Autonomous Geo-referenced Aerial Reconnaissance for Instantaneous Applications

A UAV based approach to support security and rescue forces

Axel Bürkle, Florian Segor
Matthias Kollmann, Rainer Schönbein
Department IAS
Fraunhofer IOSB
Karlsruhe, Germany
{axel.buerkle, florian.segor, matthias.kollmann,
rainer.schoenbein}@iosb.fraunhofer.de

Dimitri Bulatov, Christoph Bodensteiner,
Peter Wernerus, Peter Solbrig
Department SZA
Fraunhofer IOSB
Ettlingen, Germany
{dimitri.bulatov, christoph.bodensteiner,
peter.wernerus, peter.solbrig}@iosb.fraunhofer.de

***Abstract* - The Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB) deals with the interoperability of stationary and mobile sensors and the development of assistance systems, which optimize and simplify the operation of such systems. One focus is research on swarms with airborne miniature drones and their applications. The photo flight presented in this paper is one of the applications developed to bring the advantages of a swarm into a realistic scenario. With the aim to support rescue or security forces in action, the photo flight generates an immediate up-to-date situation picture by using an autonomous swarm of miniature drones. If the videos taken from these swarms are geo-referenced, a significantly better coordination and distribution of missions and tasks can be achieved. Therefore, the ability of geo-referencing in unknown terrain is highly demanded. In the absence of an onboard internal navigation system, an image-based approach is proposed. It consists of registration of video frames or even whole sequences stemming from different UAVs (Unmanned Aerial Vehicles) into a common coordinate system and matching with an orthophoto. These processes are called quasi-intrasensorial and intersensorial registration. For image-based geo-referencing, a differentiation must be made between scenes with negligible spatial depth (2D situation) and those where the depth cannot be neglected any longer (2.5D situation). Several applications of geo-referenced UAV-borne videos as well as ideas how the task of image-based geo-referencing can be accelerated for online processing are presented.**

Keywords - aerial situation image; unmanned aerial vehicles; swarm; geo-referencing

I. INTRODUCTION

This paper presents our recent work on a universal ground control station called AMFIS ("Aufklärung mit Miniatur-Fluggeräten im Sensorverbund", or reconnaissance with miniature aerial vehicles in a sensor network) [1]. AMFIS is a component-based modular construction kit that supports various aerial reconnaissance and surveillance tasks. It has served as the basis for developing specific products in the military and homeland security market. Applications have been demonstrated in several exercises for the PASR (Preparatory Action for Security Research) program of the European Union, the German Armed Forces, and the defense industry. The surveillance system AMFIS is an adaptable modular system for managing mobile as well as stationary sensors. The main task of this ground control station is to work as an ergonomic user interface and a data integration hub between multiple sensors mounted on light UAVs or UGVs (Unmanned Ground Vehicles), stationary platforms (network cameras), ad hoc networked sensors, and a superordinated control center.

Several software modules assist the user in obtaining aerial situation pictures. One module is the photo flight tool. This is a special property of the flight route planning in AMFIS that allows creating highly up-to-date aerial pictures of a predefined area in a short time. The software module itself is designed to work both as independent standalone software and as a part of the complex control system AMFIS. Another important property of an airborne drone is its ability to orientate autonomously in an unknown terrain. This is not only necessary to guarantee situation awareness and flexibility, but also to achieve a better coordination with other drones in the swarm. If precautions are not made, commonplace applications such as mosaicking and detection of moving objects usually suffer from error accumulation and false alarms caused by 3D structures. These problems can be solved if, ideally, each pixel of each frame is assigned a 3D coordinate. The 2D registration can be achieved if this frame is geo-referenced onto an orthophoto while the 3D component could be obtained from photogrammetric or architectural databases. Unfortunately, the obvious method of registration, namely, the use of the navigation equipment onboard of the UAV, becomes less reliable for rather inexpensive and miniaturized MUAVs (Micro UAVs). The reason is the low accuracy data stemming from such a lightweight, inexpensive navigation unit. In addition, GPS data (putting aside photogrammetric data-bases) are hardly reliable in a considerable number of applications, for example, near building walls, due to multi-path propagation. Therefore, an alternative, image-based group of approaches was developed to perform geo-referencing. We will describe the methods where the spatial depth can be neglected (2D applications) and a projection of pixels is given by a

perspective transformation of plane (2D homography). However, in the case of a relatively low sensor altitude and a moderate focal length, needed in order to achieve a satisfactory resolution of the acquired images and videos, the presence of buildings and vegetation cannot be interpreted as a disturbing factor for the upcoming computations any longer without a significant loss of accuracy. As a consequence, it is important for urban terrain to extend the existing concepts by 3D or, at least, a 2.5D method. By 2.5D situation, we understand the function of the terrain altitude depending from its longitude and latitude. This assumption is reasonable for videos taken from nadir perspective. The counterpart of homography is, in this case, a depth map that assigns a depth value to almost every pixel. An intermediate result is given by a Euclidean reconstruction of sensor trajectory and a sparse point cloud.

