

DATA FUSION FOR CLASSIFICATION AND OBJECT EXTRACTION

B. C. Gruber-Geymayer, A. Klaus, K. Karner

VRVis Research Center for Virtual Reality and Visualization, Graz, Austria – (gruber, klaus, karner)@vrvis.at

KEY WORDS: Classification, Extraction, Fusion, Orthoimage, DEM/DTM

ABSTRACT:

This paper describes the fusion of information extracted from multispectral digital aerial images for land use classification. The proposed approach integrates spectral classification techniques and spatial information. The multispectral digital aerial images consist of a high resolution panchromatic channel as well as lower resolution RGB and NIR channels and form the basis for information extraction. Our land use classification is a 2-step approach that uses RGB and NIR images for an initial classification and the panchromatic images as well as a digital surface model for a refined classification. The digital surface model is generated from the high resolution panchromatic images of a specific photo mission. Based on the aerial triangulation using area and feature based points of interest we are able to generate a dense digital surface model by a dense matching procedure. This approach produces the desired land use classification results and exploits the high redundancy of the source data set to automatically perform after a short interactive training phase.

1. INTRODUCTION

Digital aerial cameras can be used to produce images with high degree of image overlap in flight direction at almost no additional costs. A terrain point may be visible in 5 to 15 or even more images. Currently, aerial photogrammetry is undergoing a “paradigm shift” (Leberl, 2004) which means the transition from minimizing the number of film photos due to human operator intensive processing to maximizing the robustness of automation due to high redundant image information using new large format digital aerial cameras.

This contribution is based on images from the UltraCamD camera from Vexcel Imaging with its multispectral capability. UltraCamD offers simultaneously sensing of high resolution panchromatic information and additional multispectral - thus red, green, blue and near infrared (NIR) - information. The benefit of the multispectral sensing is obviously the simultaneous recording of all bands, i.e. any classification can be performed without cumbersome registration of different scenes. Additional ideas about photogrammetric color sensing may be found in (Leberl, 2002). The image data used comprise the panchromatic high resolution images as well as the low resolution multispectral images, see Fig. 1.

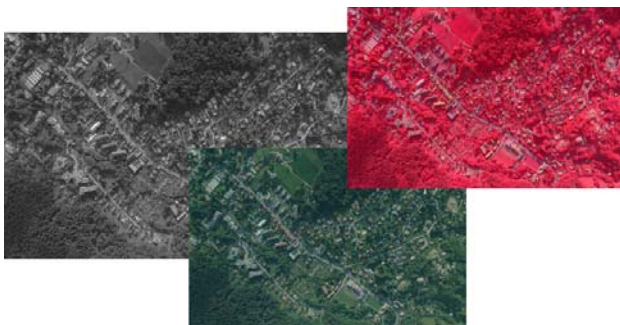


Fig. 1: Image data used
 (left) panchromatic high resolution image
 (middle) RGB low resolution image
 (right) NIR-R-G low resolution image

The proposed workflow includes the following steps:

- initial classification of all images, see section 2,
- the aerial triangulation (AT), see section 3,
- dense matching to generate a dense digital surface model (DSM), see section 4,
- ortho image production, see section 5,
- refined classification using the DSM and the ortho images, see section 6.

This paper will focus on the information fusion in classification.

2. INITIAL CLASSIFICATION

The initial classification is a supervised classification performed on each of the overlapping color images with 4 color channels RGB and NIR. The classified classes mainly rely on color and infrared and will be refined later using the information from the DSM.

The following types of classifiers for supervised classification can be found in literature:

- maximum likelihood classifier
- neural network classifier
- decision tree classifier
- support vector machine

Many - especially newer - papers propose support vector machines for use in multispectral data. The purpose of (Huang, 2002) is to demonstrate the applicability of support vector machines to deriving land use from operational sensor systems and to evaluate systematically their performances in comparison to other popular classifiers.

The support vector machine (SVM) represents a group of theoretically superior machine learning algorithms, see (Vapnik, 1995). The SVM employs optimization algorithms to locate the optimal boundaries between classes. Statistically, the optimal boundaries should be generalized to unseen samples with least errors among all possible boundaries separating the classes, therefore minimizing the confusion between classes.

