IMAGE-BASED NAVIGATION OF FOREST HARVESTERS

Marc Schulze

Institute of Photogrammetry and Remote Sensing
Technische Universität Dresden
Dresden, Helmholtzstrae 10, Germany
marc.schulze@tu-dresden.de
www.tu-dresden.de/geo/photo

Commission III/1

KEY WORDS: Photogrammetry, Forestry, Matching, Orientation, Point Cloud, Structure from Motion

ABSTRACT:

The focus of this paper the use of multi-image matching techniques in forestry applications. Background of the study is the problem of navigating heavy harvesters through skidder trails on their way to harvesting individual trees. Maneuvering these heavy vehicles over unprotected forest ground leads to irreversible soil compression and degradation effects. Therefore, harvester operators strive to navigate in a way that exactly the same (already compressed) path is used when they enter a skidder trail for a second time. For this task, vehicle navigation on a decimeter accuracy level is required. Data of existing techniques, such as GPS, IMU and/or odometry are error prone, because of difficulties like fluctuating signal strength of satellites caused by dense plant canopy, drift of IMU without update, and slippery, rough ground for wheel decoding. A camera, as a passive sensor, may avoid these problems, as it is largely independent to those outer influences.

ZUSAMMENFASSUNG:

Diese Arbeit beschreibt die Verwendung einer Mehrbildauswertung in Forstanwendungen. Der Hintergrund dieser Arbeit ist die Navigation schwerer Baumerntemaschinen entlang von Rückegassen zu alleinstehenden Bäumen. Diese Fahrt über ungeschützten Waldboden schädigt diesen durch irreparable Bodenverdichtung nachhaltig. Aus diesem Grund muss die Fahrspur eines bereits gefahrenen Pfades unbedingt eingehalten werden. Dies erfordert eine Fahrzeugnavigation im Dezimeterbereich. Die Daten bestehender Verfahren, wie GPS, IMU oder Odometrie, sind durch Signalverlust, Drift und rutschigen Boden sehr fehleranfällig. Eine Kamera, als passiver Sensor, könnte diese Probleme vermeiden, da sie grösstenteils unabhängig von äusseren Einflüssen ist.

1 INTRODUCTION

Modern forest industries need heavy and fast harvest machines to be economical. Although, vehicles are equipped with modern navigation systems, forest ground may be compressed and damaged irreversibly. GPS, IMU and odometry are error-prone, caused by dense plant canopy, IMU drift and slippery ground. Therefore, an alternative or an additional system should support existing devices. An image based approach might help to avoid these problems and increase the accuracy of localization. Captured images can be oriented by photogrammetric algorithms and deliver information about motion of used camera. In combination with GPS, the estimated motion track of harvesters could be transformed into a global coordinate system. Then, all subsequent forest machines may follow this more accurately tracked path to avoid irreversible soil compression beyond the predefined skidder trails.

The aim of this survey is to evaluate an image-based navigation for forest harvesters. In a first step, images have been captured manually along a nearly straight path within a forest to simulate a harvester path without vibration and much motion changes. Fast and reliable feature detectors like SIFT and SURF have been used to find tie points in overlapping images areas. Further, tie points have been used to estimate the relative orientation between all images and to calculate their outer orientation. As a result, the driven track can be visualized as a 3D trajectory, when the images taken along the path are processed in the corresponding order.

2 RELATED WORK

State-of-the-art forest navigation is based only on GPS and delivers an accuracy about 5 to 10m (Hamberger, 2001), caused by shielding of plant canopy. Vision based approaches might aid error-prone systems, like GPS and IMU, to navigate in spite of signal loss or drift effects. Requirements for that kind of navigation are successively captured images with overlapping regions. Image position and orientation can be reconstructed by relative orientation based on feature points. Feature detectors such as SIFT (Scale Invariant Feature Transform) (Lowe, 2004) or SURF (Speeded Up Robust Features) (Bay et al., 2006), are scale and rotation invariant and adequate for tie point detection. Structure from Motion (SfM) software packages, like Bundler (Snavely et al., 2007), take unordered sets of images and produce 3D reconstructions of all camera positions and scene geometry. SfM tools mostly use a sparse bundle adjustment to compute the results and can therefore be used only in retrospect. Another approach reported in (Davison, 2003) offers a real time application for vision based simultaneous localization and mapping (VSLAM). Hence, it is even possible to reconstruct a path with a monocular device.

