Online Geocoding and Evaluation of Large Scale Imagery without GPS

WOLFGANG FÖRSTNER, RICHARD STEFFEN, Bonn

ABSTRACT

Large scale imagery will be increasingly available due to the low cost of video cameras and unmanned aerial vehicles. Their use is broad: the documentation of traffic accidents, the effects of thunderstorms onto agricultural farms, the 3D-structure of industrial plants or the monitoring of archeological excavation. The value of imagery depends on the availability of (1) information about the place and date during data capture, (2) of information about the 3D-structure of the object and (3) of information about the class or identity of the objects in the scene. Geocoding, problem (1), usually relies the availability of GPS-information, which however limits the use of imagery to outdoor applications. The paper discusses methods for geocoding and geometrical evaluation of such imagery and especially adresses the question in how far the methods can do without GPS.

1. INTRODUCTION

Large scale imagery will be increasingly available due to the low cost of video cameras and unmanned aerial vehicles (UAVs). Their use is broad: the documentation of traffic accidents, the effects of thunderstorms onto agricultural farms, the 3D-structure of industrial plants or the monitoring of archeological excavation.

1.1. UAV specific requirements for image evaluation

There are different levels of user requirements when using UAVs.

- Images taken from an UAV often are a *value on their own*, as they provide visual information about the object which otherwise cannot be obtained. Spatio-temporal reference may be crude and available without additional effort.
- In many cases, however, one is interested in 3D information about the objects. Here again we may distinguish two levels of requirements
    - The user is only interested in a *few measures in 3D space*. She therefore needs an interactive tool for image evaluation, where she can identify points, lines or areas which then are geometrically characterized by the system. Calibration and orientation of the images needs to be derived without requiring user interaction.
    - The user is interested in a *3D visualization* of the scene. Besides calibration and orientation a module for deriving a more or less accurate 3D-model of the scene is required, again not needing assistance by the user. Of course, providing a tool for interactive evaluation may be required additionally.

  In both cases highly automatic and reliable techniques need to be available which are applicable by non-experts.
- *Scale information* needs to be provided by the system or by the user.
- Finally, the user may require *georeference* of the images or the derived 3D information. Examples for such services are Google-Earth, providing worldwide georeferenced image and partially 3D information or Pictometric, providing oblique image libraries with tools for georeferenced 3D evaluation.

The procedures for solving these tasks are well known in photogrammetry. There are aspects which require additional attention when evaluating images of UAVs:

- *Real-time evaluation* is inherent to the evaluation of images of UAV. Real time is realized in case the evaluation is performed in time for decisions based on the image content. Real-time requirements directly influence all evaluation steps.
- Image sequences taken from UAVs may show *very irregular orientation*. Though UAVs may be used in a mode similar to classical aerial triangulation with regular strip patterns, the intended flexibility of their control by a non-expert will, even in case the UAV is stabilized with an inertial system and contains a GPS receiver, show large irregularities at least in azimuth and scale. In case the camera is tilted, e. g. in order to capture facades, at least two angles will show large deviations from zero.

Finally, due to both, the low altitude of the flight line and the range of applications, *general 3D structures* to be recovered in general cannot be represented with a graph surface $z=z(x,y)$, where each *xy*-position only is represented by one z-coordinate, which definitely is insufficient in case of overhangs, vegetation or even in-door applications

1.2. Outline of paper

This paper discusses techniques for the geometric evaluation of the images taken from UAVs. We first want to discuss in how far geocoding can be performed without use of GPS which may be required in centers of cities or in indoor applications. We then discuss the problem of orientation of short or wide base-line image sequences in batch and real-time mode. Finally we discuss the problem of reconstructing complex 3D surfaces.

1.3. Hardware, Software and data used for the experiments

1.3.1. Hardware

The experiments shown below are based on image sequences taken with the UAV from Microdrones GmbH. The drone shown in fig. 1 is an electric powered quad copter, which is manually controlled. It can carry up to approximately 200 g payload. This drone is equipped with a Panasonic Lumix with a resolution of 848 x 480 pixels, a viewing angle of appr. 90° and a frame rate of 30 Hz in video mode. The camera can be tilted from 0° to 90° nadir angle. The battery allows a flying time up to appr. 30 minutes. The image sequence is stored on a flash card and compressed as quick time movie.

