Combined Airborne Sensors in Urban Environment

Alwin Dimmeler[(1)], Hendrik Schilling[(1)], Michal Shimoni[(2)], Dimitri Bulatov[(1)], Wolfgang Middelmann[(1)]

[(1)] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Gutleuthausstrasse 1, 76275 Ettlingen, Germany
[(2)] Signal and Image Centre, Royal Military Academy (CISS-RMA), Av. de la Renaissance 30, Brussels, Belgium

ABSTRACT

Military operations in urban areas became more relevant in the past decades. Detailed situation awareness in these complex environments is crucial for successful operations. Within the EDA (European Defence Agency) project on "Detection in Urban scenario using Combined Airborne imaging Sensors" (DUCAS) an extensive data set of hyperspectral and high spatial resolution data as well as three dimensional (3D) laser data was generated in a common field trial in the city of Zeebrugge, Belgium, in the year 2011.

In the frame of DUCAS, methods were developed at two levels of processing. In the first level, single sensor data were used for land cover mapping and the detection of targets of interest (i.e. personnel, vehicles and objects). In the second level, data fusion was applied at pixel level as well as information level to investigate the benefits of combining sensor systems in an operational context.

Providing data for mission planning and mapping is an important task for aerial reconnaissance and it includes the creation or the update of high quality 2D and 3D maps. In DUCAS, semi-automatic methods and a wide range of sensor data (hyperspectral, LIDAR, high resolution orthophotos and video data) were used for the creation of highly detailed land cover maps as well as urban terrain models. Combining the diverse information gained by different sensors increases the information content and the quality of the extracted information.

In this paper we will present advanced methods for the creation of 2D/3D maps, show results and the benefit of fusing multi-sensor data.

Keywords: hyperspectral, LIDAR, spatial resolution, 3D, spectral matching, classification, data fusion, terrain model, land cover

1. INTRODUCTION

Military operations in urban areas became more relevant in the past decades. Detailed situation awareness in these complex environments is crucial for successful operations. The objective of the EDA project DUCAS is to investigate the potential benefit of combined high spatial and spectral resolution airborne imagery for defense applications in urban areas. Seven countries are participating in this project: Sweden, Belgium, The Netherlands, Italy, France, Norway and Germany. DUCAS started in 2010 and will end in 2013. For the scope of DUCAS an extensive data set of hyperspectral and high spatial resolution data was generated in a common field trial in Zeebrugge, Belgium [1, 2], in the year 2011. Additionally, 3D-laser data were acquired earlier to the field trial. The attempt of that trial was to realize as much as possible realistic military scenarios like the support for conventional or special forces comprising tasks like detection, tracking and retrieval of objects.

In the frame of DUCAS methods were developed at two levels of processing. In the first level, single sensor data was used for land cover mapping and the detection of targets of interest (i.e. persons, vehicles and objects). The processing covers classification, anomaly detection, change detection and spectral matching [3, 4, 5]. The processing used hyperspectral data from the visible light to the short wave infrared data and in the long wave infrared, broad band sensor data in the visible, near infrared and mid wave infrared and airborne 3D-laser scanner as well. Due to the heterogeneous urban environment with many types of backgrounds and mixed objects the tasks are difficult to be resolved by a single

Electro-Optical Remote Sensing, Photonic Technologies, and Applications VII; and Military Applications in Hyperspectral Imaging and High Spatial Resolution Sensing, edited by G. W. Kamerman, O. K. Steinvall, G. J. Bishop, J. D. Gonglewski, Proc. of SPIE Vol. 8897, 88970U · © 2013 SPIE · CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2028648

sensor and often processing yields unsatisfying results. Therefore, in the second processing level, data fusion was applied at pixel level and at information level as well, and the benefits of combining sensor systems in an operational context were investigated.

Providing data for mission planning and mapping is an important task for aerial reconnaissance and it includes the creation or the update of high quality 2D and 3D maps. In DUCAS, semi-automatic methods and a wide range of sensor data (hyperspectral, LIDAR, high resolution orthophotos and video data) had been used for the creation of highly detailed land cover maps as well as urban terrain models. Combining the diverse information gained by different sensors increases the information content and the quality of the extracted information.

2. SENSOR DATA

The main data source used for the data analysis described in this paper is the hyperspectral "ASIA Eagle II" of the company SPECIM. The "ASIA Eagle II" operates in the spectral regions from the visible (VIS) to the near infrared (NIR), 400 nm to 970 nm. The hyperspectral data cube consists of 130 bands with a spectral width of 4.5 nm. At DUCAS the spatial resolution was 40x40 cm². All data are radiometrically (µW/(cm² nm sr) corrected and georeferenced. To reach the required accuracy of georeferencing in the order of one pixel, an interactive optimization algorithm was used to improve the inertial navigation data. An example of the data acquired on 27 June 2011, at UTC 11:17 under clear sky and sunny conditions is given in Fig. 1.

