Resource-Aware Image Mosaicking on Networks of Small-Scale UAVs

Daniel Wischounig-Strucl
Institute of Networked and Embedded Systems
Klagenfurt University, Austria
daniel.wischounig-strucl@uni-klu.ac.at

Bernhard Rinner (Advisor)
Institute of Networked and Embedded Systems
Klagenfurt University, Austria
bernhard.rinner@uni-klu.ac.at

ABSTRACT

In this approach we estimate the depth structure of images captured by small-scale UAVs to guide the mosaicking of an overview image. We focus on efficient methods where initially only metadata and descriptors of corresponding points are transferred over the network. The complete image is presented later when sufficient communication resources are available.

1. INTRODUCTION

Providing actual overview images captured by unmanned aerial vehicles (UAVs) is of special interest for various applications such as disaster response, accident scenes or building sites. In particular, to guide first-time responders in disaster recovery situations quickly aggregated and georeferenced overview images are very helpful. To cover wide areas such overview image is generated by mosaicking a set of individual images. While video capturing is well established for monitoring selected spots in the disaster area, resource limitations are in favor of keeping the image set small for mosaicking.

In our approach, networked small-scale UAVs are autonomously flying at low altitudes and are capturing images at predefined points to cover the area of interest. The UAVs are affected by wind turbulences due to their low weight and limited thrust. Cost-effective IMU (Inertial Measurement Unit) and GPS (Global Positioning System) modules are used for controlling the UAV; however these sensors provide the camera's position and pose only at limited accuracy.

Compared to aerial images from planes or satellites, the low flight altitude (typically less than $100\,m$ above ground) is very small compared to the elevation of objects at the ground level. Hence, these objects and the non-planar surface induces a significant perspective distortion at individual images. Since we keep the overlap between adjacent images small due to the resource limitations, the perspective distortion increases within these overlaps.

The required image transformations are different from classic panorama mosaicks, where the camera is rotated only around the image axis. In our setting we have to deal with general translation and rotation of the cameras, and our knowledge about the camera position and orientation is very limited due to the inaccurate onboard sensors.

ICDSC 2010 August 31 – September 4, 2010, Atlanta, GA, USA

2. RESEARCH PROBLEM AND APPROACH

The mosaicking of n nadir aerial images can be represented as optimization problem of finding a particular image transformation T_i for each image I_i. These single images I_i are transformed and merged with function $\biguplus$ to the overview image I.

$$I = \biguplus_{i}^{n} T_i I_i \qquad Q = \sum_{i=1}^{n} (\alpha\, G_i(I_i) + (1-\alpha) C_i(I_i, I)) \quad (1)$$

Q is the objective function of the optimization that weights the geospatial accuracy function $G_i(I_i)$ and the pixel correlation function $C_i(I_i, I)$ by $\alpha, 0 \leq \alpha \leq 1$ defined in Equation 1. An additional challenge is introduced when projecting and maintaining the georeferenced data from cameras, considering the device's position and pose inaccuracy.

To solve this optimization problem we have to find appropriate transformation functions for each image. However, there exist various image transformations T_i, e.g., image data based (SIFT, SURF, ...) or metadata based approaches (Position, ...). Due to resource limitations, e.g., communication bandwith, computational performance, flight time and many more, the hybrid approach presented in [3] is preferred. For the approximation of each image transformation T_i a trade-off between the pixel correlation and the preservation of geospatial relations is required.

One may think, instead of using an image transformation approach for rough stitching, texture mapping, i.e., terrain relief transformations on a 3D model would present a seamless solution. However, these 3D strategies are too expensive for delivering the overview image online, i.e., computed on the UAVs during flight.

$$T_i(x, y) \quad = \quad T_{\mathrm{pose},i} \cdot T_{\mathrm{pos},i} \cdot T_{\mathrm{match},i} \quad (2)$$

Equation 2 presents the composition of the proposed transformation. First, images are perspectively corrected by the camera's estimated pose $T_{\mathrm{pose},i}$ to transfer them into nadir view. Second, these images are aligned by the assumed

global position $T_{\text{pos},i}$ to reduce the search space significantly. Finally, the images are mosaicked with neighboring images, by approximating the transformation $T_{\text{match},i}$ to optimize the output quality. Hence, omitting perspective distortions that may propagate over images is one benefit of using the similarity transformation. To enhance the results it is necessary to limit the search space for selecting an *optimal* subset of correspondences.