An important benefit of the support vector machine approach is that the complexity of the resulting classifier is characterized by the number of support vectors rather than the dimensionality of the transformed space. As a result, SVMs tend to be less prone to the problem of over-fitting than some other methods. Initial classification uses the SVM library LIBSVM developed at the National Taiwan University, see (Chang, 2005) for software details and (Hsu, 2003) for a practical guide to support vector classification.

Initial classification discriminates all classes that are more significantly described by color and NIR values than by texture and spatial relationship. These classes are:

- Solid: man made structures like streets, buildings with gray or non-colored roofs
- Colored roofs
- Soil, bare earth
- Lake, river, sea
- Vegetation: wood, grassland, fields
- Dark shadows
- Swimming pools
- Snow/ice

The first step in classification is feature extraction, i.e. the process of generating spectral feature vectors from the 4 input planes. The selection of the features to be extracted is important because it determines the amount of features that have to be computed and processed. In addition to the improved computational speed in lower dimensional feature spaces there might also be an increase in the accuracy of the classification algorithm. The features computed for initial classification include

- Single pixel values of all input planes
- Normalized ratio between image planes
Ratio images may be used to remove the influence of light and shadow on a ridge due to the sun angle. It is also possible to calculate certain indices which can enhance vegetation or geology. NDVI - Normalized Difference Vegetation Index - is a commonly used vegetation index which uses the red and infrared bands of the spectrum.
- Values computed in a circular neighborhood of given radius like minimum, maximum or standard deviation

Fig. 2, illustrates a subset of feature values for some classified image regions. The colors represent the following features:

- R,G, and B colored red, green and blue
- NIR is colored violet
- NDVI is colored in yellow

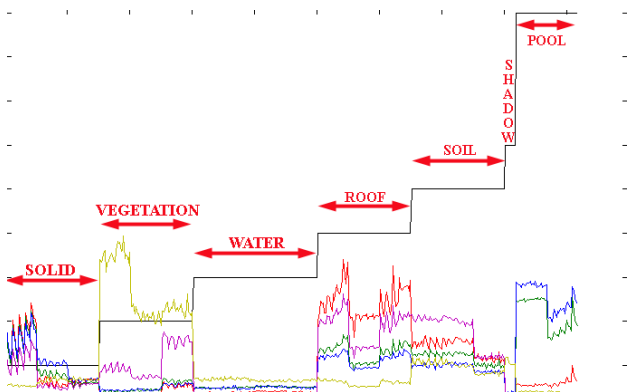


Fig. 2: A subset of feature values used in initial classification

Additional features computed are not included as they would reduce the clearness of Fig. 2. The distribution of features can be depicted as shown in Fig. 3 for each two features. The SVM is trained to find optimal boundaries between the classes represented by these image features. Initial classification is performed by applying the trained SVM on each pixel.

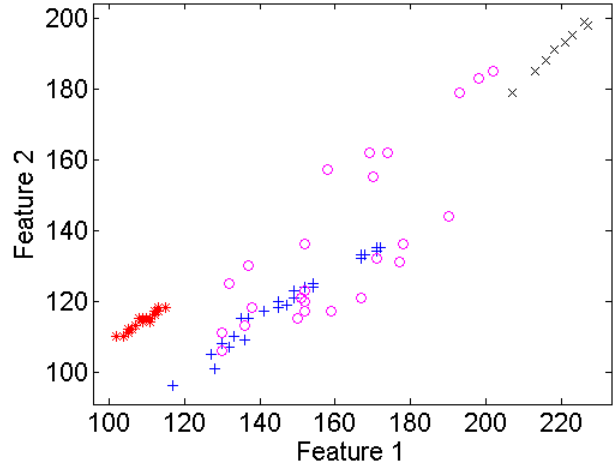


Fig. 3: Distribution of two image features related to the predefined classes

The result of the initial classification is for each pixel the most probable class including its probability and additionally a second class and its probability if there are two classes with high probabilities. The two classes and their probabilities will be used when the fusion of several initial classification results will be performed, see section 5.

Fig. 4 lists the color representation of the classes that are classified in initial classification and that are used in the following classification images.