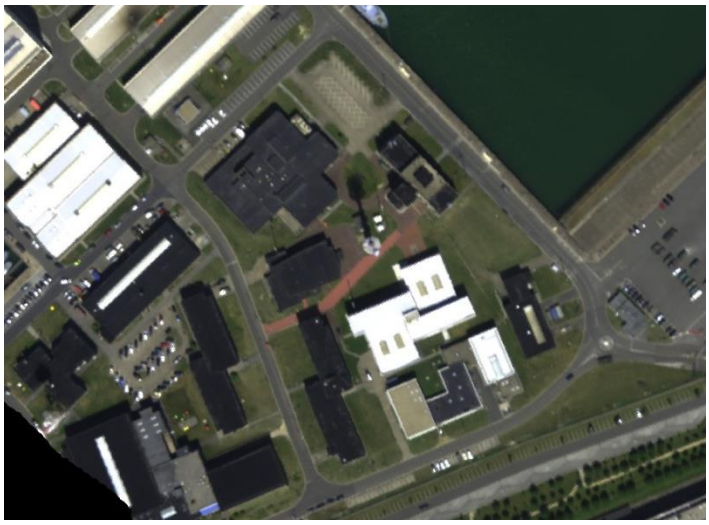

Figure 1: AISA Eagle II image 27 June 2011, UTC 11:17, red=637 nm, green=549 nm, blue=459 nm.

The second data set was taken by a scanning 3D laser, the Light Detection And Ranging (LIDAR) "ALTM Gemini" from the company Optech. However, the data acquisition by TopScan GmbH took place three month prior to the DUCAS trial, on 27 March 2011, to avoid tree canopies. The point density is approximately 100 points/m². Range and the respective height resolution is about 5 cm. These LIDAR data were mainly used for building extraction and terrain modeling, see chapter 3.2 and 4 respectively.

For the creation of the highly detailed land cover maps and urban terrain models two other data sources were used additionally, an orthophoto created by RMA from overlapping high resolution images and video sequences acquired by the Netherland Aerospace Laboratory (NLR) on the 27 June 2011, UTC 15:50. In chapter 4 a test case is described in which the data were fused for the classification and the reconstruction of buildings as well as texturing the horizontal and vertical structures with these data.

3. CLASSIFICATION AND MAPPING

This chapter describes the processing done for classification and mapping. Chapter 3.1 describes the classification results for the hyperspectral images and chapter 3.2 the pixel based fusion of the classification results with the buildings extracted from the 3D laser data which were used to update a map in chapter 3.3.

3.1. Classification using Hyperspectral Imagery

The aim of the spectral classification is to find the best spectrum match in order to assign a label to a pixel. This spectrum can either be taken from a spectral library or directly from a region of interest (ROI) in the data itself. Several pixel-based supervised methods like Support Vector Machine (SVM) [6], Minimum distance [7], Mahalanobis distance [7], Maximum Likelihood [7] and Spectral Angle Mapper (SAM) [8] were applied. Supervisely, for each class regions of interest were defined.

The Minimum Distance classification calculates the Euclidian distance between the mean spectra of the class. The Spectral Angle Mapper calculates the angle between the mean spectra and the Maximum Likelihood and Mahalanobis distance methods use class statistics for the classification. For each method, a threshold for maximum distance can be applied to ignore outliers. For the SVM, the selected training samples are used to define the decision boundaries between the classes. These boundaries are calculated such that the margins between the classes are maximized and the data points closest to the boundaries are named support vectors. If the classes cannot be separated linearly, it is also possible to use nonlinear kernel functions to define more complex boundaries. A pixel is assigned to a class depending on its position with respect to the boundaries.

The results of the different methods were compared to a validation map, which was generated by an operator using a map and very high resolution color images of 5 cm and 8 cm resolution as well as the Digital Elevation Model (DEM). Fig. 2 shows the validation map (left) and the classification result obtained using SVM (right). The overall accuracy was calculated as the ratio of correctly classified pixels to total number of pixels. Table 1 contains these accuracy results of the different methods. As one can see, the pixel based supervised classification by SVM performs best. This result was expected for a complex (i.e. urban areas) and very high spectral data.

