

ON VISUAL REAL TIME MAPPING FOR UNMANNED AERIAL VEHICLES

Richard Steffen and Wolfgang Förstner

Department of Photogrammetry
University of Bonn
rsteffen@uni-bonn.de, wf@ipb.uni-bonn.de

Commission III

KEY WORDS: orientation, reconstruction, image sequence, real time, UAV

ABSTRACT:

This paper addresses the challenge of a real-time capable vision system in the task of trajectory and surface reconstruction by aerial image sequences. The goal is to present the design, methods and strategies of a real-time capable vision system solving the mapping task for secure navigation of small UAVs with a single camera. This includes the estimation process, map representation, initialization processes, loop closing detection and exploration strategies. The estimation process is based on the Kalman-Filter and a landmark based map representation. We introduce a new initialization method for new observed landmarks. We will show that the initialization process and the exploration strategy has a significant effect on the accuracy of the estimated camera trajectory and of the map.

1 INTRODUCTION

Unmanned Aerial Vehicles (UAVs) as low cost and flexible platforms for image acquisition become more and more important. Monitoring changes in agriculture, inspection of buildings, the wide field of surveillance and documentation in archeology are only a few civil applications using UAVs.

We can distinguish between high-altitude systems such as presented in the PEGASUS Project (Everaerts et al., 2004) and low-altitude systems conducted by small helicopters or quad-copters. Compared to high-altitude systems, low-altitude systems have the disadvantage of a small range and a small payload capacity. Both systems require position and orientation information for the purpose of navigation.

The position of the UAV is usually determined by a combination of GPS and INS (Inertial Navigation Systems) measurements, in which GPS will account for the long term instability of the inertial sensors. Moreover, GPS is not available in indoor environments and cannot measure directly the orientation of the UAV or can be even jammed. Considering the small payload capacity of small UAV systems and the fact that GPS and INS measurements might not be available, the use of localization techniques from image sequences can be very beneficial.

This paper addresses the problem of the localization and mapping for navigation purposes using monocular vision only. There is no focus on a dense reconstruction of an elevation map in this paper. This can be performed by a post processing step of the image data as we will show at the end of this paper. The described techniques are designed for real time processing.

The technical part of the paper addresses the problem of initialization of new points within the Kalman-Filter: The classical method of initializing a point based on a measured ray and an approximate distance has shown to give poor results. We propose a new solution based on the condition of the achieved triangulation accuracy.

The paper is organized as follows: First we will give a short overview about our available UAV hardware and the meaning of real-time capability in our sense. We will discuss the state of the art estimation approaches for localization and mapping and their commonly used map representation. In the next section we will briefly describe our Kalman-Filter based model. Furthermore we

will explain our stable initialization solution of new landmark coordinates. Regarding the exploration strategy a spiral path is presented as well as an idea for an efficient loop closing detection. In an experimental section we discuss the influence of a bad initialization and the benefit of our solution. We also present results based on simulated data for long strips and for a spiral path exploration. Finally we will show first results for trajectory determination and surface reconstruction based on real data. An outlook to future works will conclude this paper.

2 OVERVIEW

In this section we present a short overview about our available UAV and vision based localization and mapping methods in the context of real time possibility.

2.1 UAV Hardware

The real data experiment shown below is 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 of payload. This drone is equipped with a Panasonic Lumix camera 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 a quick time movie.



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

2.2 On the Notion of Real Time

In our perception, real time means the capability to react in a required time on sensor data. The results of the computation of the sensor data should be available within a reasonable time. In

our task this is the time for secure navigation and for instance a surveillance function.

But not all parts of the calculations need the same time. For example, secure navigation demands a local accurate map with an update of 1 Hz. The absolute position can be used for path planning and close loop detection and is not necessary for every second.

2.3 Vision based Navigation

Vision based navigation or simultaneous localization and mapping becomes a very important research field in robotics for the last years. These techniques can be used almost anywhere, for instance on ground vehicles, in a subsea environment or in the air. Depending on the carrier platform and the navigation requirements different vision hardware systems are applied.