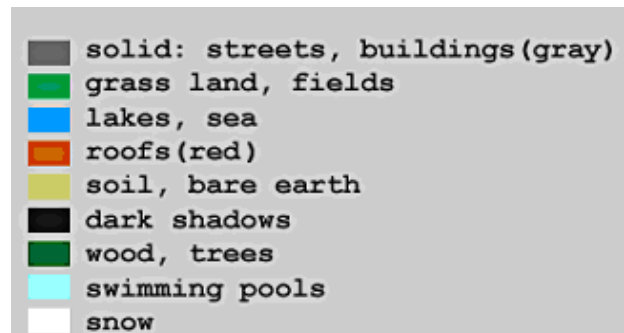


Fig. 4: Classes for initial classification and their color representation

Fig. 5 illustrates a detail of the classification result based on the color scheme in Fig. 4.

A more complex example for the input and output of initial classification is depicted in Fig. 6. The input is RGB and NIR data of a flight stripe and output is a sequence of initial classification results.

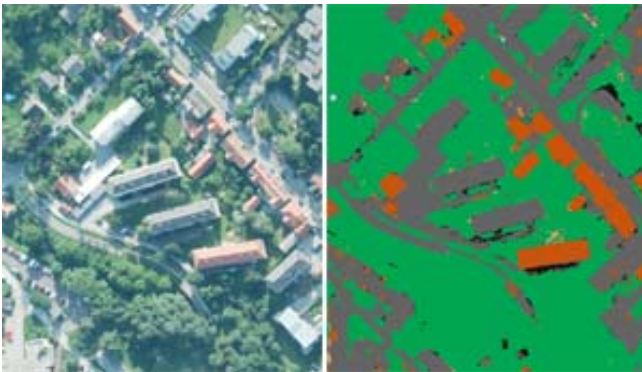


Fig. 5: (left) RGB image detail
(right) Initial classification result

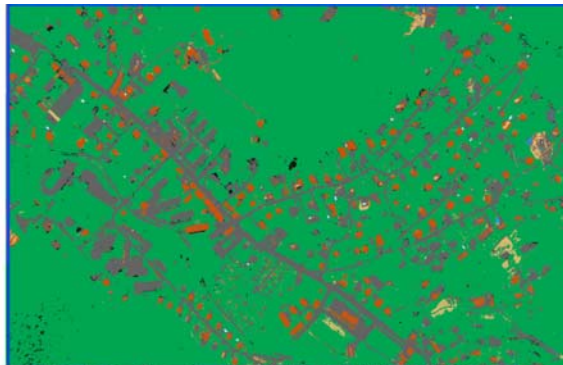


Fig. 6: (bottom) RGB images of a flight stripe
(top) Initial classification results of flight stripe

In a supervised classification the analyst identifies several areas in an image which represent known features or land use. These known areas are referred to as 'training sites' where groups of pixels are a good representation of the land cover or surface phenomenon. Using the pixel information the classification procedure then looks for other areas which have a similar grouping and pixel value. The analyst decides on the training sites and thus supervises the classification process.

The identification of training sites has to be done

- after radiometric camera calibration
- when the weather and lightening conditions or the land properties significantly change

3. AUTOMATIC AERIAL TRIANGULATION

Digital airborne cameras are able to deliver high redundant images which result in small baselines. Normally, the stripes of images have at least 80% forward overlap and at least 20% side overlap (in urban areas 60% side overlap). This high redundancy and the constraint motion of a plane help to find good starting solutions needed for a fully automated AT. Nevertheless, an accurate extraction of tie points is needed for a robust and accurate AT (Thurgood, 2004). Our POI extraction is based on Harris points and POIs from line intersections (Bauer, 2004).

After the POIs extraction in each image we calculate feature vectors from the close neighborhood. These feature vectors are used to find 1 to n correspondences between POIs in two images. The number of candidates is further reduced using affine invariant area based matching. In order to fulfill the non-ambiguous criteria, only matches with a high distinctive score are retained. The robustness of the matching process is enhanced by processing a back-matching as well.

Another restriction is enforced by the epipolar geometry. Therefore the RANSAC method is applied to the well known five point algorithm (Nister, 2003). As a result we obtain inlier correspondences as well as the essential matrix. By decomposition of the essential matrix the relative orientation of the current image pair can be calculated.