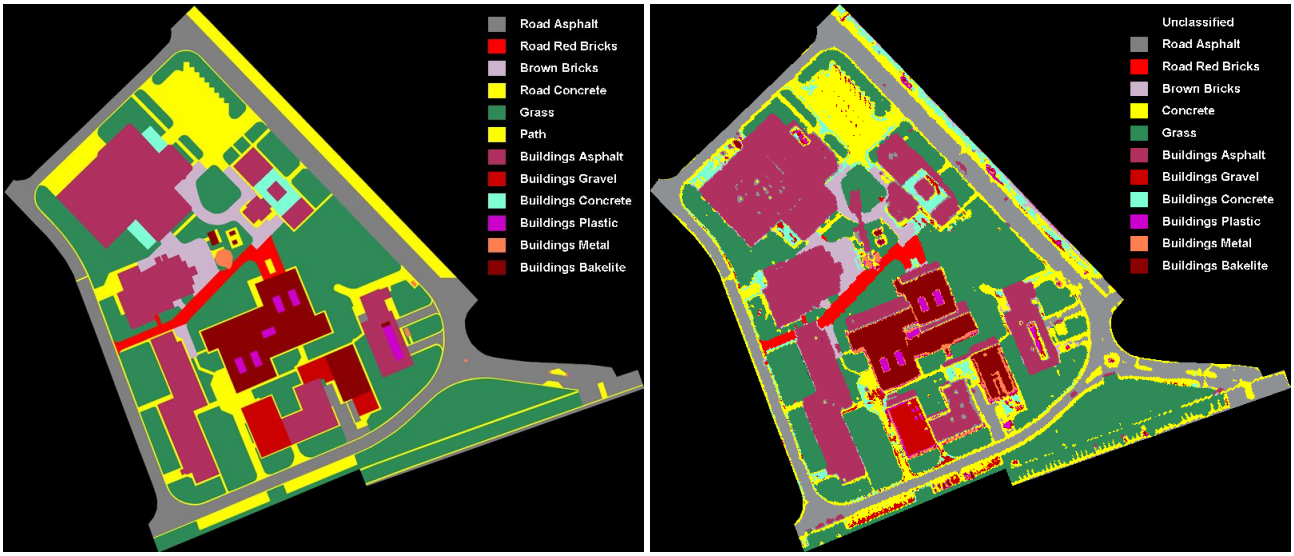

Figure 2: Validation map (left) and SVM classification result before post-processing (right).

Table 1: Overall accuracy for different classification algorithms applied to Eagle data.

Minimum Distance	67.4
Mahalanobis Distance	69.9
Maximum Likelihood	56.8
Spectral Angle Mapper	72.9
Support Vector Machine	78.8

For further improvement, majority analysis can be used to reduce the high segmentation in the classification result (suppress single pixel in homogenous areas). This post-processing improved the overall accuracy by approximately 10%

for the "Mahalanobis Distance" classifier and about 5% for the SVM. Excluding the SVM and SAM, the obtained accuracy by the classifiers was not as high as expected (i.e. over 85% accuracy). One reason for this unsatisfying result is the small number of pixels that were selected for the training regions in relation to the number of spectral bands. For "Maximum Likelihood" classifier, there should be around 10 times the number of bands of training pixels for separating correctly the segments. Due to the rare appearance of some classes (e.g. building metal) this was not possible with the available data sets. Moreover, it is obvious that some errors were originated from objects which are not present in the validation set (as vehicles and sky-roof windows and many different small objects available in the scene). Other classification confusions and errors were originated from shadows and misclassification at the edges of objects.

Fig. 3 presents the error image for the SVM classifier before post-processing. White pixels indicate the misclassified ones. As can be observed, most errors occur in shadow areas and at the edges of objects. The ones in shadow areas are due to low contrast and nonlinear behavior of the spectra. Errors at the contour of objects are due to mixed pixels and misalignment with the validation map.

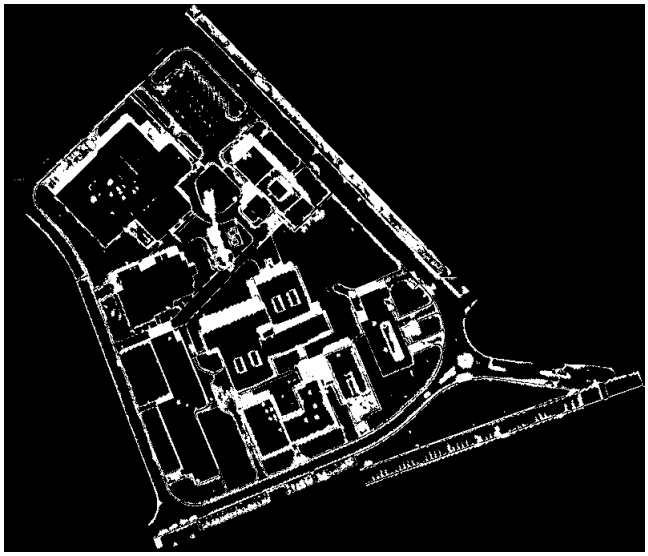

Figure 3: Error image of SVM classification before post-processing.

