FROM PIXELS TO BUILDINGS

B. C. Gruber-Geymayer, L. Zebedin, K. Karner

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

KEY WORDS: Land Use**,** Classification, Orthoimage, DEM/DTM, Building, Visualization

ABSTRACT:

This paper describes the land use classification of multispectral digital aerial images, the removal of buildings from the digital surface model and the visualization of the textured digital surface model with reconstructed buildings. The proposed approach applies spectral classification techniques to multispectral digital aerial images with RGB and NIR channels. The results of the land use classification are used both in dense matching and building extraction. Dense matching uses the knowledge of water areas to prevent wrong matches due to non-Lambertian reflections. The obtained digital surface model is used to generate the corresponding RGB ortho images. Further, in building extraction the classification is used as starting point for searching exact building borders and textures for building facades. The buildings in the digital surface model are replaced by refined building models using feature information. Results from a huge test area in city center of Graz are presented and analyzed. This approach produces automatically appealing visualization after a short interactive training phase for classification.

1. INTRODUCTION

This contribution deals with images from the large format digital camera UltraCamD from Vexcel 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 high resolution panchromatic images have a size of 11500 pixels across-track and 7500 pixels along-track. The multispectral low resolution images haves a size of 3680 by 2400 pixels. The panchromatic high resolution images as well as the low resolution multispectral images are used for this approach.

The data set used in this paper was acquired in the summer of 2005 from the city center of Graz, Austria. It consists of 155 images flown in 5 strips. The along-track overlap of this data set is 80%, the across-track overlap is approximately 60%. The ground sampling distance is approximately 8cm.

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,
- removal of buildings from the DSM, see section 7,
- building reconstruction, see section 8, and finally
- visualization of results, see section 9.

The focus of this contribution lies on detection and description of buildings in the classification and reconstruction process.

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 classes rely on color and infrared and will be refined later using height information from stereo matching.

The classifier used for the supervised classification is a support vector machine (SVM). 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. SVMs have - after careful pre-processing – results that have similar accuracy like nearest neighbor or neural networks but the SVMs are more stable in use. Initial classification relies on 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. The classes trained for the Graz dataset are:

- Solid: man made structures like streets or buildings
- Water: lakes, rivers, sea
- Vegetation: wood, grassland, fields
- Dark shadows

Other classes like red roofs, bare earth, snow or swimming pools were not trained as they are not relevant for the following processing steps that concentrate on building reconstruction in a landscape.

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 can 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

The feature values are scaled to prevent that features with greater numeric ranges may dominate. The scaling factors are determined during training and applied for later classification.

We use a supervised classification: the analyst decides on the training sites and thus supervises the classification process. The training sites are several areas in an image which represent known features or land use.

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.

A complex example for the output of the initial classification is illustrated in Fig. 1. The color scheme used to represent the classification results is listed in the lower right corner of Fig. 1. Details on initial classification are described in (Gruber-Geymayer, 2005).

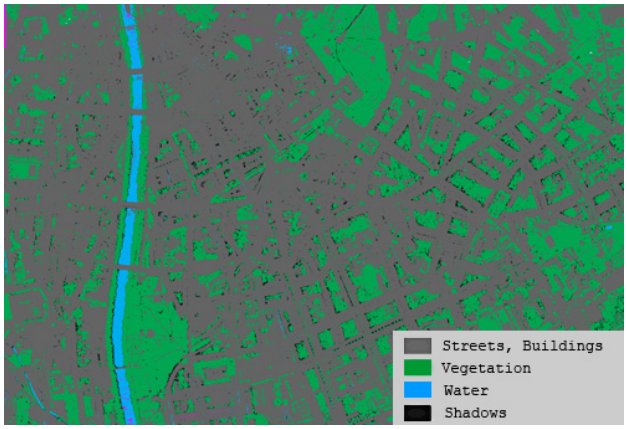


Fig. 1: Initial classification results of Graz dataset

3. AUTOMATIC AERIAL TRIANGULATION

The UltraCamD camera is able to produce highly overlapping images at very short baselines. Such dense block of images serves well for a robust automated aero triangulation. We start with a POI extraction in each image and 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 applying a back-matching as well.

This step is accomplished for all consecutive image pairs. In order to compute the orientation of the entire image 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. 2 shows an oriented block of images.