With respect to matching tools, there must be a trade-off between a reliable registration and a reasonable computational load. We differentiate between a (quasi)-intrasensorial registration that links neighboring frames of video sequences or even different video sequences recorded by different UAVs and an intersensorial registration that allows matching scenes of a rather different radiometry. Here, a particular contribution of this work consists in creation of synthetic images for 3D-registered mosaics and modification of the well-known adaptive self-similarity approach [2] for such synthetic images. All situations mentioned in this paragraph are summarized in Table 1.

TABLE I. GEO-REFERENCING, OVERVIEW OF RELEVANT SITUATIONS

	Intrasensorial registration	***Quasi-intrasensorial registration***	***Intersensorial registration***
Participate	Neighboring frames of the same video	Different subsequences of the same video or different, but similar videos	(Mosaicked) video and orthophoto
Geometry	From frame to frame: almost the same	Different in scaling, angle of view	Can be adjusted, rasterization in 3D see Section VI.B)
Radiometry	From frame to frame: negligibly small differences	Very small differences	Large differences
Matching tool	KLT-tracking, [3]	SIFT, [4]	SIFT for moderate differences in radiometry or adaptive Self-similarities, [2] (otherwise)
2D case	Registration by 2D homography	Registration by 2D homography, [5]	Registration by 2D homography, see Section VI.A
2.5D resp. 3D case	Fundamental matrix, Euclidean reconstruction, depth maps extraction, [6], [7]	Registration by 3D homography, Section VI.B	Creating a synthetic image and registration with orthophoto by 2D homography, see Section VI.B

In the current implementation, the methods for geo-referencing can only be performed offline. However, efforts are being made to integrate them into the AMFIS station.

The paper is structured as follows: After a short survey of related work an overview of the application scenarios is presented in Section III, followed by a description of the airborne platform in Section IV. Section V introduces the used algorithms followed by the description of the post processing in Section VI. The paper concludes with a summary (Section VII) and an outlook on future work (Section VIII).

II. RELATED WORK

Related work is discussed regarding the two focuses of this paper, namely flight platforms and geo-referencing.

A. Flight Platform

With respect to the photo flight, there are some projects with a similar scope.

At the "Universität der Bundeswehr" in Munich, a UAV for precision farming [8] is currently being developed. It is used to analyze agricultural areas from the air to find the regions that need further manuring to optimize the growth of the crop. A commercial off-the-shelf fixed wing model is equipped with an autopilot and either a near infrared or a high quality camera. This technique allows monitoring the biomass development and the intensity of the photosynthesis of the plants.

The AirShield project (Airborne Remote Sensing for Hazard Inspection by Network Enabled Lightweight Drones) [9][10], which is part of the national security research program funded by the German Federal Ministry of Education and Research (BMBF), focuses on the development of an autonomous swarm of micro UAVs to support emergency units and improve the operational picture in case of huge disasters. The aim is to detect potentially leaking CBRNE (Chemical, Biological, Radiological, Nuclear, Explosives) contaminants in their spatial extent and to carry out danger analysis without endangering human life. The swarm is supported by a highly flexible communication system, which enables communication between the swarm members and between the swarm and the ground station.

The precision farming project as well as the AirShield project are very promising and show first results. However, the application aim of both projects differs from ours although we plan to extend the photo flight to scenarios similar to the ones of AirShield (see Section VIII).

B. Geo-referencing

In the field of geo-referencing, we refer to our previous work ([5], see also references therein), where only 2D-relevant situations were taken into account. If there is a large parallax between the frame and the orthophoto, triangular networks can be mapped onto the orthophoto, which allows creating multi-homography-based mosaics [6]. This has advantages in situations where no 3D reconstruction can be performed from the video (e.g., in the case of zooming and rotating cameras). The relevant work on registration of 3D point clouds to 2D images and the closely related problem of

pose estimation can be found in the recent work of [11] and in references cited there. In order to learn about producing synthetic images for matching, we refer to [12]. Finally, for creating dense depth maps from a set of images that allow rendering 3D clouds, a survey [13] can be recommended. Examples on algorithms for multi-view dense matching recently developed can be found in [7] and [14].

III. Application Scenarios

The sense of security in our society has significantly changed over the past several years. Besides the risks arising from natural disasters, there are dangers in connection with criminal or terroristic activities, traffic accidents or accidents in industrial environments. Especially in the civil domain in case of big incidents there is a need for a better data basis to support the rescue forces in decision making. The search for people buried alive after a building collapses, or the analysis of fires at big factories or chemical plants are possible scenarios addressed by our system.

Many of these events have very similar characteristics. They cannot be foreseen in their temporal and local occurrence so that situational in situ security or supervision systems are not present. The data basis, on which decisions can be made is rather slim and therefore the present situation is often very unclear to the rescue forces at the beginning of a mission. However, these are exactly those situations, for which it is extremely important to understand the context as fast as possible to initiate suitable measures.