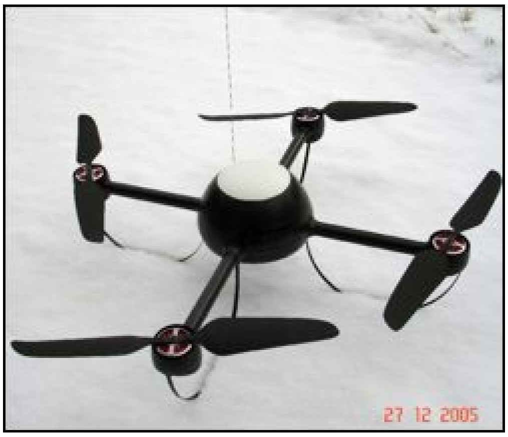

Fig. 1: Used hardware. Drone MD 4-200 from Microdrones ©
equipped with a Panasonic Lumix video camera

1.3.2. Software

We use several open source and commercial software packages. For the georeference tests we use the Lowe-SIFT point extractor (Lowe 2004), provided by D. Lowe. The tracking in the images sequences uses the OpenCV implementation of the Kanade-Lucas-Tomasi-tracker. We integrated the freely available bundle adjustment program SBA of Lourakis and Argyros (2004). The point cloud derived from multiple images uses the software package Match-T of INPHO GmbH.

1.3.3. Data

The experiments were performed near the "Drachenfels" close to the city of Bonn. From this area a 20 cm orthophoto and lidar scans with a mean point density of 1 meter was available, provided by the State Department of Geodesy and Geoinformation, Northrhine Westphalia. Furthermore we acquired an image sequence of a small area with the drone with the support of Microdrones GmbH. The average flying height was approx. 30 m. We chose two subsequences for the experiments. The first image subsequence consists of vertical views. It contains a building and vineyards with their typical rows. Fig. 2 shows every 100th image to give an impression on the roughness of the flight path. The camera was calibrated offline.

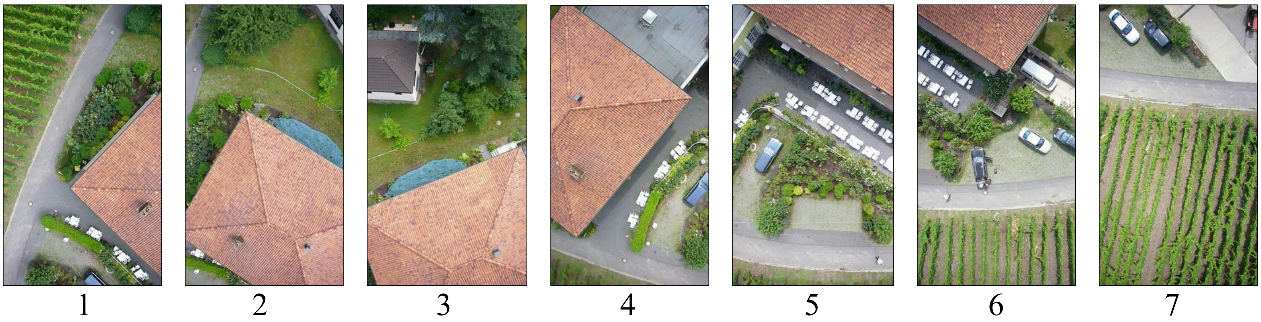

Fig. 2: Every 100th image of subsequence 1,
observe the rotations (2←→3) and scale differences (6←→7)

The second image subsequence was taken with the camera tilted by appr. 45°. It contains mainly vegetation, especially high trees.

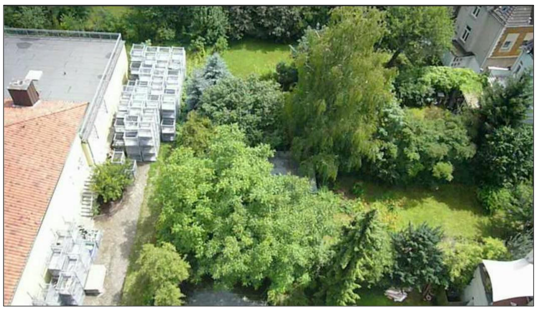

Fig. 3: Image taken from second subsequence taken with a tilted camera showing vegetation

2. GEOCODING OF LARGE SCALE IMAGERY WITHOUT GPS?

2.1. General

Geocoding of images needs external information. This usually is provided either by control points, i. e. points with known 3D coordinates which are visible in the images, or by directly by measuring the position of the camera during the exposure using GPS, which due to the availability of low cost GPS receivers has become the primary choice, see the previous presentations on this conference.