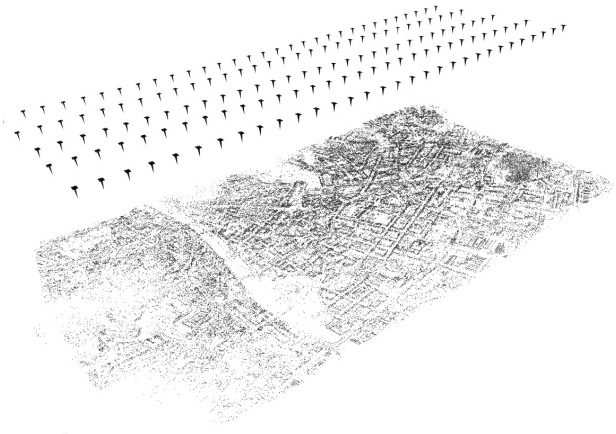


Fig. 2: Oriented block of 5 stripes of 31 images: arrows denote images and black dots denote 70.000 tie points

The 5 x 31 aerial images are oriented to each other using about 70.000 tie points on the ground which are shown as black dots in Fig. 2. The whole block of images was processed without any human interaction. Details are described in (Bauer, 2004).

4. DENSE MATCHING

Once the AT is finished we perform a dense area based matching to produce a dense DSM (digital surface model). 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 level where the calculation proceeds. This procedure continues until the full resolution level is reached. Dense matching uses the results of initial classification: running water areas are excluded from computation to prevent wrong matches due to reflections. A detailed description of this algorithm is given in (Zebedin, 2006b). The DSM for the Graz dataset is illustrated in Fig. 3.

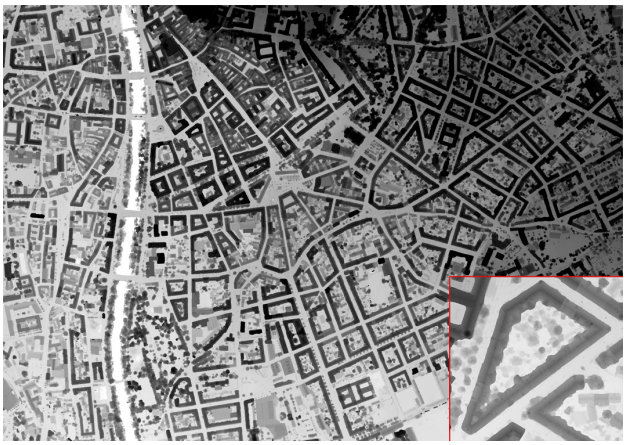


Fig. 3: DSM computed from 155 images of the Graz dataset; Note the enlargement at the lower right.

5. DATA FUSION AND ORTHO IMAGE GENERATION

An ortho image is obtained by the ortho projection onto the DSM, Fig. 4 shows an RGB ortho image including an image detail. The color information of the ortho image is calculated using all available aerial images and is based on view-dependent texture mapping. The color information may either be panchromatic, RGB or CIR (NIR-R-G).

Fig. 4: Ortho RGB image of the Graz dataset. Note the enlargement at the lower right.

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

- Determine if a pixel of the initial classification result is visible in the ortho image, i.e. not hidden by an object
- For each visible pixel select 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 a majority voting using the classes with highest and second highest probability for all visible pixel

The fusion of several initial classification results from multiple source images improves the quality of the classification.

6. REFINED CLASSIFICATION

The following refinement of the initial classification results is performed during refined classification:

- Solid gets refined into streets or parking areas and buildings
- Vegetation gets refined into grass land or fields and wood or trees

The refined classification of objects of class solid implies the specification of a minimal building height to distinguish between objects of lower height like cars and small huts. The minimal building height is used to compute building blocks. Building blocks are defined as local height maxima 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 in initial classification:

- Compute the significant minimum height value in a region with specified radius
- Compute the maximum height difference, i.e. the difference between height value and significant minimum height value
- If the maximum height difference is higher than the specified minimum building height, then the pixel belongs to a building
- Remove small buildings up to a specified size to prevent small but high objects like street-lamps to be classified as buildings

Refined classification for objects of class solid or roof is based on the computed building blocks. See Fig. 5 for an example on building blocks and on refined classification results in which simple as well as complex buildings are correctly classified and are represented in yellow.