The *optimum* results for the desired transformation are composed from a subset of points from the same plane in the 3D domain. To determine these points an estimation on the surface structure is required before fitting a common plane to these points in adjacent images. To estimate the elevation levels and the refined camera pose we utilize epipolar geometry [2] and structure-from-motion (SfM), respectively.

In this approach image correspondences with their descriptors are shared together with the surface structure, refined camera pose and position among all contributing neighbors (determined by their image overlaps) over the network. The following processing is executed distributed over all neighboring nodes. Based on that information common planes are fitted into different elevation levels of the image data. The reduced set of point correspondences on these planes is used to compute transformation function. In latter steps additional knowledge gained from the processing could be used to generate a detailed depth model or mark objects in the scene.

3. CASE STUDY AND PRELIMINARY RESULTS

Point features are extracted from images in their overlapping areas, that are determined prior to limit the search area by projecting the camera's view defined by the camera's intrinsics, position and IMU data. The extracted points within this area are matched by the nearest neighbor search to find keypoints with minimum Euclidean distance (cp. Figure 1). To further reduce this set of points from structure outliers, the fundamental matrix is fit using RANSAC. From these inliers the epipolar geometry, defined by the fundamental matrix F and the epipoles e_1 and e_2 is used to reconstruct the Euclidian coordinates [1] with the instrinsics from the camera calibration.

In Figure 2 the result of the rough 3D point reconstruction of the overlapping area is presented, while two cameras are also included with their computed orientation and position. The position and pose of the cameras of the contributing images are aligned with the camera motion retrieved from the epipolar geometry. Point features are projected to absolute coordinates in the 3D space building the surface structure.

Within this structure we search for an *optimum* subset of points, which fits to planes in the whole set. This subset of correspondences is used for the computation of the transformation. If all selected points are on the same elevation level or plane the similarity transformation for these points is able to preserve geospatial relations. Thus it is important to spend effort on selecting those points to avoid perspective distortions.

4. FUTURE WORK

The camera pose gained by SfM will be fused with IMU and GPS data. This could also help to improve the data accuracy on the UAV.

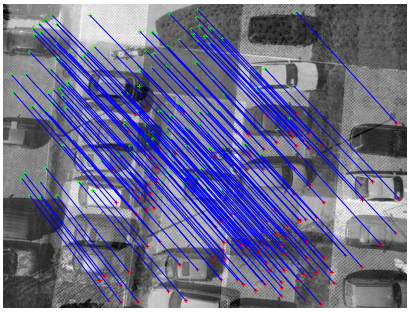

Figure 1: Overlapping image area with camera motion vectors

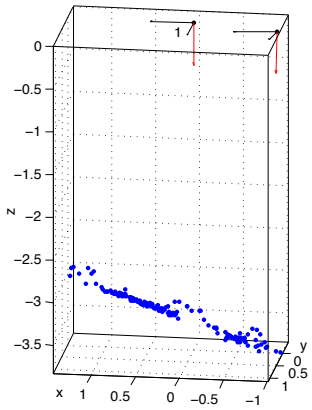

Figure 2: 3D structure constructed in Euclidean coordinates

We will analyze different SfM estimation algorithms and optimization strategies for fitting common planes in adjacent images in the 3D domain.

Due to SfM that is used to estimate the 3D structure of the surface, we can apply texture mapping on this surface structure to generate the mosaick.

Acknowledgment

This work was performed in the project *Collaborative Microdrones (cDrones)* of the research cluster Lakeside Labs and was partly funded by the European Regional Development Fund, the Carinthian Economic Promotion Fund (KWF), and the state of Austria under grant KWF-20214/17095/24772.

5. REFERENCES

[1] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[2] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, page 131 Vol. 1, 1999.

[3] S. Yahyanejad, D. Wischounig-Strucl, M. Quaritsch, and B. Rinner. Incremental Mosaicking of Images from Autonomous, Small-Scale UAVs. *7th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, Sept. 2010.