Several reasons motivate to explore the use of external information other than GPS for geocoding: GPS may give poor or no information e. g. in the vicinity of high buildings, in indoor application, or in case the image data are taken without access to GPS.

In the context of UAVs the availability of control information needs to be discussed anew. The low altitude leads to a comparably high resolution in terms of ground sampling distance and to a small footprint. We only want to discuss control information which allows automatic mensuration, matching and camera orientation. There are basically three types of ground control:

- *control points*, artificial or natural. In both cases the control point is to represented by an image template as this representation poses least restrictions on the type of point, see e. g. Rauhala et al. 1995, in contrast to roof corners (Förstner 1988) or man holes (Drewniok and Rohr 1997). Natural control points represented as image templates are only useful in case the image content between the template and the actual image is not too large. This hardly can be guaranteed for very large scale images of outdoor scenes, e. g. when matching images to orthophotos, usually being derived from imagery older than a year. Moreover the footprint of a very large scale image may only have a diameter of 10 m, say (see fig. 2 and 3). This is not enough for a reliable search as it only represents an image patch 30 x 30 pixels of an orthophoto having a resolution of 30 cm. However, indoor applications may rely on images of control points.
- *control features*, especially lines. They have been proposed and used as an alternative, as lines appear more frequently, see Paderes et al. 1984, Mikhail 1993, Habib et al. 2004. Here the situation is not really better.
- *digital surface model* (DSM) as control (Ebner/Strunz 1988). The idea of this proposal is to exploit the 3D form of the terrain to register two DSMs. The method requires sloped terrain with areas of different aspect in the template. Therefore the method is very well suited for hilly terrain or for urban areas. The method appears not to be used regularily due to the lack of 3D form information from digital images. However, there has been a large progress in automatic derivation of point clouds and DSMs from images (Scharstein/Szeliski 2002, Ton/Mayer 2007, Heinrichs et al. 2007, Match-T INPHO). The quality of recent results will stimulate the discussion on the relation between intensity image and range image based DSM derivation. In our context, we need a stable surface form over time in order to allow georeferencing of a DSM automatically derived from the image sequence and the reference DSM, e. g. provided by some state department. Obviously, one needs to investigate the effect of varying vegetation on DSMs.

Altogether, georeferencing without GPS needs stable visual or form information. For outdoor applications this may be a DSM. For indoor applications, both, visual and form information may be used.

2.2. Experiment

We want to demonstrate the feasibility of georeferencing using a reference DSM and a template DSM derived from the image data to be georeferenced. However, we cannot apply the method of Ebner and Strunz (1988) directly, as they require very precise approximate values, due to the used least square matching approach. We follow the line of object location in image analysis proposed by Lowe (2004). This method is based on salient features automatically derived from the data and which are described with rotation and scale invariant attributes. These features are derived in both, the template and the master data set. The attributes are used to match the features. The method has proven to be very robust object location in image analysis. A first transfer to use it for the registration of range data has been proposed by Wessel et al. (2006). We show that the method of Lowe directly can be applied to DSMs.

For this we match the DSM of app. 1400 x 900 m^2 with a rotated and scaled cutout of 200 x 200 m^2, simulating a DSM derived from an image sequence taken with an UAV. We use the LIDAR point cloud of the terrain covering an area, computed a grid representation with a grid size of 1 m. As the matching software of Lowe only can handle integer images with a range of 0-255 we use a scaled version of the highpass filtered image $g'=g-B*g$ for the matching, here B is a boxfilter of the size of the template of 200 x 200 pixels. The template then was rotated by 70^o and scaled by 1.5 to simulate the total uncertainty about the azimuth and the partial knowledge of the scale, transferred from the starting area of the UAV. Reference and template DSM are shown in the fig. 4.