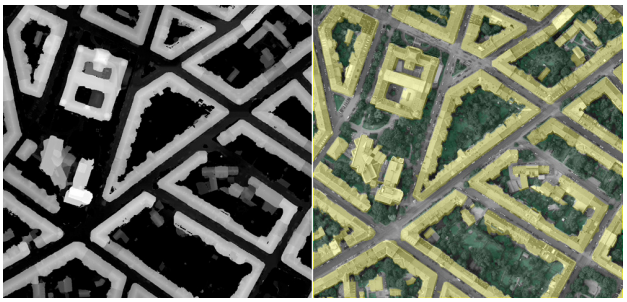


Fig. 5: (left) building blocks
(right) refined classification with buildings in yellow

The computation of building blocks in such way as described above implies that there will be solid objects classified as buildings that have a height difference to the neighborhood but are not considered to be buildings, for example

- construction sites
- undercrossings
- embankments

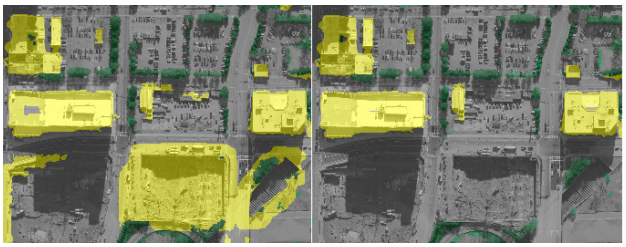


Fig. 6: Removal of objects with height differences that are misclassified as buildings

The algorithm considers the border of a building and computes the number of border pixels with

- a height difference to the neighbor pixels that is larger than the specified minimum building height
- no significant height difference to neighbor pixels from the class solid
- no significant height difference to neighbor pixels that are not classified as solid, e.g. trees near a building

Depending on the number of border pixels assigned to these classes the building is removed, or marked as building with low probability or accepted.

Refined classification not only detects buildings but refines the class vegetation into the class grass and the class wood or trees. Refined classification performs in a way that the classification results are less scattered and regions up to a minimal size are removed.

The refined classification result for the whole city of Graz dataset – the related RGB ortho image of the scene is presented in Fig. 4 - is given in Fig 7**.** The area marked in red is the one visualized in Fig. 12 with a textured DSM and the reconstructed buildings. Details on refined classification are described in (Gruber-Geymayer, 2005).

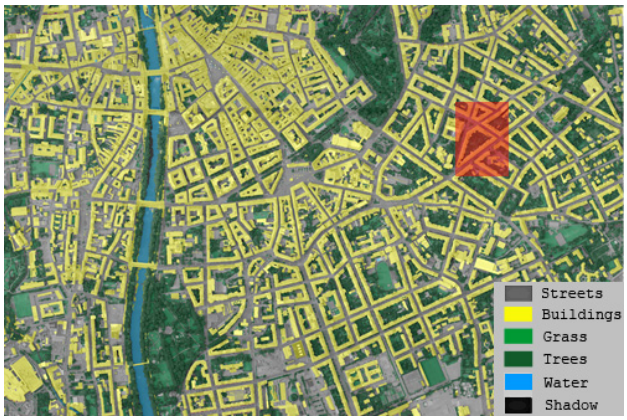


Fig. 7: Refined classification results for Graz dataset; the area marked in red is visualized in Fig. 12

7. REMOVAL OF BUILDINGS FROM THE DSM

The DSM represents the entire landscape with buildings and trees. The generation of a Digital Terrain Model (DTM) from the DSM can be performed using the refined classification results. Objects of type building and tree have to be removed from the DSM to get the DTM. The height for buildings and trees is computed in the local neighborhood: the difference between height value in the DSM and a significant minimal height value in the local neighborhood.

In our current approach we are less interested in the DTM but in a DSM where only buildings are removed. The DSM with removed buildings can be textured by the ortho RGB image and the reconstructed buildings can be placed into this textured landscape. The concept to remove buildings from the DSM is as follows: for each pixel classified as building in refined classification

- Determine in the 8 main directions the next pixel "on earth", i.e. solid or grass pixel, see Fig. 8
- Determine the height value for the "on earth" pixel as well as its distance from the pixel considered
- Compute the weighted mean value to get the height value for the building pixel

A more sophisticated interpolation by hyperplanes is not necessary as the removed buildings are replaced by reconstructed buildings. The buildings will be modeled by facades and roofs extracted from aerial images, see section 8.

Fig. 8: Removal of building: 8 points used for height interpolation

The removal of all buildings from DSM for the Graz dataset generates a DSM as depicted in Fig. 9. Please compare to the input DSM in Fig. 3.