Conceptual we distinguish between monocular (Davison, 2003) and stereo or multi-stereo vision concepts (Agrawal et al., 2007). In the case of monocular vision there is no direct depth estimation possible. Assuming a motion of the camera in a static world, depth information can be achieved. In the case of stereo vision we can estimate the object coordinates at any point of time. Furthermore a time based detection of motion in the object space is also possible (Franke et al., 2005).

Furthermore, we have to decide whether a local or a global map should be obtained. Local maps do not require advanced map building techniques (Estrada et al., 2005), but they are only useful for safe navigation in small environments, not for navigation in the context of path finding. The representation and the update of global maps is a challenge depending on the real time capability of the estimation processes and an intelligent memory storage.

Commonly used techniques to estimate the exterior camera orientation are based on the identification of landmarks and the estimation of their coordinates. They presume an exact data association between measurements at different viewpoints. This is done by interest points used as landmarks which are usually tracked through the image sequence by a KLT-Tracker (Tomasi and Kanade, 1991) or by descriptor matching (Lowe, 2003). The map is represented by the landmark coordinates and their uncertainties, as well as their correlations between different landmarks.

Various techniques are available to estimate the trajectory and the map. The well known bundle adjustment is one of the traditional algorithms. In the past years, bundle adjustment has become famous in the robotic research community often referred as *information filter* (Thrun et al., 2005). By using the sparseness and solving the linear normal equation system according to a factorization technique (Grün, 1982), (Triggs et al., 2000), the complexity is nearly equivalent to the well known Kalman-Filter based approaches. But, the elimination of needless parameters from the equation system using *schur complement* is time consuming with respect to computation.

The Kalman-Filter (Kalman, 1960) updates the unknown parameters, stored in the so called state vector, in an iterative manner at every time step. The complete information about the map and the trajectory is stored in the state vector and their covariance matrix. Needless parameters can be eliminated by discarding them from the covariance matrix and from the state vector. The accuracy of the Kalman-Filter depends on the accurate parameter initialization. This effect can be reduced by using an inverse depth representation for object points (Montiel et al., 2006), but is very difficult to handle. We will introduce an alternative solution on this problem later in this paper.

Both, the Kalman-Filter approach and the bundle adjustment technique take into account that the map and trajectory parameters are linked together. The *FastSLAM* (Montemerlo et al., 2002)

approach factorizes the full localization and mapping problem exactly into a product of a robot path posterior, and landmark posteriors conditioned on the robot path estimate. This factored posterior can be approximated efficiently using a particle filter. The landmark positions can be estimated independently by a set of small Kalman-Filters for every path hypothesis. This technique has a high memory consumption, but can deal with up to 100.000 landmarks.

3 ALGORITHMS AND STRATEGIES

3.1 Pose estimation with a Kalman Filter

As described in section 2.3 the parameter estimation for the position and the orientation as well as the object coordinates is a core module for navigation. We use a Kalman-Filter based approach for this task. The interior camera parameters are assumed to be known.

We use the modeling according to (Davison, 2003). The state vector \mathbf{p}^t contains the camera parameters, their velocities and the 3D-points:

$$\mathbf{p}^t = \begin{bmatrix} \mathbf{r} \\ \mathbf{q} \\ \mathbf{v} \\ \boldsymbol{\omega} \\ \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{N_t} \end{bmatrix}^t \quad (1)$$

The uncertainty is coded in the covariance matrix Σ_{pp}^t . The camera trajectory is represented by its actual position \mathbf{r}^t , its orientation quaternion \mathbf{q}^t , its velocity vector \mathbf{v}^t and its angular velocity vector $\boldsymbol{\omega}^t$. The 3d point coordinates are represented by their N_t Euclidean points \mathbf{X}_i^t . The list of points will vary over time, requiring elimination and inclusion of points at possibly every time step.