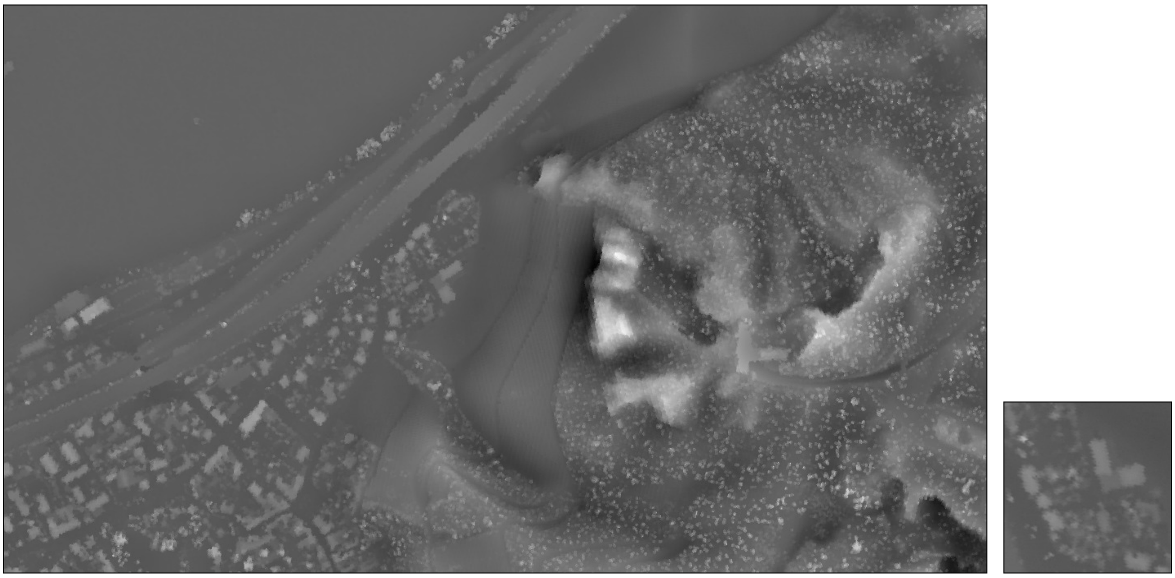

Fig. 4: DOM-Georeferencing. Left: high pass filtered reference DOM.
Right: template DOM, 200 x 200 m^2. Observe the rotation and scale of the small template.

The result of the matching experiment is the following: Altogether 2127 points were extracted in the reference-DOM, 31 in the template-DOM. The matching eliminated 12 points due to dissimilarity, and yielded 19 correct and 3 incorrect matches. This allows an easy determination of the parameters (translation, rotation and scale) using a robust procedure like RANSAC (Fischler/Bolles 1981). Refinement of these approximate values may use least squares matching as proposed by Ebner and Strunz (1988).

3. EVALUATION OF VERY LARGE SCALE IMAGE SEQUENCES

3.1. General

The evaluation of images usually consists of two steps:

- The determination of the relative and possibly absolute orientation
- The 3D reconstruction of the visible surface.

3.1.1. Orientation of the images

In the context of UAVs we may acquire

1. image sequences with a large frame rate of, say, 30 Hz. We refer to this case as small baseline imagery.
2. a sequence of images with a low frame rate of, say, 1 Hz as in classical photogrammetry. We refer to this case as wide baseline imagery.

In both cases the orientation may be achieved using *bundle adjustment* (see Triggs et al. 2000), which is the usual procedure for wide base line imagery. In case of small baseline imagery the point transfer between images can reliably be performed by tracking of feature points. However, the orientation of large image sequences soon gets prohibitive, as ½ hour of video already produces 54000 images, when assuming a frame rate of 30 Hz. Therefore it is useful to transfer the points within the image strip by a tracking procedure using all images, but using only a sparse set of key frames (images) for the orientation of the complete strip. This procedure is much more stable, than matching points between the key frames.

Alternatively one may determine the orientation of the images using a *Kalman filter*. This recursive estimation is a well known technique in all kinds of automatic control applications (Kalman 1960). A. J. Davison (2004) shows a robustly working system with a single camera based on a linear camera motion model for in-door scenes. Due to the small base line geometry the representation of 3D points with their coordinates has been shown to be unstable, therefore the representation uses the inverse depth, which in a first approximation is proportional to the parallaxes (Montiel et al. 2006).

3.1.2. 3D-Surface generation

The generation of 3D-surfaces usually is a second step based on the recovered orientation parameters. Of course, the bundle adjustment as well as the Kalman filter yields highly stable 3D-point, which may be used as a first approximation for a 3D surface representation. However, the point density usually is not exploiting the image content, which would require an intensity based reconstruction of the surface or a much higher density of the 3D points together with a dense interpolation. In addition, the surface reconstruction would require (1) the simultaneous use of multiple images (see Heinrichs et al. 2007) and (2) a truly 3D representation allowing overhangs or even holes (Ton/Mayer 2007). Here techniques known from the evaluation of LIDAR data may be used to advantage.