The results of the SVM classification obtained for the AOI were encouraging but they also highlighted the presence of noise in the scene. To reduce the variability in the hyperspectral data and the noise, Minimum Noise Fraction (MNF) [9] transformation was performed earlier to the SVM. The resulting of the overall accuracy shows an improvement of 2.6% to 81.4%.

3.2. Fusion of Classification Results with LIDAR Data

To further improve the classification results for the AOI, georeferenced 3D laser point cloud data which were interpreted as DEM were fused with the obtained results. The fusion of hyperspectral and DEM data allows in general reducing of ambiguities between ground and elevated features which are covered by similar materials (i.e. road and roof asphalt). Fusing DEM and hyperspectral data was done in two ways. The first one was a pixel-based fusion by integrating height as additional band to the hyperspectral data; the second one was an object-based fusion. However, only small improvement in the accuracy was achieved in both cases (up to about 2%, see Table 2). In complex urban areas the geometrical and materials separation in the scene usually is high; however, the investigated scene is relatively simple. In the AOI there are large buildings with large spectral distances in between and different roof and background materials (i.e. the asphalt of roof and pavement are different). Hence, the improvement by using DEM is rather marginal.

Nevertheless, if the origin of the errors in the hyperspectral image is the presence of shadows (see also Fig. 1) a significant increase in the classification accuracy is gained by including a shadow mask in the evaluation process. The shadow mask was calculated using the DEM and the sun position [10]. As Table 2 shows, the overall accuracy is 89.5% (improvement of 8%) with supplementary fusion of height and spectral information as well as using shadow masks to exclude the areas from classification (labelled as unclassified class).

Table 2: Overall accuracy for SVM and different levels of data fusion.

SVM with MNF	81.4
SVM with MNF & height as 1 band	83.0
SVM with MNF and DEM	83.5
SVM with MNF, DEM and shadow mask	89.5

The images in Fig. 4 illustrate this improvement. The left image presents the SVM classification map after MNF. The right image shows the result after object-based fusion and applying the shadow mask. It can clearly be seen that the classification areas are much more homogenous, the edges and corners of the buildings are sharpened and the shadows are marked as a separate, undefined class.

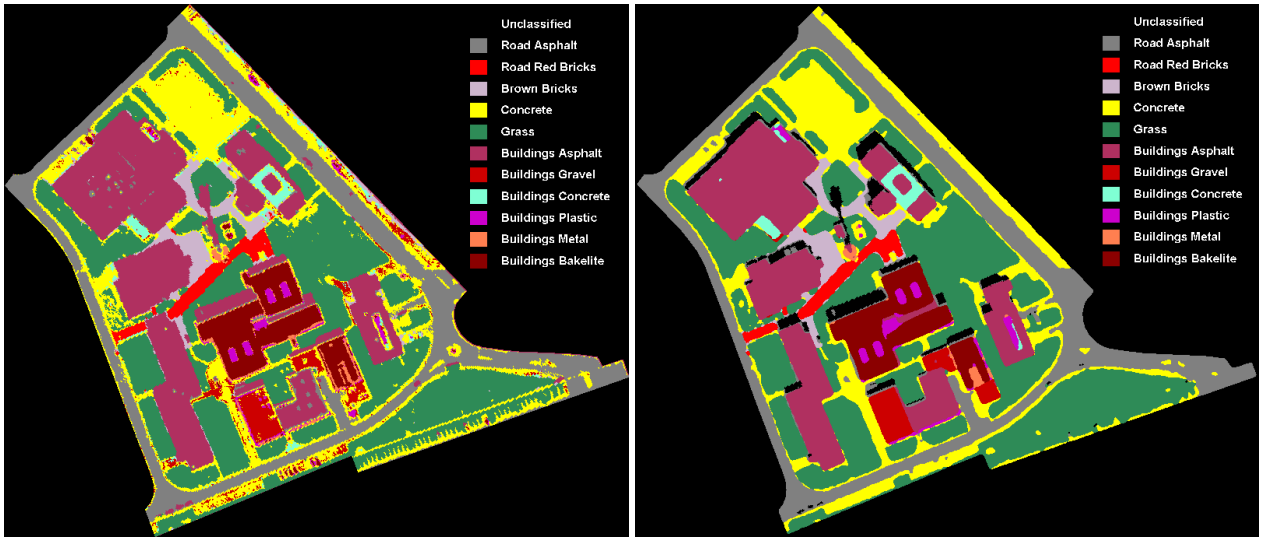

Figure 4: SVM Classification after MNF without fusion (left) and after data fusion with DEM and shadow mask (right).

Fig. 5, left image, presents the SVM error map before fusion and after object-based fusion and applying the shadow mask (right image). Nearly all the remaining errors are present at the contour of objects; errors are caused by mixed pixel and the evaluation process.