At the moment, our approach uses the same camera and structure representation. We assume a linear time update model, which can easily be computed by

$$\mathbf{p}^{t+1} = \begin{bmatrix} \mathbf{r}^{t+1} \\ \mathbf{q}^{t+1} \\ \dots \end{bmatrix} = \begin{bmatrix} \mathbf{r}^t + \mathbf{v}^t \Delta t \\ \Delta \mathbf{q}^t \circ \mathbf{q}^t \\ \dots \end{bmatrix} + \boldsymbol{\eta}^t \quad (2)$$

with $\Delta \mathbf{q}^t = (1, \boldsymbol{\omega}^t \Delta t)$ and \circ representing the quaternion multiplication. Velocity, angular velocity and Euclidean points are assumed not to change over time.

The measurement equations are based on the co-linearity equations, which can be written as homogeneous equations

$$\mathbf{x}_i^t = \lambda_i^t \mathbf{P}^t \mathbf{X}_i^t \quad \text{with} \quad \mathbf{P}^t = \mathbf{KR}(\mathbf{q}^t) [I_3 | -\mathbf{r}^t] \quad (3)$$

indicating, that we assume the calibration of the camera not to change over time.

3.2 Initialization Procedure

Each Kalman-Filter needs to be initialized. In the context of simultaneous localization and mapping at each time step new points may be included into the state vector and therefore then need to be initialized.

Initial values for the camera position and orientation and their velocities have to be introduced. The complete orientation of the object coordinate system can be derived from a small set of available control points. These points need to be introduced into the Kalman-Filter in order to define the scale of the complete reconstruction.

Initial values for new points have to be provided continuously. This turns out to be a problem, as bad initializations of 3D-points may prevent the system from convergence.

One way for the initialization of points is to calculate the Euclidean coordinates on the first visible projection ray presuming an a priori distance with a large covariance matrix and from thereon update the point in the Kalman-Filter. But already the intersection with the next ray, which due to the small parallax, leads to an uncertainty which cannot really be modeled using a normal distribution as shown in (Montiel et al., 2006) and therefore lead to biases in the estimation.

This problem can be circumvented by representing the 3D-point using the inverse depth, which intuitively corresponds to the parallax and therefore is well represented by a normal distribution. This inverse representation usually requires six parameters as the 3D-point needs to be linked to the first camera the points is visible. After a stabilization of the estimated coordinate it is necessary to transform the inverse representation to an Euclidean representation, in order to keep the number of parameters in the state vector small. This procedure turns out to be cumbersome.

We here propose, to solve the problem of biased estimation in a different manner. We suggest postponing the initialization until the geometry guarantees the 3D-point is stable. This requires to store for each 3D-point \mathbf{X}_n , which is observed in frame t for the first time, the orientation parameters $\mathbf{r}^t, \mathbf{q}^t$ and the measured image coordinate \mathbf{x}_n^t . As the initialization need not to be perfect in terms of accuracy, we do not need to introduce these data into the Kalman-Filter at later stages $> t$.

Each new point is tracked leading to a sequence of points $\{\mathbf{x}_n^{t'}, t' = t + 1, \dots\}$ in the following frames. At each time τ we select that pair of frames (t', τ) for determining the 3D-point where the length of the basis $|\mathbf{r}^\tau - \mathbf{r}^{t'}|$ is largest, as we expect the determination of the 3D-point to be most stable. If the covariance matrix of the 3d-point is round enough, we initialize the 3D-point and include into the Kalman-Filter. The roundness R of the covariance matrix $\Sigma_{\mathbf{X}_n \mathbf{X}_n}$ is determined from the eigenvalues using $\sqrt{\lambda_3/\lambda_1}$, indicating the maximum ratio of uncertainty at that point, c. f. (Beder and Steffen, 2006). We use as threshold of $T_R = 0.1$ for this ratio, i. e. new points are guaranteed to fulfill $R > T_R$.

