available.

ACKNOWLEDGEMENTS

The authors wish to thank ESA AO3-277 for providing the SAR data, the National Land Survey of Finland for providing the map data and Matine, Academy of Finland and the Finnish Geodetic Institute for financial support.

REFERENCES

- Baatz, M., Schäpe, A., 2000. Multiresolution Segmentation – an optimization approach for high quality multi-scale image segmentation. In: Strobl, J. et al. (eds.): *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg 2000*. Herbert Wichmann Verlag, Karlsruhe, pp. 12-23.
- Dammert, P. B. G., Askne, J. I. H., Köhlmann, S., 1999. Unsupervised segmentation of multitemporal interferometric SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5), pp. 2259-2271.
- Definiens, 2000. eCognition, Object Oriented Image Analysis. User Guide, Version 1.0, Delphi2 Creative Technologies GmbH – DEFiNiENS, München, Germany.
- Definiens Imaging GmbH, 2002. <http://www.definiens-imaging.com/>
- Engdahl, M. E., Hyypä, J., 2002. Land-cover classification using multitemporal ERS-1/2 INSAR data. Manuscript.
- Quegan, S., Yu, J. J., 2001. Filtering of multichannel SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11), pp. 2373-2379.
- Strozzi, T., Dammert, P. B. G., Wegmüller, U., Martínez, J.-M., Askne, J. I. H., Beaudoin, A., Hallikainen, M. T., 2000. Landuse mapping with ERS SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2), pp. 766-775.
- Wegmüller, U., Werner, C., 1997. Retrieval of vegetation parameters with SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 35(1), pp. 18-24.