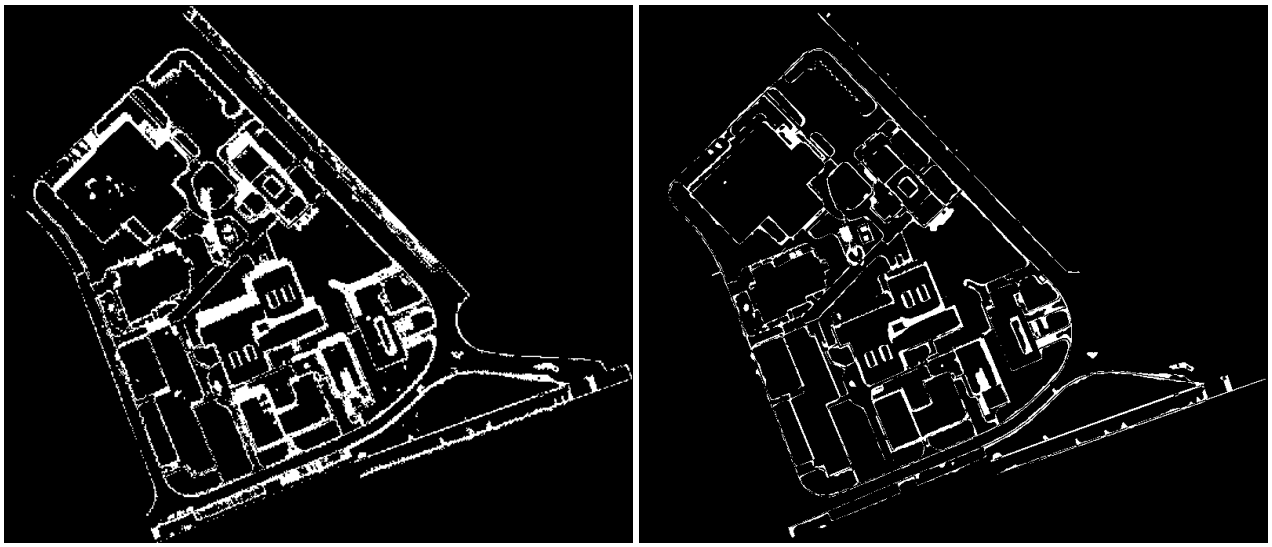

Figure 5: Error images after MNF without fusion (left) and after data fusion with DEM and shadow mask applied (right).

3.3. Update of Map

Having the classified surface, these results can be merged with an existing map for update. In Fig. 6, left, the original map of the AOI is shown. In the merged map with classification results, right, much more details are present. It can clearly be seen that additional buildings were constructed at the bottom half of the image. Details include roads and parking areas (red), paths (yellow) and additional information on grass cover (green) and gravel areas (skin color). Additionally, there is also a particular construction (purple) detected, which is a small, low building. However, the tower (grey circle in the center of left image) is labeled as road and hence is not present any more.