The covariance matrix can be determined assuming the orientation parameters to be error free. In our experiments the triangulation and the error propagation is performed using the unscented transformation (Julier and Uhlmann, 1997). Therefore the precision is over estimated and, what is even more severe, the correlations between the new point and the state of the system are zero. Thus a blind inclusion of the point together with its covariance matrix, derived this way, would make this new point acting like a control point. For this reason we need to couple the uncertainty of the parameters of the state vector and the uncertainty of the new point. The idea is to determine the new point \mathbf{X}_n via a point \mathbf{X}_r which is close to the new point via:

$$\mathbf{X}_n = \mathbf{X}_r + \Delta \mathbf{X}_n \quad (4)$$

and perform error propagation assuming the given covariance matrix of \mathbf{X}_r and its correlation with all parameters of the state vector. However, the coordinate difference $\Delta \mathbf{X}_n = \mathbf{X}_n - \mathbf{X}_r$ has covariance matrix which in the worst case is

$$\Sigma_{\Delta \mathbf{X}_n \Delta \mathbf{X}_n} = 2 \Sigma_{\mathbf{X}_n \mathbf{X}_n}$$

and is assumed to be independent of \mathbf{X}_r with. This is a realistic assumption, as the reference point would probably have a similar accuracy in case of triangulated from the same two images.

Thus we obtain the extension of the state vector in the following manner

$$\mathbf{p}^{\text{ext}} = \begin{bmatrix} \mathbf{p}_- \\ \mathbf{X}_r \\ \mathbf{X}_n \end{bmatrix} = \underbrace{\begin{bmatrix} I \\ I_3 \\ I_3 \quad I_3 \end{bmatrix}}_A \underbrace{\begin{bmatrix} \mathbf{p}_- \\ \mathbf{X}_r \\ \Delta \mathbf{X}_n \end{bmatrix}}_g \quad (5)$$

where \mathbf{p}_- is the state vector without point \mathbf{X}_r . The generating vector has the covariance matrix

$$\Sigma_{gg} = \begin{bmatrix} \Sigma_{\mathbf{p}_- \mathbf{p}_-} & \Sigma_{\mathbf{p}_- \mathbf{r}} & 0 \\ \Sigma_{\mathbf{r} \mathbf{p}_-} & \Sigma_{\mathbf{r} \mathbf{r}} & 0 \\ 0 & 0 & \Sigma_{\Delta \mathbf{X}_n \Delta \mathbf{X}_n} \end{bmatrix} \quad (6)$$

The new covariance matrix of the state vector including the new landmark can now be calculated as follows

$$\Sigma_{pp}^{\text{ext}} = A \Sigma_{gg} A^T = \begin{bmatrix} \Sigma_{\mathbf{p}_- \mathbf{p}_-} & \Sigma_{\mathbf{p}_- \mathbf{r}} & \Sigma_{\mathbf{p}_- \mathbf{r}} \\ \Sigma_{\mathbf{r} \mathbf{p}_-} & \Sigma_{\mathbf{r} \mathbf{r}} & \Sigma_{\mathbf{r} \mathbf{r}} \\ \Sigma_{\mathbf{r} \mathbf{p}_-} & \Sigma_{\mathbf{r} \mathbf{r}} & \Sigma_{\mathbf{r} \mathbf{r}} + \Sigma_{\Delta \mathbf{X}_n \Delta \mathbf{X}_n} \end{bmatrix}. \quad (7)$$

This covariance matrix is guaranteed to be positive semidefinite and reflects the new situation realistically. Obviously, the linear transformation is sparse and except for the addition of the covariance matrices of the reference and the point difference only copies need to be realized, which is extremely efficient.

The algorithm for initialization of new points thus can be summarized as follows:

1. Detecting that a new point \mathbf{x}_n^t is observed in image t .
2. Set $\tau = t + 1$.
3. Determine whether this frame τ leads to a stable triangulation:
 - (a) Determine the frame k with the longest base length $k = \text{argmax}_{t' \in \{t, \dots, \tau-1\}} (|\mathbf{r}^\tau - \mathbf{r}^{t'}|)$.
 - (b) Determine \mathbf{X}_n and its covariance matrix using frame pair (k, τ) and its roundness $R_{k\tau} = \sqrt{\lambda_3/\lambda_1}$.
 - (c) If $R_{k\tau} > T_R$ then stop, else increase τ by one and go to 3.
4. Search for the point \mathbf{X}_r closest to \mathbf{X}_n .
5. Include the new point into the state vector using (5) and (7).

The proposed algorithm is able to initialize new points in a stable manner considering the correlation structure. In section 4 we will show the feasibility of this approach.