This step is accomplished for all consecutive image pairs. In order to get the orientation of the whole set, the scale factor for additional image pairs has to be determined. This is done using corresponding POIs available in at least three images. A block bundle adjustment refines the relative orientation of the whole set and integrates other data like GPS or ground control information. Fig. 7 shows an oriented block of images.

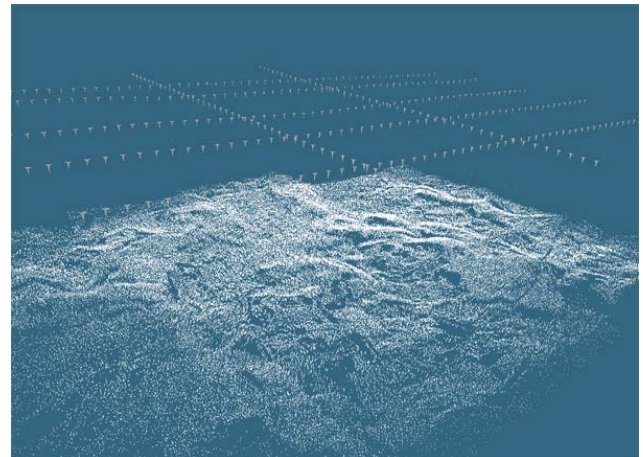


Fig. 7: Oriented block of 7 stripes of about 50 images each denoted by a small arrow

The 7 x 50 aerial images are oriented to each other using about 70.000 tie points on the ground which are shown as white dots in Fig. 7. The whole block of images was processed without any human interaction.

4. DENSE MATCHING

Once the AT is finished we perform a dense area based matching to produce a dense DSM (digital surface model). During the last few years more and more new dense matching

algorithms were introduced. A good comparison of stereo matching algorithms is given in a paper by Scharstein et al. (Scharstein, 2002). Recently, a PDE based multi-view matching method was introduced by Strecha et al. (Strecha, 2003). In our approach we focus on an iterative and hierarchical method based on homographies to find dense corresponding points. For each input image an image pyramid is created and the calculation starts at the coarsest level. Corresponding points are determined and upsampled to the next finer level where the calculation proceeds. This procedure continues until the full resolution level is reached. A more detailed description of this algorithm implemented on graphics hardware can be found in (Zach, 2003).

The workflow for aerial triangulation and dense matching is as follows: the input dataset consist of one or more flight stripes of panchromatic data, camera parameters and approximations for the exterior orientation of the images. The resulting output is a DSM, see Fig. 8.

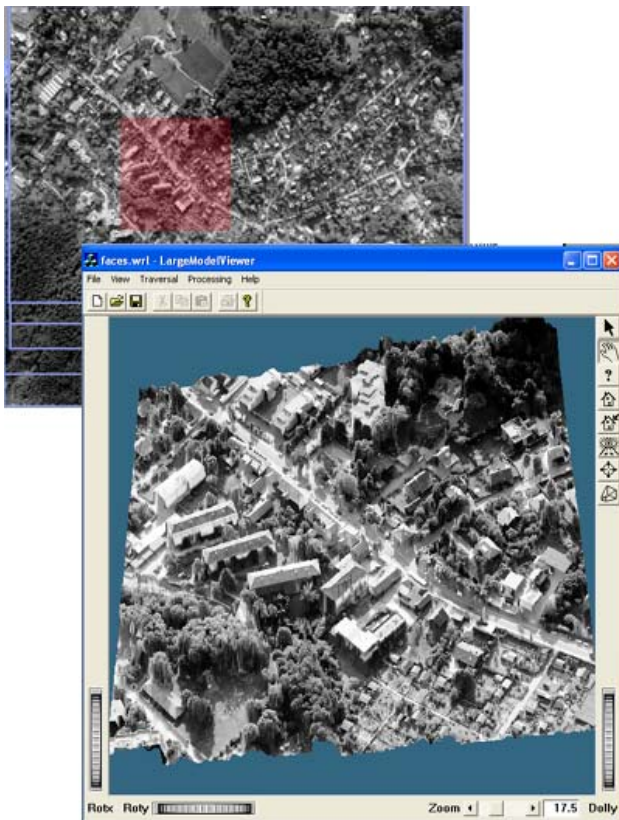


Fig. 8: (top) panchromatic images of flight stripe
(bottom) DSM textured with panchromatic image data of the area marked red in the top image

5. DATA FUSION AND ORTHO IMAGE GENERATION

An ortho image is obtained by the ortho projection of the DSM, see Fig. 9 for an RGB ortho image. The color information of the ortho image is calculated using all available aerial images of one or more flight stripes and is based on view-dependent texture mapping described in (Bornik, 2001). The color information may either be panchromatic, RGB or CIR (NIR-R-G).