In the following we want to demonstrate the feasibility of classical bundle adjustment and Kalman filter and subsequent multi-image matching techniques applied to very large scale imagery.

3.2. Examples

3.2.1. Orientation Experiments

In a first step we want to show the orientation of an image sequence can be performed fully automatically in both modes, batch with a bundle adjustment and sequentially with a Kalman filter for real time applications.

Our implementation of a Kalman filter for the orientation and reconstruction form video sequences also uses an inverse depth representation for object points. To achieve optimal accuracy we implemented an iterative extended Kalman filter. For better robustness we, in the follow up version, will use an iterative implicit Kalman filter (Steffen/Beder 2007).

The bundle adjustment solution is performed with every 10th frame of the two image subsequences. Appr. 2500 object points were determined. The datum is defined automatically by choosing that frame which yields an optimum accuracy for the coordinate system, see Läbe/Förstner 2005. The scale is fixed using a known distance between two object points. The estimated σ_0 was 0.5, indicating the tracker to yield a standard deviation of 0.5 pixels. The Kalman filter approach uses a subset of around 1000 object points. Here, the datum is typically defined by the first camera position and orientation. All frames are used. The initialization of the Kalman filter is performed according to Beder/Steffen (2006). In order to be able to compare the two results we transformed the Kalman filter results into the same coordinate system as the bundle adjustment results. As can be seen in Fig. 5 and 6 the trajectories are similar; up to now we did not perform a statistical analysis of the differences.

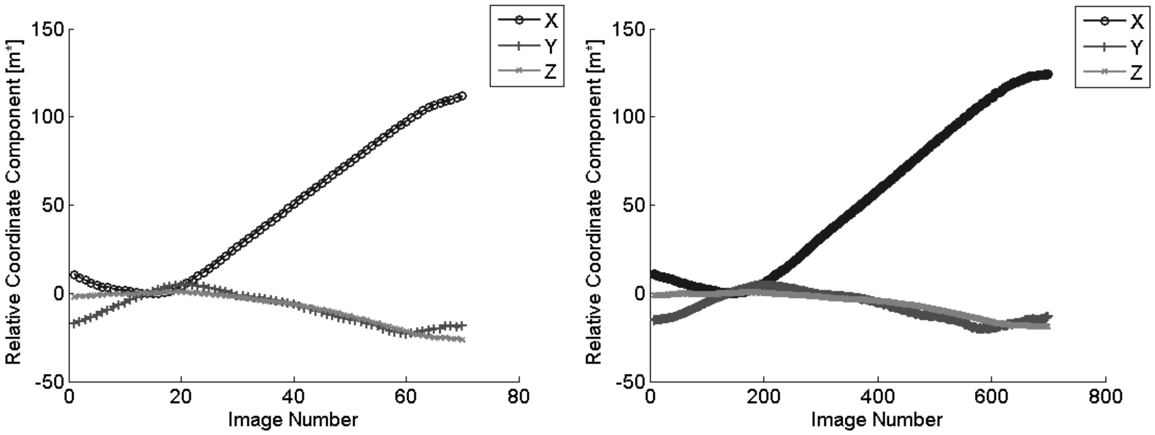

Fig. 5: Trajectory of sequence 1: position (left: bundle adjustment, right: Kalman filter)