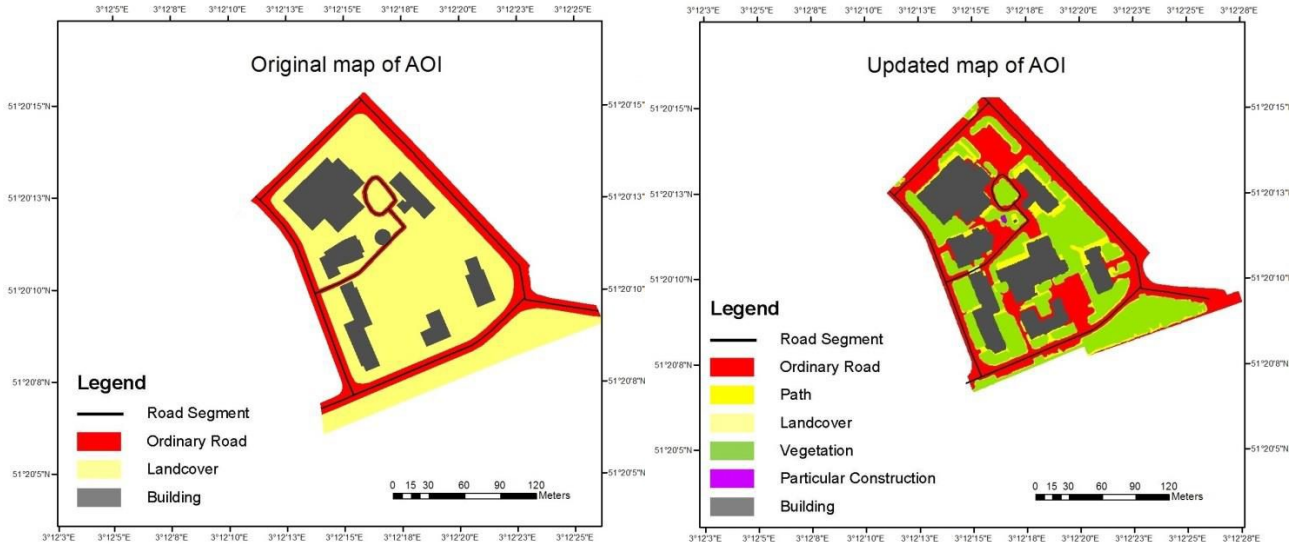

Figure 6: The original map (left) and the updated map (right).

4. TERRAIN MODELING

A very detailed 3D terrain models is very helpful for training purposes and crucial for mission planning in complex urban areas. In urban scenarios, buildings are of particular interest as they can be used for ambushes, offer cover for snipers and impede the line of sight. A good knowledge of the building shapes and positions is also important for trafficability analysis.

Detection of buildings and vegetation was done in three steps: computation of Digital Terrain Model (DTM), separation of buildings and vegetation as well as filtering and labeling [11]. The difference between DEM and DTM gives elevated points, which are either buildings or vegetation, like trees or bushes. In a second step it would be possible to filter out the vegetation by analyzing the multiple return signal of the laser pulse caused by vegetation. However, in the AOI the presence of vegetation is negligible. In the third step the detection of buildings is made by filtering elevated regions by their altitude, area and elongation. This is to suppress false pixels. The remaining pixels are the buildings hypothesis. The reconstruction of the buildings is applied to these pixels for the creation of an accurate 3D model.

The reconstruction of buildings is again done in three steps: building boundary extraction, roof detail analysis and texturing. First the building contour as a rectangular polygon (optionally) is computed. Straight lines in the DEM are obtained by an edge detector [12]. Their slopes are stored in an histogram modulo 90° for determining the main building orientation. Afterwards the height is thresholded to get the lowest boundary rectangle: The contours of the building are refined by increasing the threshold yielding additional rectangular subparts. Neglecting very small and very narrow parts is achieved by reducing the number of vertices.

The second step of the building reconstruction is the roof detail analysis. Several dominant planes from the 3D point clouds within every building are computed by a modified random sample consensus (RANSAC-) procedure [11]. After calculating a course set of planes and applying some morphological operations, roof polygons are built and intersected between each other and the outlines of the buildings. However, with this procedure small structures like chimneys and

dormers are smoothed out. These structures could be identified in advance by determining small elevated, connected regions of low gradient and by ignoring the corresponding points in the later computations. Walls are modeled by vertical trapezoids connecting roof polygons and corresponding points on the ground.

Although this 3D model already contains a lot of information needed for mission planning, the information gain can be increased by considering additional sensor data. Wall textures obtained from up-to-date sensor data contrary to the standard wall textures gives hints for access points to buildings, shows windows and provides hazard areas which should be avoided during a possible attack. For texturing the vertical polygons (walls), the conventional None-NADIR view NLR video was the available complementary data. For the texturing procedure, the exact position and the view direction of the camera for each image is needed. Due to lack of accuracy of the inertial navigation data, an interactive method was used. Having sufficient corresponding points between the 3D model and camera image, the camera orientation can be estimated fully automatically. For this at least six corresponding points are needed for a non-calibrated camera. We took around ten corresponding points to improve the accuracy of the process. Now, each polygon in the 3D model can be back-projected in the image plane thus obtaining a synthetic depth map. This depth map allows determining the foreground polygon for every pixel. Using a cost function, the best image for every wall is chosen for texturing. If the wall is not covered completely by one image, the second best image is used, and so on. Fig. 7 illustrates the occlusion analysis procedure.

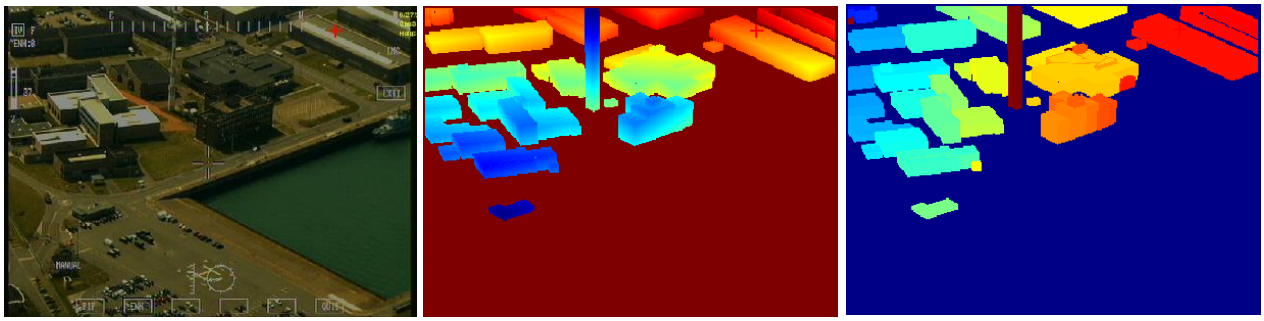

Figure 7: Illustration of the occlusion analysis procedure. The left image shows a video frame, the image in the middle shows the corresponding synthetic depth map of the buildings and the right image shows the foreground polygon.