Fig. 9: Ortho RGB image

The fusion of initial classification results is done in the following way:

- Determine if a pixel of an initial classification result is visible in the ortho image, i.e. not hidden by an object
- For each visible pixel get the class with highest probability as well as the class with second highest probability – if available
- Perform special handling of shadows: remove visible shadow results if other initial classification results have more specific results like solid or vegetation
- Perform majority voting using the classes with highest and second highest probability for all visible pixel

The advantages of the ortho classification are

- Moving objects are removed, see Fig. 10
- Very scattered classification results are improved, see Fig. 11

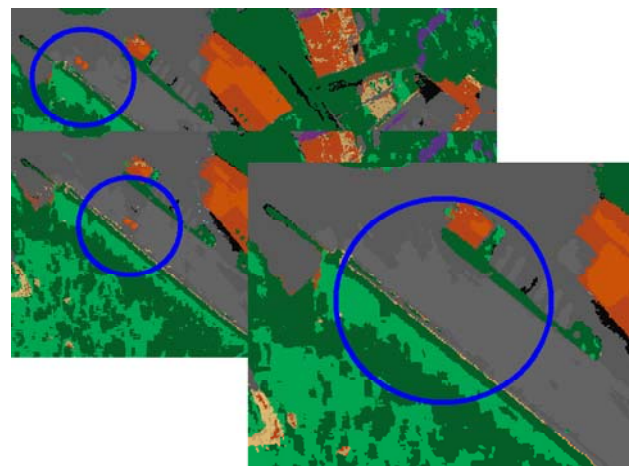


Fig. 10: Moving red car in the 2 left images is removed in the ortho initial classification in the right image

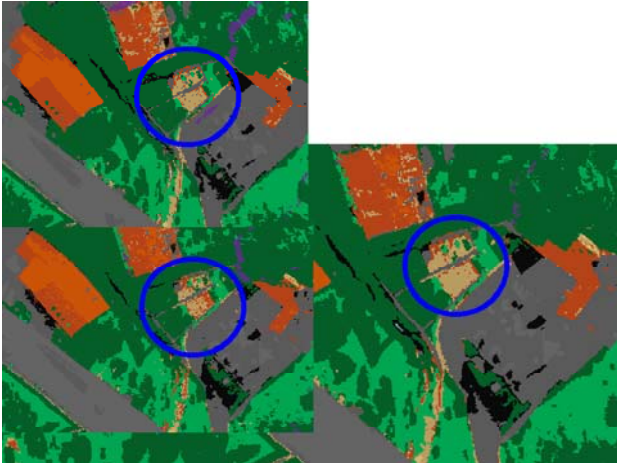


Fig. 11: Scattered classification result in a kitchen garden is improved in the ortho initial classification in the right image

The fusion of several initial classification results improves the quality of the classification. Scattered results, for example caused by chimneys or windows on roofs, are improved in refined classification when a fusion with the height data is performed.

6. DATA FUSION AND REFINED CLASSIFICATION

Data fusion and the use of multiple classifiers, see (Roli, 2002), is a topic of special interest. The following data form the basis for data fusion in refined classification:

- Ortho-initial-classification
- Ortho-panchromatic image
- Ortho-height image extracted from the DSM

Refined classification updates the results from initial classification arranged into an ortho classification using the spatial properties of the 3D features and the high resolution ortho panchromatic image. Spatial properties allow to differentiate trees from low vegetation, concrete roofs from streets, etc. Data fusion includes the computation of additional information from the ortho input data like height gradients, building blocks and texture measures.