3.3 Exploration Strategy

A common application using UAVs is the exploration of inaccessible or hazardous areas. In the classical aerial photogrammetry overlapping strips are flown. This is satisfactory to estimate the trajectory and the surface, usually done by a global optimal bundle adjustment. As we know, the standard deviations $\sigma_r(k)$ in position of a long strip with k images is growing with (Ackermann, 1965)

$$\sigma_r(k) \propto \sqrt{k^3}. \quad (8)$$

However, the estimation process using a Kalman-Filter based approach accumulates errors from image to image. Therefore, it is necessary that the actual observations are connected to already stable estimated landmarks.

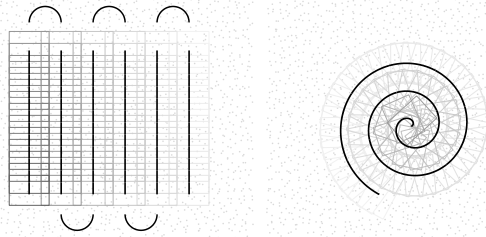


Figure 2: Left: classical area exploration by overlapping strips. Right: overlapping images using Archimedic spiral exploration strategy.

On the right side of figure 2 we show the top view of the camera path following the Archimedic spiral. In case of the Archimedic spiral the footprint of the images are overlapped depending on the distance of the arms of the spiral. The major advantage of this trajectory is that a connection of stable estimated landmarks is ensured every time.

3.4 Search Loop Closing Strategy

The identification of already observed landmarks is an important task. This is usually called as the loop closing problem. The re-identification of landmarks is often performed by image descriptor based matching, such as SIFT descriptors (Lowe, 2003), (Mikolajczyk and Schmid, 2005). For large maps the one by one matching computation is expensive and similar structures might result in wrong matches. Our key idea is to scale down the search space by using the actual reconstruction result. For new observed landmarks, SIFT descriptors in different scales can be computed by using the actual image data. The landmark can be referenced by its coordinate in an oc-tree, which divides the 3d space in small pieces. In order to detect a loop closing we have to match the descriptor for a new landmark with the referenced descriptors in the oc-tree buckets of the corresponding landmark coordinate with respect to the reconstruction accuracy. To improve the reliability we have to check the correct matching with the reconstructed neighbor landmarks for consistency.

4 EXPERIMENTAL RESULTS

To evaluate our approaches and strategies we have implemented them into a simulation environment. Furthermore we compare the Kalman-Filter approach with a bundle adjustment on a small image sequence based on real data.

4.1 Simulation Environment

Our simulation scenario has the following setup. The camera has a resolution of 640×480 pixel and a principle distance c of 640 pixel, which results in a field of view of app. 53° . The altitude over a plane surface is 50 m, the flight velocity is 5 [m/s] and the image sequence is acquired with 25 [fps]. We evaluate our algorithm with a 300 m strip and a 900 m spiral path. True point coordinates X_i are computed by a randomized sampler. The observations x_i^t are determined with the camera parameters and additive white noise of 0.5 [pel]. To compare our results with true camera orientations, the initialization of the camera orientations and the initial landmark coordinates for the first frame are set to the true value with variance zero. The system noise in the time update is set to 2 [m/s²] for accelerations and 20 [°/s²] for angular accelerations.

Variance structure of the strip. In a first experiment we will investigate, whether the general rule (8) can be confirmed by our implementation. Any deviation from (8) indicates systematic errors in the data or the implementation, e.g. in case the prior information is too strong. This can be done by initializing the new landmarks based on ground truth data with a large covariance matrix and the observations of the landmarks are considered to be error free. Figure 3 illustrates that the graphs satisfy (8).

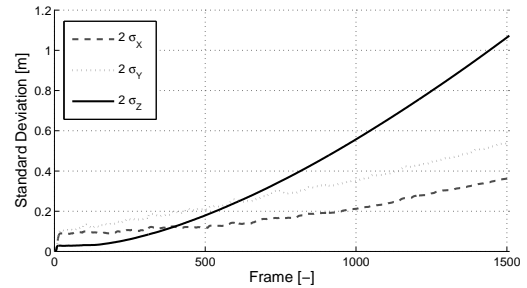


Figure 3: Theoretical precision of the positions r^t determined from the Kalman-Filter. The small roughness of the curves result from the randomness of the number of points in the different frames.