3 SENSOR AND DATA

For first experiments, manual images have been captured with the NIKON D700 and a NIKON NIKKOR 20mm wide-angle lens, as shown in figure 1 and summarized in table 1). Shutter time and aperture have been chosen so that motion blur is avoided. The

pictures have been taken in full resolution of 4256 x 2832 pixel, but had to scaled down to a smaller size (see section 4) to comply with requirements of feature detection algorithms.

(a) NIKON D700   (b) NIKON NIKKOR 20mm

Figure 1: Used camera and lens

parameter	value
resolution	4256 x 2832 pixel
sensor	24 x 36 mm CMOS chip
focal length	20mm

Table 1: Data of camera and lens

4 METHODOLOGY

In this experiments two different approaches of pose estimation and 3D data acquisition in woodland with relative orientation have been evaluated. Basically, known structure form motion algorithms have been utilized, but here they are employed in the context of forest environments, which previously did not receive adequate attention. A REAL-TIME-PROCESS estimates the orientation of two subsequent images with SURF (Bay et al., 2006) or SIFT (Lowe, 2004) feature points and calculates the parameters of motion (X,Y,Z,ω,ϕ,κ) between this sequence. As a second approach, POST-PROCESSING calculates the position of images with SIFT features in a combined orientation and bundle adjustment (Bundler (Snavely et al., 2007) and (Lourakis and Argyros, 2009)). This bundle adjustment improves position and orientation of each image in sequence by more redundancy based on by more observations and distributing residual errors in equal parts on each point of view.

Figure 2: Example image of captured stack with 36 frames overall

4.1 Post Processing Approach

A post-processing algorithm can be employ to compute the camera positions and orientations of images set after capturing over-

all. Further, some additional algorithms use those orientated images and their positions to estimate a dense point cloud of the surrounding environment. The post-processing algorithm uses the Bundler software (Snavely et al., 2007) for image orientation. Therefore, a stack of images with overlapping areas is required. Global unique feature points inside of covered regions are necessary at first. These are located by a SIFT feature detector to get a robust feature descriptor for this position. The SIFT feature descriptor specifies a global unique feature and can be compared with feature points of different images of the stack. An approximate nearest neighbors (ANN) matching method among different features delivers in combination with computed fundamental matrix and RANSAC (Random Sample consensus) (Fischler and Bolles, 1981) a robust estimator for parameters of relative orientations. Furthermore, the bundle adjustment connects all detected homologue feature points and approximated focal length to reduce the error of observations using a least squares method based on Gauss-Markov-Model. In addition to camera position and 3D points, the software computes focal length and distortion, thus it is not necessary to calibrate the camera before using. Results of Bundler can be imported into PMVS2 (Patch-based Multi-view Stereo) software (Furukawa and Ponce, 2010) to create a dense point cloud of environment. "PMVS is a multi-view stereo software that takes a set of images and camera parameters, then reconstructs 3D structure of an object or a scene visible in the images." (Furukawa, 2010)

4.2 Real Time Approach

A post processed localization and mapping assists in the estimation of the driven path and computation of a 3D point cloud of the environment after image acquisition. Nevertheless, in some cases a real time analysis is required, as for instance a simultaneous localization and driving of forest vehicles. Such a method could help restraining the forest machine to a given track and consequently avoiding further soil compression. In comparison to the approach of section 4.1, this real time algorithm computes the orientation of images sequentially. That means, the pose estimation is computed gradually, from image to image, and the actual position is accumulated by previous image positions. First of all feature points are detected and feature descriptors are computed. Following descriptors are matched to obtain homologous point pairs in an overlapping image pair. In order to improve the speed of this algorithm, SURF features are detected instead of SIFT. RANSAC based on a computed fundamental matrix, helps to find the best solution for orientation. Because of the over determined equation system, an additional adjustment of the relative orientation is performed and uses all inliers determined with RANSAC and increases the accuracy of parameters by minimizing the square of residuals. The number of tie points between two images is essential to compute an accurate relative orientation. Tests with images, that contain trees, have shown that the number of corresponding points strongly depends on the resolution of the image (Fig. 3). The reason for this behavior is founded in the computed feature descriptor size and the fast background color changing of features. So it may helpful to resample the image to a smaller size (table 2). In this approach, image coordinates are direct observations, so a calibration of the camera is necessary and feature point positions have to be corrected for lens distortion and principle point shift.