The following refinement of initial classification results is performed:

- Solid gets refined into
 - Streets depicted gray
 - Buildings depicted in yellow
- Vegetation gets refined into
 - Grass land and fields depicted in bright green
 - Wood and trees depicted in dark green

The refined classification of objects of class solid implies the training of a minimal building height to distinguish between objects with low height like cars and small huts. The minimal building height is used to compute building blocks. Building blocks are defined as local height maxima as described in (Bolter, 2001) that are restricted to all non vegetation and non water classes. The building blocks are computed in the following way for each pixel classified as solid or roof in initial classification:

- Compute significant minimal height value in a region with specified radius

- Compute maximal height difference, i.e. the difference between height value and significant minimal height value
- If the maximal height difference is higher than the trained minimal building height, then the pixel belongs to a building
- Remove small regions up to a specified size to prevent for example street-lamps or other small but high objects to be classified as buildings

Refined classification for objects of class solid or roof is based on the computed building blocks. See Fig. 12 for an example on building blocks and on refined classification results in which buildings – in yellow – are correctly classified. The RGB ortho image of the scene used is given in Fig. 9.

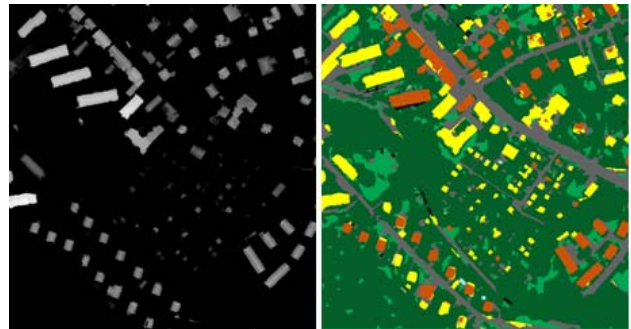


Fig. 12: (left) building blocks
(right) refined classification with buildings in red and yellow – depending on roof color – as well as trees in dark green and grass in bright green

Fig. 13 shows how small objects like street-lamps or cars are handled in refined classification. The 2 lamps are seen as solid objects with large height values which may lead to a classification as building. Due to the small size of the lamps they are not classified as building but as solid. Additionally small red objects like the red car in Fig. 13 that are misclassified as roofs are reclassified as solid in the refined classification due to their small size.

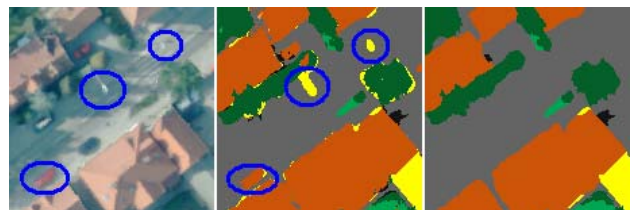


Fig. 13: (left) RGB ortho image with 2 street lamps and a red car marked
(middle) refined classification without verification of regions with small size
(right) refined classification result ignoring the street-lamps and the car

Refined classification detects not only buildings but refines the class vegetation into grass and wood or trees. The refinement is based on a SVM trained using the following features:

- the panchromatic value
- mean and standard deviation of panchromatic values in a specified neighborhood of the pixel
- mean and standard deviation of height gradient values in a specified neighborhood of the pixel

The use of height gradient values improved the detection of wood or trees compared to approaches where only texture measures are used. Fig. 14 gives an example of a house as well as trees, grass and solid. The trees are correctly classified.

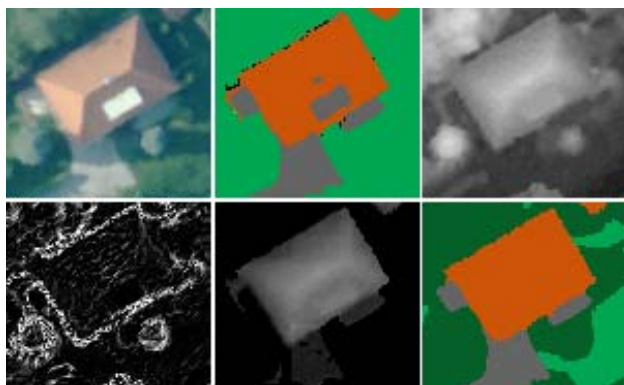


Fig. 14: (top left) RGB image of house with red roof
(top middle) initial classification
(top right) height data
(bottom left) height gradients
(bottom middle) building blocks
(bottom right) refined classification

Refined classification performs data fusion in a way that the classification results are less scattered, see again Fig. 14: the initial classification – top middle image – has the roof correctly classified but the chimney and a small roof over a window are classified as solid. The height data – top right image – as well as the height gradients – bottom left image – and the building blocks – bottom middle image – cause a classification of the whole roof as one block, see bottom right image.