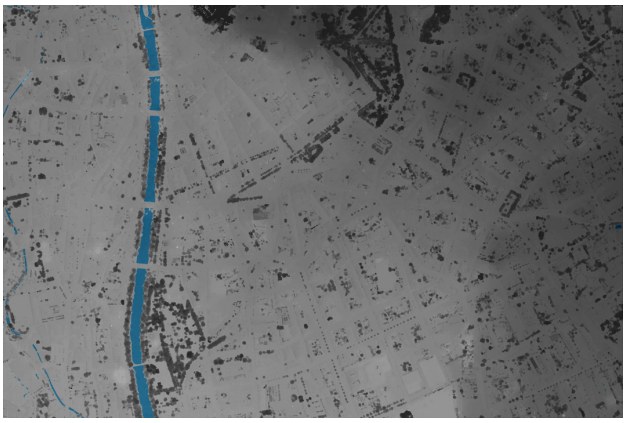


Fig. 9: DSM of Graz dataset, buildings have been removed

8. BUILDING RECONSTRUCTION

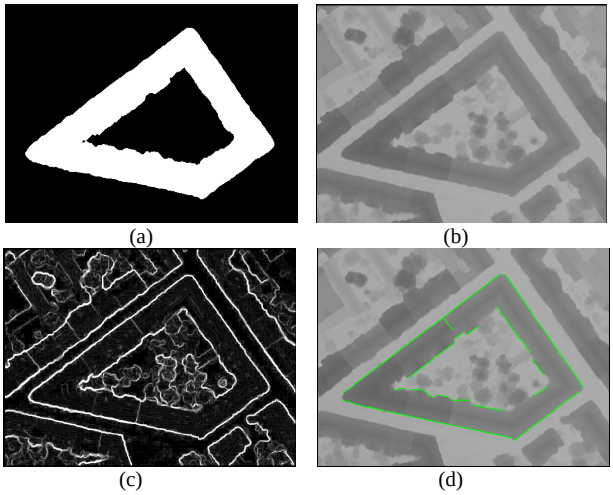
The building reconstruction is performed in 2 steps: optimized building facades are determined in one step and optimized roof planes in a second step. The algorithm for obtaining building facades can be decomposed into three steps: first some hypotheses are derived from the refined classification results. Then these facades are translated in such a way, that they are parallel to the true facade. In the last step the fine-grained optimization using multi-view correlation is performed.

The algorithm to find roof planes is based on finding clusters of points that can be optimally represented by a roof plane. The following subsections outline the building reconstruction. A detailed description of façade reconstruction is given in (Zebedin, 2006a).

8.1 Initial façade positions

A building layer that describes the position of buildings can be extracted from the refined classification results. The building layer and the height field are used to get the initial estimates for the position of facades. The search for facades can be restricted to regions near buildings using the building layer. We apply an edge detector to the height field to detect lines in the building regions. One important parameter of this line extraction is the minimum length of each line, as longer lines tend to be more stable in the ongoing optimization steps.

The result of this procedure is illustrated in Fig. 10. Note that only lines near buildings are extracted whereas there are no lines near the tree in the inner courtyard of the building, see Fig. 10(d). These lines in 2D are then extended to 3D planes by estimating the minimum and maximum height from the surrounding area in the height field. A small margin is subtracted from the top and bottom of the plane to account for possible occlusions near the roof and the ground.



(a) (b) (c) (d)

Fig. 10: Initial line extraction process in the height field
(a) building-layer as result of refined classification
(b) height field
(c) gradient image (Sobel) and
(d) height field with extracted lines in green

8.2 Line Direction Optimization

The first optimization applied to the facade planes aligns the hypothesis from the initial step to the real facade. As a result the plane should afterwards be parallel to the real facade. The algorithm relies on the fact that facades mainly contain structures which are horizontally or vertically aligned with the facade itself: windows, balconies, signs and similar structures.

For each facade plane the algorithm first makes a ranking of all available cameras and assigns each one to a score. When the optimal camera has been determined, the corresponding image is correctly resampled by taking into account the perspective view. A Gaussian filter is then applied to remove small artifacts and distortions. In the following step we compute an orientation histogram: each peak in this histogram corresponds to one strong line direction in the texture. The proposed assumption that facades contain horizontally and vertically aligned structures allows concluding that the peak closest to zero can be moved to zero to make the facade plane parallel to the real facade. This enables us to calculate an orientation change which compensates this deviation of the peak by intersecting the lines from camera center to the endpoints of the detected lines with the horizontal plane. This ensures that the new line is horizontal and therefore that the plane is parallel to the real facade.