Comparison of the initialization procedures. The second experiment proves the benefit of our initialization method. Figure 4 shows the estimated camera projection centers by initializing new points on the first visible projection ray with 20 m distance. The estimated camera projection center quickly shows a huge drift of the flying height as shown in the Z component and a resulting drift in the Y component along the flight path. An initialization using the true distances shows similar effects.

The results of our new initialization method, c. f. figure 5, is much more stable. Again, we have a drift in the Z and Y component, but only 1 % of the strip length.

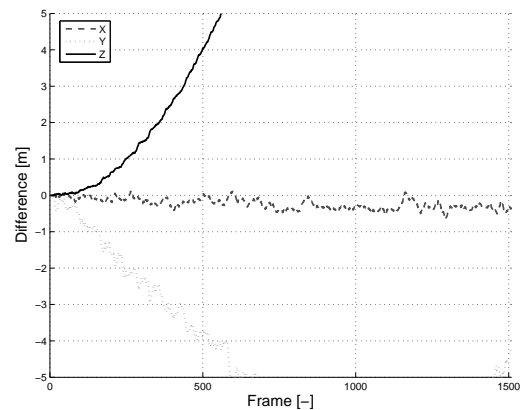


Figure 4: Estimated projection centers for a long strip by a Kalman-Filter with initialization of new observed points on the first projection ray, 20 m distance and large covariance matrix.

Comparing figure 6 with 3 we observe a decay of the estimated variances of the camera trajectory. We assume that this results from the approximation of the new introduced point correlation structure.

Evaluation of the exploration strategy. In a third experiment we calculate the camera trajectory based on the mentioned spiral path strategy, c. f. figure 2, with an overlapping of 30%. There are no systematic errors in the estimated projection centers of the camera as shown in figures 7 and 8. In addition, the trajectory is 3 times longer than the strip path.

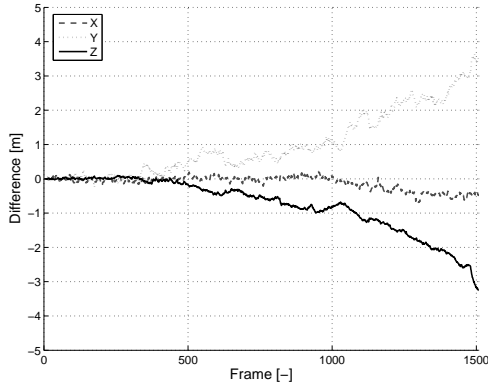


Figure 5: Estimated projection centers for a long strip by a Kalman-Filter and initialization of new observed points with our proposed stable initialization method.

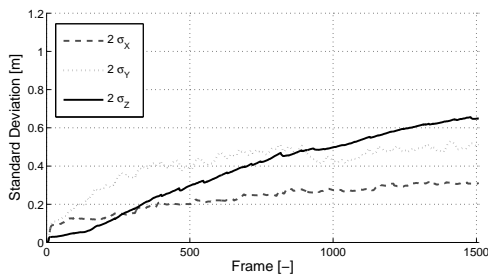


Figure 6: Standard deviation of the estimated projection centers for a long strip by a Kalman-Filter and initialization of new observed points with our proposed stable initialization method.

4.2 Real Environment

The real data experiment were performed near a vineyard. With the support of Microdrones GmbH we acquired an image sequence of a small area using our drone. The average flying height was approx. 30 m. The image sequence consists of vertical views. It contains a building and vineyards in their typical in-line form. Fig. 2 shows every 100th image to give an impression on the roughness of the flight path. The camera was calibrated offline.