7. CONCLUSIONS AND FUTURE WORK

In our approaches we use the high redundancy in the source input images to execute classification. The fusion of initial classification results to an ortho classification, the DSM generation and the fusion of ortho images and DSM are based on images with a high degree of image overlap. The classification task is performed without human interaction after an initial training phase. In the future we see the need to extract edge-based information and to construct digital elevation models from the DSM and the classification results.

8. REFERENCES

J. Bauer, H. Bischof, A. Klaus, K. Karner, 2004, Robust and fully automated Image Registration using Invariant Features, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXV, Istanbul, Turkey, ISSN 1682-1777.

R. Bolter, 2001, *Buildings from SAR: Detection and Reconstruction of Buildings from Multiple View High Resolution Interferometric SAR Data*, Dissertation, TU Graz.

A. Bornik, K. Karner, J. Bauer, F. Leberl, H. Mayer, 2001, High-quality texture reconstruction from multiple views, *The Journal of Visualization and Computer Animation*, Volume 12, Issue 5, 2001, Online ISSN: 1099-1778, Print ISSN: 1049-8907, John Wiley & Sons, pp. 263-276.

C.-C. Chang and C.-J. Lin, 2005, *LIBSVM: a Library for Support Vector Machines*, National Taiwan University.

C.-W. Hsu, C.-C. Chang, and C.-J. Lin, 2003, *A Practical Guide to Support Vector Classification*, Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan.

C. Huang, L.S. Davis, J.R.G. Townshend, 2002, An assessment of support vector machines for land cover classification, *Int. J. Remote Sensing*, Vol. 23, No. 4, pp. 725-749.

A. Klaus, J. Bauer, K. Karner, K. Schindler, 2002, MetroGIS: A Semi-Automatic City Documentation System, *Photogrammetric Computer Vision 2002 (PCV'02)*, ISPRS - Commission III, Symposium 2002, September 9 - 13, 2002, Graz, Austria.

F. Leberl, J. Thurgood., 2004, The Promise of Softcopy Photogrammetry Revisited. ISPRS 2004, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXV, Istanbul, Turkey, ISSN 1682-1777.

F. Leberl, R. Perko, M. Gruber, 2002, Color in photogrammetric remote sensing, *Proceedings of the ISPRS Commission VII Symposium*, Hyderabad, India, Vol. 34, pp. 59-64.

D. Nister, 2003, An efficient solution to the five-point relative pose problem, *CVPR 2003*, pp. 195-202.

F. Roli, J. Kittler, 2002, Fusion of multiple classifiers, *Information Fusion*, Vol. 3, pp. 243.

D. Scharstein and R. Szeliski, 2002, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1/2/3), pp. 7-42.

M. Sormann, A. Klaus, J. Bauer, K. Karner, 2004, VR Modeler: From Image Sequences to 3D Models, *SCCG (Spring Conference on Computer Graphics) 2004*, ISBN 80-223-1918-X, pp. 152-160.

C. Strecha, T. Tuytelaars, L. Van Gool, 2003, Dense Matching of Multiple Wide-baseline Views, *ICCV 2003*, Vol 2, pp. 1194-1201.

J. Thurgood, M. Gruber, K. Karner, 2004, Multi-Ray Matching for Automated 3D Object Modeling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXV, Istanbul, Turkey, ISSN 1682-1777.

V. N. Vapnik, 1995, *The Nature of Statistical Learning Theory*, Springer, 1995.

C. Zach, A. Klaus and K. Karner, 2003, Accurate Dense Stereo Reconstruction using 3D Graphics Hardware, *Eurographics 2003*, Short Presentations, pp. 227-234.

9. ACKNOWLEDGEMENTS

This work has been done in the VRVis research center, Graz/Austria (<http://www.vrvis.at>), which is partly funded by the Austrian government research program Kplus. We would also like to thank Vexcel Imaging Austria, (<http://www.vexcel.at>) and Vexcel Corporation (<http://www.vexcel.com>) for supporting this project.