5 RESULTS AND DISCUSSION

The results of post (Sec. 4.1) and real time processing (Sec. 4.2) are very different. Reasons for the different results are due to the different ways of adjustment. Bundler uses a sparse bundle adjustment to enhance all extracted observations and approximated
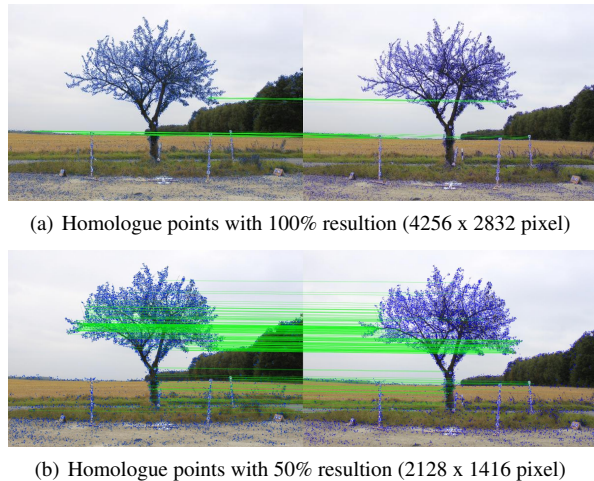
(a) Homologue points with 100% resultion (4256 x 2832 pixel)

(b) Homologue points with 50% resultion (2128 x 1416 pixel)

Figure 3: Different number of feature points, due to different resolution

image resolution	number of corresponding feature points
100%	61
75%	167
50%	197
25%	97

Table 2: Number of corresponding feature points with different image resolutions. Respective image examples Fig. 3

camera parameters. Outliers are detected and eliminated and all residuals are minimized by the method of least squares. As a result, generated errors are distributed in equal parts to each camera position and existing point of views exhibit only small errors. Bundler, utilized in the post processing method, uses a sparse bundle adjustment. In compare to the sequentially working real time approach, the estimated camera position are more accurate, but the analysis can only start after the entire whole image recording is completed. The real time algorithm directly estimates the position and orientation after capturing a single image, but the error is increasing from image to image because of variance propagation.

5.1 Post Processing Approach

Unfortunately, Bundler provides no accuracy results of obtained calculations. Therefore, an external bundle adjustment, which imports the results of Bundler (3D points, 2D points, image orientation) has been used to compute the results with accuracy details again(see table 3). All object points from Bundler are employed as control points in this bundle adjustment and previously obtained feature points are observations; therefore it is more of a multi space resection adjustment.

| parameter	mean standard deviation for 36 frames
X	1.4mm
Y	1.6mm
Z	1.8mm
XYZ	2.8mm
ω	0.010°
ϕ	0.012°
κ	0.004°

Table 3: Mean accuracy of camera orientation from Bundler

This computation achieved results based on a excellent image acquisition without any interruptions or difficulties. That means there was no interruption of image sequence, no obstacles, con-

stant velocity and without strong wind. Thus, this mentioned results are very optimistic but also possible. Bundler automatically computes a high accurate pose estimation based on a small number of parameters (focal length and image size). It detects natural feature points with an accuracy below one pixel in image space and therefore it can employed in forest applications as well. The obtained results are encouraging and demonstrate that the post processing method is indeed feasible. PMVS provides a patch-based algorithm that uses the results from Bundler to create a dense point cloud of environment. Based on a extended multi-view patch correlation, PMVS increases the number of points to four times in compare to Bundler with given image sequence of 36 frames.
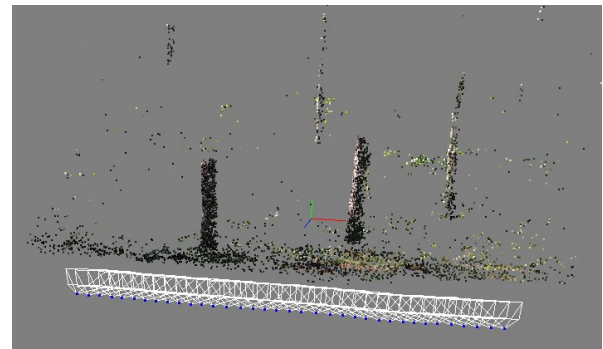
(a) Bundler results; camera position and orientation and computed 3D points
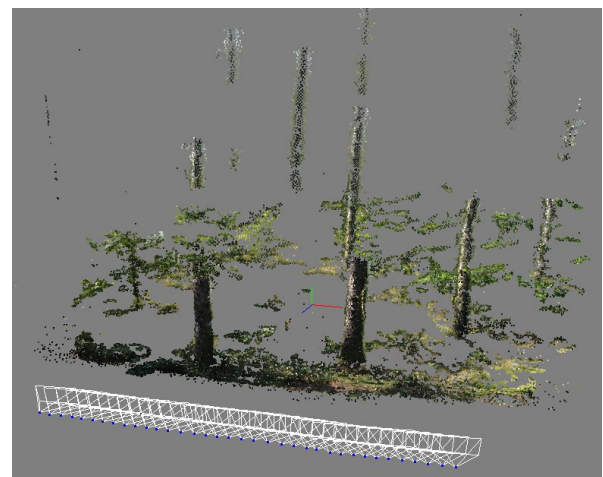
(b) PMVS2 results; dense point cloud