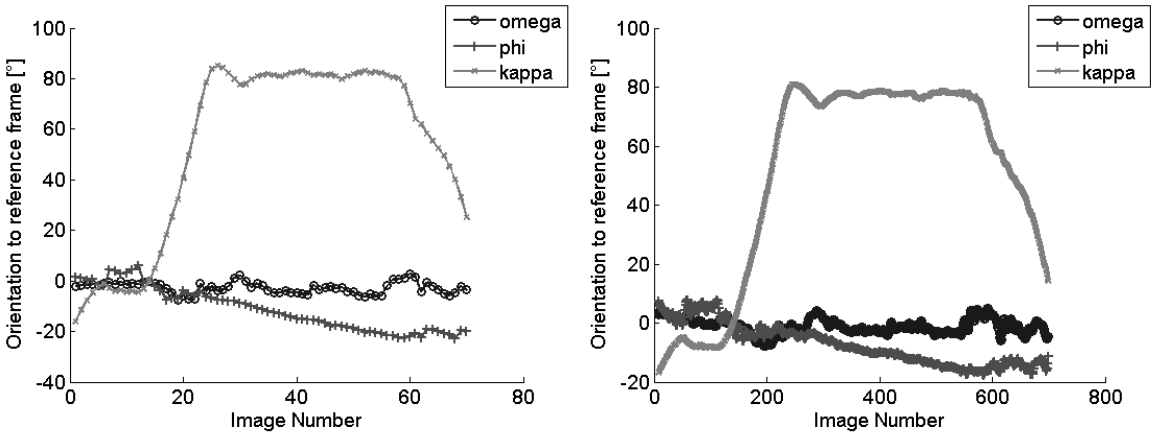

Fig. 6: Trajectory of sequence 1: angles (left: bundle adjustment, right: Kalman filter)

The overall reconstruction accuracy, namely histogram of the point error, is presented in fig. 7. The theoretical mean point error is in the range of 20 cm.

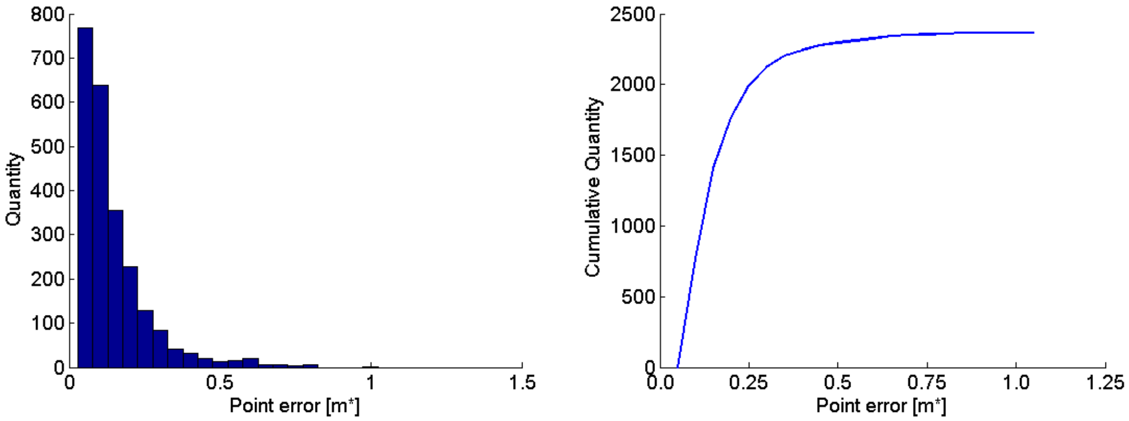

Fig. 7: Precision of sequence 1. Left: histogram. Right: cumulative histogram.
90 % of all point errors are below 30 cm, the average point error is 18 cm.

The point cloud of the 3D points for both image subsequences are shown in fig. 8. In the left subfigure the points on the roof of the building are clearly visible. In the right subfigure one can see the highly different Z-values of the trees in the middle of the figure, see the image shown in fig. 3.