To compare our results of the Kalman-Filter with a fully automated bundle adjustment, we use the same observations for both approaches, derived from a KLT-Tracker of the image sequence. The bundle adjustment solution is performed with every 10th frame of the image sequences. 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 and Förstner, 2005). The scale is fixed using a known distance between two object points. The estimated σ_0 was 0.5, indicating that the tracker yields a standard deviation of 0.5 pixels. The Kalman-Filter approach uses a subset of around 1000 object points. In this case 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 our initialization method, c.f. 3.2. 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 it can be seen in figure 10 and 11 the trajectories are nearly similar; up to now we did not perform a statistical analysis of the differences. The bundle adjustment results for camera position and orientation can be used as the input for a dense surface reconstruction, here performed with a 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 . We computed a grid repre-

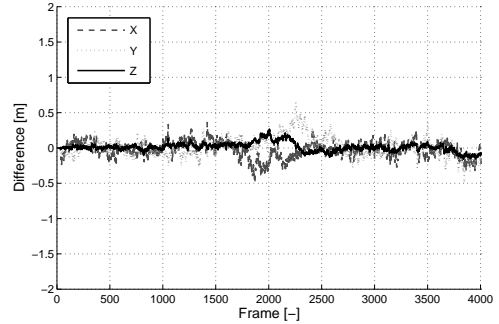


Figure 7: Estimated projection centers for the spiral path example.

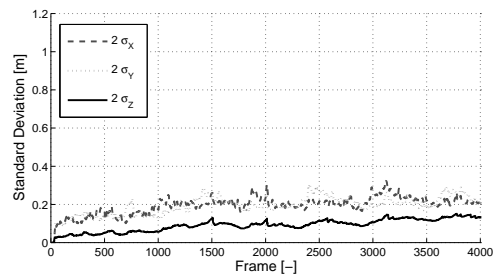


Figure 8: Standard deviation of the estimated projection centers for the spiral path example.

sentation, in which holes in the point cloud were interpolated in a hierarchical way. The shaded 3D surface in figure 12 demonstrates the high resolution achievable with low resolution images, even the rows of the vineyards can be identified.

5 CONCLUSION AND FUTURE WORK

In this paper we discussed the benefit of UAVs as a platform for image acquisition under the condition of real time computation of trajectory and surface information. We presented methods for an effective exploration strategy and for the efficient detection of the loop closing problem using SIFT descriptors under the knowledge of the location and scale of a landmark. An adaptation of an approach for camera motion detection was developed for a stable initialization of new landmarks in the Kalman-Filter by considering the structure of accuracy. In an experimental section we presented results of the influence of this strategies to the estimated camera projection centers in a simulated environment as well as for a real image sequence. The differences between the results of a bundle adjustment and the Kalman-Filter based approach were marginal with respect to a navigational task. Our conclusion is that the real-time computation of a camera trajectory and landmark coordinates from an UAV without GPS is still a complex challenge. As we have shown, the accuracy is suitable.

In the future we will apply a so called sliding-window representation of the camera trajectory for a better approximation of motion in the prediction step of the Kalman-Filter instead of the linear prediction model represented by actual camera orientations and their velocities. This could also stabilize the estimation process with outliers in the observations. Besides a detection of non stationary objects is highly recommended.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. A. Grimm, IGI GmbH and Thorsten Kanand, Microdrones GmbH for providing the required hardware and Felix Müller for his support.



Figure 9: Every 100th image of the sequence, observe rotations and scale differences

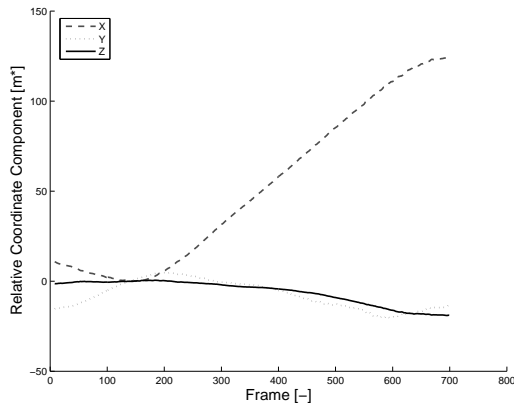


Figure 10: Results of a Kalman-Filter of a real image sequence for the estimated camera projection centers.

REFERENCES

Ackermann, F., 1965. Fehlertheoretische Untersuchungen über die Genauigkeit photogrammetrischer Streifentriangulationen. Deutsche Geodätische Kommission.