Fig. 8 shows the results of the wall texturing for one building. In this 3D image the orthophoto was used for horizontal areas on ground and roofs instead of the classification results. In such an image important details like windows or entrance points are visible. However, due to the poor quality of the video data, these details are not very clear in Fig. 8.

Figure 8: 3D-terrain model of area with texture from orthophoto for pixel on horizontal areas and from slant view video for pixel of vertical areas (walls).

Information about the materials covering the surface is essential for trafficability analyses or gives hints which roofs can be accessed by enemy forces. Given a good spatial registration, the classification results can be used for dressing the horizontal areas as roofs and the ground. This result contains information about materials as well as object types. Fig. 9 shows a sample result for this work. At the top of the image there are also buildings outside the AOI, where classification data are missing.

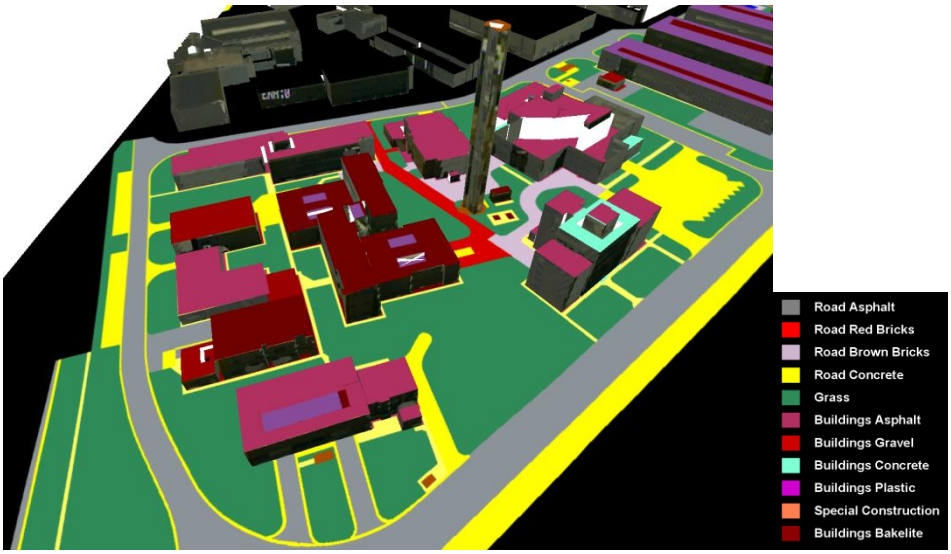

Figure 9: 3D model with horizontal planes (ground and roofs) textured by manually improved classification results.

5. DISCUSSION AND CONCLUSIONS

For military operations in urban areas, detailed situation awareness is crucial for the success of the operations in these complex environments. Therefore, providing data for mission planning and mapping is an important task for aerial reconnaissance. This also includes the creation or update of high quality 2D and 3D maps (which is also one of the tasks of the EDA project DUCAS).

Advanced (semi-) automatic classification algorithm in a combination with hyperspectral data can help operators to create land cover maps of large area in a raid procedure. Several pixel-based supervised classification methods were applied on hyperspectral data for the creation of 2D land cover maps. The best results were obtained with the Support Vector Machine classifier, which is a powerful and robust method. However, the accuracy for radiance data was not exceeded 79% (lower than expected). The main reasons for this unsatisfying result are the small number of pixels selected for the training regions in relation to the number of spectral bands, and the presence of shadows. The classification results were improved slightly for MNF transformed data with reduced variability. Results were also improved with additional data of a DEM, because the combined data allowed the separations between geometrical objects and materials. However, the research in this study shows that in urban scenes where the separation between man-made objects is clear, fusion of optical data and DEM are not mandatory. A significant increase in the classification accuracy is gained by including a shadow mask in the evaluation process. With supplementary fusion of height and spectral information as well as using shadow masks the overall accuracy was improved to 90% resulting in a highly detailed 2D land cover map. The remaining errors are at the contour of objects, which are due to mixed pixels and misalignment between validation map and the data used for classification.