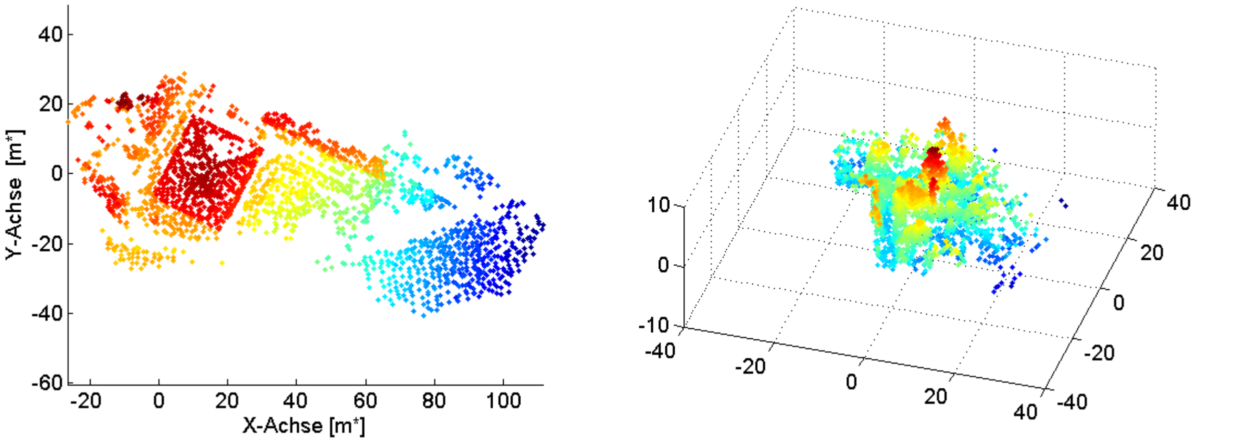

Fig. 8: Point cloud from AT (Z is color coded), Left: Point cloud of image subsequence 1 from bundle adjustment estimation as top view. Right: Point cloud of image subsequence 2 from bundle adjustment estimation as perspective view. Observe the points of the trees having very different height, see fig. 3.

3.2.2. Surface reconstruction

The bundle adjustment results for camera position and orientation can be use as input for the new version of MATCH-T, which can handle multiple images. We use all 70 key frames to derive a dense 3D point cloud with 120546 points. This corresponds to app. 10 points per m^2 and is comparable to the highest resolution obtainable with LIDAR systems. We compute a grid representation, interpolating holes in the point cloud with a hierarchical interpolation. The shaded 3D surface in fig. 9 demonstrates the high resolution achievable with low resolution images, even the rows of the vineyards can be identified.

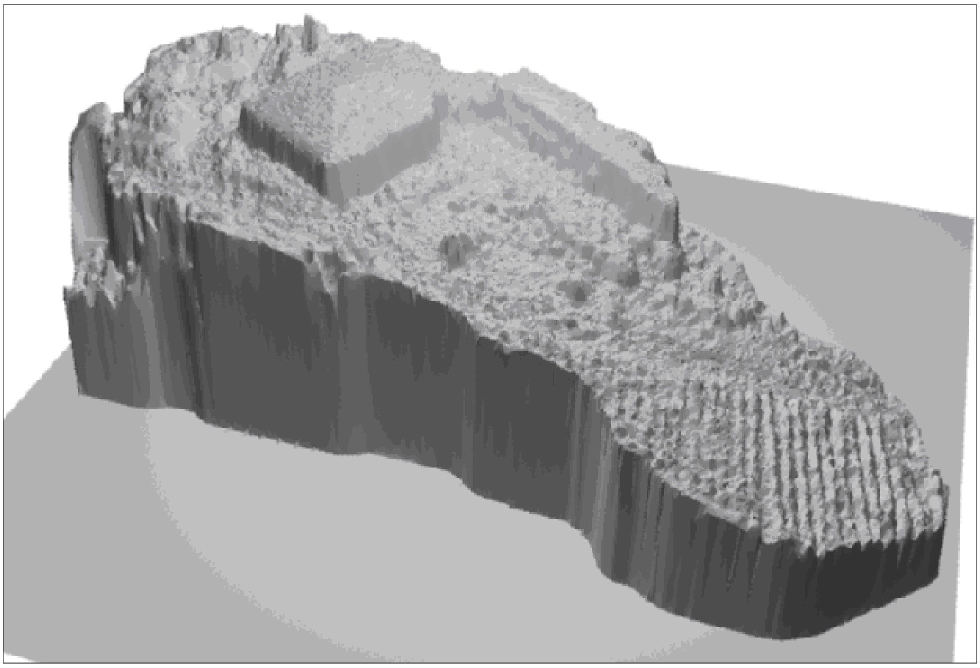

Fig. 9: Rendered surface from Match-T-point cloud

4. CONCLUSION

The paper discusses automatic methods for the evaluation of very large scale imagery acquired from new cheap sensor platforms. They open new fields of applications of photogrammetry. Geocoding can be realized using digital surface models derived from the images in case a reference DSM with enough local structure is available. Indoor application may also use the image content itself, which is not explicitely demonstrated here. Automatic orientation of image sequences can be reliably and robustly performed using tracking procedures for point transfer and batch or sequential estimation. The automatically derived 3D point clouds show a comparable density and quality as LIDAR data. High resolution DSMs may be derived from such low resolution image data automatically, providing a basis for simplified 3D data interpretation. Investigations into the accuracy and the speed of the methods are urgently necessary. However, the potential of the methods lie in the degree of user friendly automation, which could make very large scale photogrammetry a tool for day-to-day use of everyone.