8.3 Correlation Optimization

In the third and last step of façade optimization each facade plane is moved forward and backward to increase the correlation of warped textures from different views. A hierarchical approach is used to overcome problems with mismatches caused by inaccurate initial positions. This means that each warped texture is turned into an image pyramid and starting with the coarsest level the correlation optimization is performed until the highest resolution level is reached. The quality of the optimization is usually good except in those cases where there are occlusions in all images, e.g. trees in inner courtyards, or the facade can not be satisfyingly be approximated with one plane.

8.4 Roof reconstruction

The roof planes are detected using a RANSAC approach which is similar to (Samadzadegan, 2005). First three random points from the local point cloud are selected and used as the hypothesis. Then points are classified as inliers or outliers depending on their distance to the hypothetical plane. Two separate roofs may lie on the same plane, therefore a clustering is applied to the inliers to retain only points which really pertain to one roof. Once a cluster is accepted, those points are removed from the point cloud and the process is repeated until no further planes with sufficient support are found. Some points could be assigned to two or more planes because they lie on an edge of the roof. This ambiguity is alleviated by initially accepting planes with a high support only. This ensures that dominant planes are fitted first.

As a final result we show a block of buildings from the Graz dataset (marked red in Fig. 7). The visualization is presented in Fig 11. Roofs and outer facades are good but some facades in the courtyard suffer from occlusions by trees. These shortcomings are overcome when these building models are placed into the landscape and the trees are visualized too, see Fig. 12.

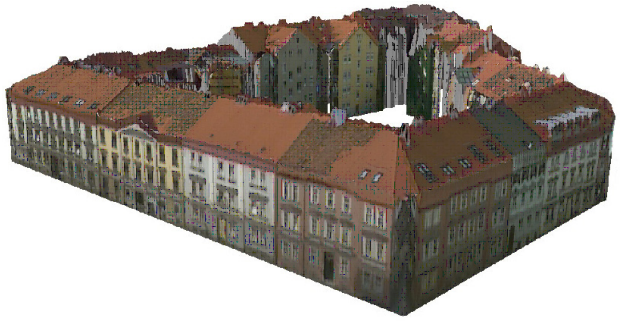


Fig. 11: Building representation with optimized facade and roof planes

9. VISUALISATION OF RESULTS

Finally the reconstructed buildings are placed into a model of the landscape as follows: the DSM from which the buildings have been removed is textured with the ortho RGB image. Thus streets with cars or vegetation get their representation. The reconstructed buildings with facades and roofs are now placed into the textured landscape. A small area of Graz represented by 265000 triangles can be seen in Fig. 12. The area represented is marked red in Fig. 7.

Fig. 12: Reconstructed buildings placed into textured DSM; visualised region is marked red in Fig. 7

10. CONCLUSIONS

In our approach we use the high redundancy in the source input images to generate city models fully automatically. The steps performed are a land use classification, aerial triangulation, dense matching, ortho image generation, removal of buildings from the DSM as well as building reconstruction. The whole task is performed without human interaction after an initial training phase for classification. The algorithms are outlined and their results are demonstrated using a dataset from Graz. Further test sites were cities with skyscrapers on the one hand and suburban areas with a mixture of small houses and gardens with trees.

11. REFERENCES

Bauer J., Bischof H., Klaus A., Karner K., 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.

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

Gruber-Geymayer B. C., Klaus A., Karner K., 2005. Data fusion for classification and object extraction, *Proceedings of CMRT05*, Joint Workshop of ISPRS and DAGM, August 29-30, 2005, Vienna, Austria, pp. 125-130.

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

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

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

Leberl F., Thurgood J., 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.

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

Samadzadegan F., Azizi A., Hahn M., Lucas C., 2005. Automatic 3D object recognition and reconstruction based on neuro-fuzzy modelling, *ISPRS Journal of Photogrammetry and Remote Sensing*, 2005, Volume 59, pp. 255-277.

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

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

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

Thurgood J., Gruber M., Karner K., 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.

Zebedin L., Klaus A., Gruber B., Karner K., 2006a. Façade Reconstruction from Aerial Images by Plane Sweeping, to appear in Proceedings of PCV.

Zebedin L., Klaus A., Gruber-Geymayer B., Karner K., 2006b. Towards 3D Map Generation from Digital Aerial Images, to appear in *ISPRS Journal of Photogrammetry and Remote Sensing Theme Issue "Digital Aerial Cameras"*.

12. 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 (http://www.vexcel.com) for supporting this project.