Figure 4: results of Bundler and PMVS2

5.2 Real Time Approach

As presented in table 4, the mean accuracy is comparable with the post processing method (section 5.1). However, illustrated results are just the accuracy values for one image pair. Because of variance propagation, the uncertainty of all following images increases. It is noticeable, that rotation values (ω,ϕ,κ) are worst in comparison to translation. The basic cause for this effect is a clustered tie point allocation: In some images, homologous points are very close together and concentrated in only one region of the image. As a result the rotation parameters can not be computed with the expected accuracy.

To avoid the problem of variance propagation, successive image have to be connected to each other. Detected and stored feature points can be used to obtain the relative orientation between two or more images. In that way, preceding images could be linked with current images to stabilized pose estimation. One possible

parameter	mean standard deviation for 36 frames
Y	0.04mm
Z	0.13mm
ω	0.04°
ϕ	0.19°
κ	0.06°

Table 4: Mean accuracy of relative orientation between one image pair

implementation is discussed in (Davison, 2003) and (Davison et al., 2007). It is an advanced model of VSLAM with an Extended Kalman Filter (EKF). At the moment, our work does not support a prediction and SLAM algorithm. Nevertheless first results and existing works demonstrate the potential of this method, which we strive to develop further.

6 CONCLUSIONS AND FUTURE WORK

First results of this pilot study show the feasibility of theoretical background and it is worthwhile to pursue this approach further. A vision based method can estimates a position of a forest harvester with an adequate accuracy. Especially the post-processing approach delivers stable and precise results. The Pose estimation results below one cm in experiments using post and real time processing exceed the expectations and form the basis for further research. Furthermore, the achieved results meet accuracy requirements of about 15cm, i.e. half of a tire width, for forest navigation to avoid soil compression. Influences on accuracy, such as strong wind and fast movements, have not been analyzed yet. The calculated point cloud of Bundler and PMVS2 can be used to create a virtual forest. Current approaches for automatic tree detection ((Heurich and Weinacker, 2004),(Schilling et al., 2011)) for laser scanner data can be applied on this data. In this context, another possible navigational aid could be a virtual forest model in which cylinders have been fitted to appropriate point cloud subregions. Assuming all trees of this virtual forest have at least 2D coordinates, a camera on a harvester and a corresponding software have to solve the assignment of trees in object and image space and could compute their position by spatial resection.

REFERENCES

Bay, H., Tuytelaars, T. and van Gool, L., 2006. Surf: Speeded up robust features. Technical report, ETH Zurich.

Davison, A., 2003. Real-time simultaneous localisation and mapping with a single camera. In: Proc. International Conference on Computer Vision, Nice.

Davison, A. J., Reid, I. D., Molton, N. D. and Stasse, O., 2007. Monoslam: Real-time single camera slam. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, pp. 1052–1067.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24(6), pp. 381–395.

Furukawa, Y., 2010. Patch-based multiview stereo software (pmvs - version 2). http://grail.cs.washington.edu/software/pmvs/.

Furukawa, Y. and Ponce, J., 2010. Accurate, dense, and robust multi-view stereopsis. IEEE Trans. on Pattern Analysis and Machine Intelligence 32(8), pp. 1362–1376.

Hamberger, J., 2001. GPS als Mittel zum umweltschonenden Maschineneinsatz: Navigation von Forstmaschinen und Dokumentation ihrer Fahrbewegung. PhD thesis, Technische Universitt Mnchen.

Heurich, M. and Weinacker, H., 2004. Automated tree detection and measurement in temperate forests of central europe using laserscanning data. Int. Arch. Photogramm. Remote Sens. XXXVI, pp. 198–203.

Lourakis, M. A. and Argyros, A., 2009. Sba: A software package for generic sparse bundle adjustment. ACM Trans. Math. Software 36(1), pp. 1–30.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, pp. 91–110.

Schilling, A., Schmidt, A. and Maas, H.-G., 2011. Automatic tree detection and diameter estimation in terrestrial laser scanner point clouds. In: 16th Computer Vision Winter Workshop.

Snavely, N., Seitz, S. M. and Szeliski, R., 2007. Modeling the world from internet photo collections. International Journal of Computer Vision 80, pp. 189–210.