Agrawal, M., Konolige, K. and Bolles, R. C., 2007. Localization and mapping for autonomous navigation in outdoor terrains: A stereo vision approach. In: WACV '07: Proceedings of the Eighth IEEE Workshop on Applications of Computer Vision, IEEE Computer Society, Washington, DC, USA, p. 7.

Beder, C. and Steffen, R., 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 (eds), Pattern Recognition, LNCS, Springer, pp. 657–666.

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

Estrada, C., Neira, J. and Tardos, J., 2005. Hierarchical slam: Real-time accurate mapping of large environments. In: Transactions on Robotics, Robotics and Automation, Vol. 21, pp. 588 – 596.

Everaerts, J., Lewyckij, N. and Fransaer, D., 2004. Pegasus: Design of a stratospheric long endurance uav system for remote sensing. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV Part B2.

Franke, U., Rabe, C., Badino, H. and Gehrig, S., 2005. 6d-vision: Fusion of stereo and motion for robust environment perception. In: DAGM '05, Vienna.

Grün, A., 1982. An optimum algorithm for on-line triangulation. In: International Society for Photogrammetry and Remote Sensing, Volume 24 - III/2, pp. 131–151.

Julier, S. and Uhlmann, J., 1997. A new extension of the Kalman filter to nonlinear systems. In: Int. Symp. Aerospace/Defense Sensing, Simul. and Controls, Orlando, FL.

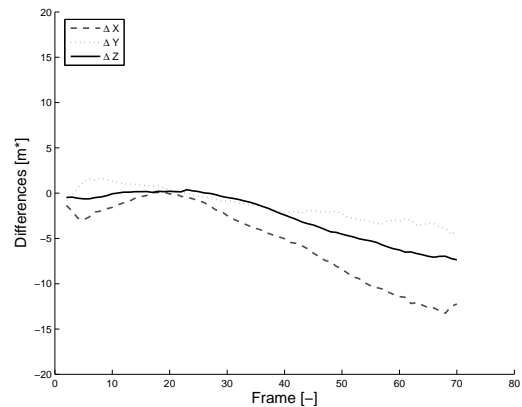


Figure 11: Differences between the results of a Kalman-Filter and a bundle adjustment of a real image sequence for the estimated camera projection centers.

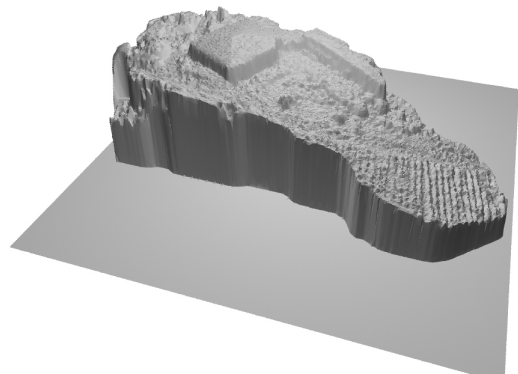


Figure 12: 3d view of a dense elevation map of the real image sequence

Kalman, R., 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME - Journal of Basic Engineering on Automatic Control 82(D), pp. 35–45.

Läbe, T. and Förstner, W., 2005. Erfahrungen mit einem neuen vollautomatischen Verfahren zur Orientierung digitaler Bilder. In: Proceedings of DAGM, Rostock, Germany.

Lowe, D., 2003. Distinctive image features from scale-invariant keypoints. In: International Journal of Computer Vision, Vol. 20, pp. 91–110.

Mikolajczyk, K. and Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis & Machine Intelligence 27(10), pp. 1615–1630.

Montemerlo, M., Thrun, S., Koller, D. and Wegbreit, B., 2002. Fastslam: A factored solution to the simultaneous localization and mapping problem. In: Proceedings of the 18th National Conference on Artificial Intelligence (AAAI), pp. 593–598.

Montiel, J., Civera, J. and Davison, A. J., 2006. Unified inverse depth parametrization for monocular slam. In: Robotics Science and Systems.

Thrun, S., Burgard, W. and Fox, D., 2005. Probabilistic robotics. MIT Press.

Tomasi, C. and Kanade, T., 1991. Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University.

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