5. ACKNOWLEDGEMENTS

We appretiate the support of Prof. A. Grimm, IGI Systems, T. Kanand, Microdrones GmbH and B. Weber for their support of the project.

6. REFERENCES

C. Beder and R. Steffen (2006). Determining an initial image pair for fixing the scale of a 3d reconstruction from an image sequence. In K. Franke, K.-R. Müller, B. Nickolay, and R. Schäfer, editors, Pattern Recognition, number 4174 in LNCS, p. 657-666. Springer.

A. J. Davison (2003). Real-time simultaneous localisation and mapping with a single camera. In Proceeding of the 9th International Conference on Computer Vision, p. 674-679.

C. Drewniok and K. Rohr (1997). Exterior Orientation -- An Automatic Approach Based on Fitting Analytic Landmark ModelsSpecial Theme Issue "Automatic Image Orientation", ISPRS J. of Photogrammetry & Remote Sensing *52*, p. 132-145.

H. Ebner, G. Strunz (1988). Combined point determination using Digital Terrain Models as control information, In: Int. Arch. of Photogrammetry and Remote Sensing, Vol. 27, Part B11, p. 578-587.

W. Förstner (1988). Model Based Detection and Location of Houses as Topographic Control Points in Digital Images. Int. Archives of Photogrammetry and Remote Sensing, Vol 27, B11, Kyoto.

A. Habib, M. Morgan, E.M. Kim, R. Cheng (2004). Linear features in phtogrammetric activities. Int. Arch. of Photogrammetry and Remote Sensing, Vol. 35, part B2, p. 610-615, Istanbul.

M. Heinrichs, O. Hellwich and V. Rodehorst (2007). Efficient Semi-Global Matching for Trinocular Stereo, Photogrammetric Image Analysis PIA'07, 2007, Munich.

R. E. Kalman (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, p. 35-45.

T. Läbe and W. Förstner (2005). Erfahrungen mit einem neuen vollautomatischen Verfahren zur Orientierung digitaler Bilder. In Proceedings of DGPF Conference Rostock, Germany, p. 21-23.

D. Lowe (2004). Distinctive image features from scale-invariant keypoints. In International Journal of Computer Vision, volume 20, p. 91-110.

M. I. A. Lourakis and A. A. Argyros (2004). The design and implementation of a generic sparse bundle adjustment software package based on the levenbergmarquardt algorithm. Technical Report 340, Institute of Computer Science- FORTH, Heraklion, Crete, Greece. Available from http://www.ics.forth.gr/~lourakis/sba.

E. M. Mikhail (1993). Linear features for photogrammetric restitution and object completion. in E. B. Barrett; D. M. McKeown (Eds.) Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision. Proc. SPIE Vol. 1944, p. 16-30.

J. Montiel, J. Civera, and A. Davison (2006). Unified inverse depth parametrization for monocular slam. In Proceedings of Robotics: Science and Systems, Philadelphia, USA.

F. C. Paderes, E. M. Mikhail, W. Förstner (1984). Rectification of Single and Multiple Frames of Satellite Scanner Imagery using Points and Edges ad Control. NASA Symp. On Mathematical Pattern Recognition and Image Analysis, Houston.

U. A. Rauhala, W. J. Mueller (1995). Feature entity least squares matching: a technique for the automatic control of imagery, in D. M. McKeown, I. J. Dowman; (Eds.) Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II, Proc. SPIE Vol. 2486, p. 60-73.

D. Scharstein, R. Szeliski (2002). A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, IJCV(47), No. 1-3, p. 7-42.

R. Steffen, C. Beder (2007). Recursive estimation with implicit constraints. Pattern Recognition, 29th DAGM Symposium, Heidelberg, Germany, Proceedings. Lecture Notes in Computer Science, Springer, to appear.

D. Ton and H. Mayer (2007). 3D least-squares-based surface reconstruction. , Photogrammetric Image Analysis PIA'07, 2007, Munich.

R. Wessel, M. Novotni, R. Klein (2006). Correspondences between Salient Points on 3D Shapes, in L. Kobbelt, T.Kuhlen, T. Aach, R. Westermann: Proceedings of Vision, Modeling, and Visualization, p. 365-372.

B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon (2000). Bundle adjustment – A modern synthesis. In W. Triggs, A. Zisserman, and R. Szeliski, editors, Vision Algorithms: Theory and Practice, LNCS, p. 298-375, Springer.