Beside 2D maps, highly detailed 3D terrain models are very helpful for training purposes and crucial for mission planning in complex urban areas. A better support can be achieved by integrating the computed model into combat simulation software [11]. During operations 3D models also build the basis for target handoff between different video systems. For urban areas, buildings are of particular interest. To be able to create such a 3D model, the detection of buildings (and vegetation, if present) is necessary. Having the buildings reconstructed, these 3D models were enriched in a data fusion process with additional information gained by other sensors like texture of visible images and/or classification results created from hyperspectral sensors. The later allowed automatic labelling of objects with basic classes like road and vegetation.

Overall, a wide range of sensor data (hyperspectral, LIDAR, high resolution orthophotos and video data) had been used for the creation of highly detailed land cover maps as well as urban terrain models. Combining the diverse information gained by the different sensors increases the information content and the quality of the extracted information. However, it must be mentioned that scenes with many different classes are not qualified for automatic classification due to low training pixels and high operational effort.

REFERENCES

[1] Renhorn, I., Benoist, K., Bourghys, D., Boucher, Y., Briottet, X., de Ceglie, S., Dekker, R., Dimmeler, A., Dost, R., Friman, O., Kåsen, I., Maerker, J., van Persie, M., Resta, S., Schwering, P., Shimoni, M. and Haavardsholm, T., "Detection in urban scenario using combined airborne imaging sensors", Proc. SPIE 8353, Infrared Technology and Applications XXXVIII, 83530I (May 1, 2012); doi: 10.1117/12.921473

[2] Renhorn, I., Axelsson, M., Benoist, K., Bourghys, D., Boucher, Y., Briottet, X., de Ceglie, S., Dekker, R., Dimmeler, A., Friman, O., Kåsen, I., Maerker, J., van Persie, M., Resta, S., Schwering, P., Shimoni, M. and Haavardsholm, T., "Detection in urban scenario using combined airborne imaging sensors", 5th International Symposium on Optronics in Defence and Security (OPTRO)", 1-6 (2012).

[3] Dekker, R.J., Schwering, P.B., Benoist, K.W., Pignatti, S., Santini, F. and Friman, O., "LWIR Hyperspectral Change Detection for Target Acquisition and Situation Awareness in Urban Areas", Proc. SPIE 8743, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX, 874306 (May 18, 2013); doi: 10.1117/12.2015761

[4] Renhorn, I., Achard, V., Axelsson, M., Benoist, K., Borghys, D., Briottet, X., Dekker, R., Dimmeler, A., Friman, O., Kåsen, I., Matteoli, S., Lo Moro, M., Opsahl, T., van Persie, M., Resta, S., Schilling, H., Schwering, P., Shimoni, M., Haavardsholm, T. and Viallefont, F.: „Hyperspectral Reconnaissance in Urban Environment", Proc. SPIE 8704, Infrared Technology and Applications XXXIX, 87040L (June 11, 2013); doi: 10.1117/12.2019348

[5] Borghys, D., Idrissa, M., Shimoni, M, Friman, O., Axelsson, M, Lundberg, M. and Pernee, C., "Fusion of multispectral and stereo information for unsupervised target detection in VHR airborne data", Proc. SPIE 8745, Signal Processing, Sensor Fusion, and Target Recognition XXII, 874514 (May 23, 2013); doi: 10.1117/12.2015968

[6] Chang, C.-C., Lin, C.-J., "LIBSVM: A Library for Support Vector Machines" (2001).

[7] Richards, J.A., [Remote Sensing Digital Image Analysis], Springer-Verlag, Berlin, 240 (1999).

[8] Kruse, F. A., Lefkoff, A. B., Boardman, J. B., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., Goetz, A. F. H., "The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging spectrometer Data", Remote Sensing of the Environment, v. 44, 145-163 (1993).

[9] Green, A. A., Berman, M., Switzer, P., Craig, M. D., "A transformation for ordering multispectral data in terms of image quality with implications for noise removal", IEEE Transactions on Geoscience and Remote Sensing, v. 26, no. 1, 65-74 (1988).

[10] Borghys, D, Kasen, I, Achard, V, Perneel, C, "Comparative evaluation of hyperspectral anomaly detectors in different types of background", Proc. SPIE 8390, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, 83902J (May 8, 2012); doi:10.1117/12.920387

[11] Bulatov, D., Solbrig, P. and Wernerus, P., "Ad-hoc model acquisition for combat simulation in urban terrain", Proc. SPIE 8538, Earth Resources and Environmental Remote Sensing/GIS Applications III, 85380G (October 25, 2012); doi:10.1117/12.974486

[12] Burns, J.B., Hanson, A.R. and Riseman, E.M., "Extracting straight lines," Transactions on Pattern Analysis and Machine Intelligence, 8(4), 425-455 (1986).