An up-to-date aerial image can be a valuable additional piece of information to support the briefing and decision making. However, helicopters or supervision airplanes that can supply this information are very expensive or even unavailable. Up-to-date high-resolution pictures from an earth observation satellite would provide the best solution in most cases. But under normal circumstances these systems will not be available. Nevertheless, it would usually take too long until a satellite reaches the desired position to provide this information. A small, transportable and, above all, fast and easily deployable system that is able to produce similar results is proposed to close this gap.

The AMFIS tool “photo flight”, explained in Section V, can provide the missed information by creating an overview of the site of interest in a very short time. The application can be used immediately at the beginning of the mission with relative ease and the results provide a huge enhancement to already available information.

Applications include support of fire-fighting work with a conflagration, clarification the debris and the surroundings after building collapses, and search for buried or injured people. Additionally the system can be used to support the documentation and perpetuation of evidence during the cleaning out of the scene at regular intervals.

Non-security related application scenarios are also conceivable, as for example the use of infrared cameras to search large cornfields for fawns before mowing or to document huge cultivated areas or protective areas and biotopes.

The photo flight tool shows excellent results in the production of up-to-date aerial situation pictures in ad hoc scenarios. The intuitive and ergonomic graphic user interface allows the operator to define an area of interest and start the photo flight. The results are a number of images depending on the size of the area of interest. They are merged and geo-referenced by suitable tools.

Since AMFIS is capable of controlling and coordinating multiple drones simultaneously [15], the photo flight tool was designed to make use of a UAV swarm. By using more than one UAV, the same search area can be covered in less time or respectively a bigger area can be searched in the same time.

The biggest problem when working with multiple UAVs is the dwindling clarity for the operator, especially when there are different types of UAVs and payloads. The more drones used in an application, the more complicated the control of the single systems gets. That is why it is most essential to reduce the work load on the user as much as possible. Therefore the idea of a self-organizing swarm is transferred to the photo flight application in order to reduce the efforts for controlling this tool to a minimum. The user only has to define the area of interest and decide which drones he would like to use.

The application is responsible for all additional work, such as the composition of the respective flight routes, the control of the single UAVs including the observation of the aerial security to avoid collisions as well as setting the return flight.

IV. Flight Platform and System Architecture

The primary aim of the photo flight is the clarification of certain areas. The used drones do not necessarily have to be identical. They also can differ in their technical configurations. Nevertheless, in this first research attempt to build a swarm, UAVs of the same type were used.

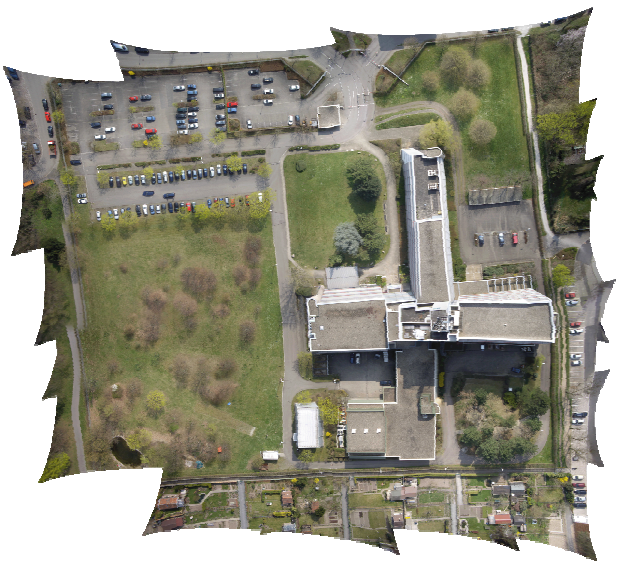

Figure 1. Situation picture from photo flight (ca. 9500 x 9000 pixel)

Figure 2. Sensor platform AirRobot 100-B.

A. *Flight Platform*

Enormous effort has been put into the selection of this flight platform. A platform that already comes with a range of sensors, an advanced control system and autonomous flight features significantly reduces the effort for cooperative swarm of micro drones. Furthermore, when it comes to flying autonomously, the system has to be highly reliable and possess sophisticated safety features in case of malfunction or unexpected events.

Other essential prerequisites are the possibility to add new sensors and payloads and the ability to interface with the UAV's control system in order to allow an autonomous flight. A platform that fulfils these requirements is the quadrocopter AR100-B by AirRobot (see Figure 2). It can be controlled both from the ground control station through a command uplink and by its payload through a serial interface.

To form a heterogeneous swarm from different UAVs, new systems were gradually integrated. Currently, beside the AR100-B there is also a Microdrones MD4-200 as well as a MikroKopter with eight rotors (MK Okto). The user can identify the system by its call sign – the operation of the drone, however, remains identical, rendering the complexity of the heterogeneity transparent.

B. *Software Architecture*

The AMFIS ground control station's software architecture is basically 3-tiered, following a pattern similar to the MVC (Model-View-Controller) paradigm best known from web application development. The central application is the so-called AMFIS Connector (see Figure 1), a message broker responsible for relaying metadata streams within the network. Metadata is transmitted using an XML-based

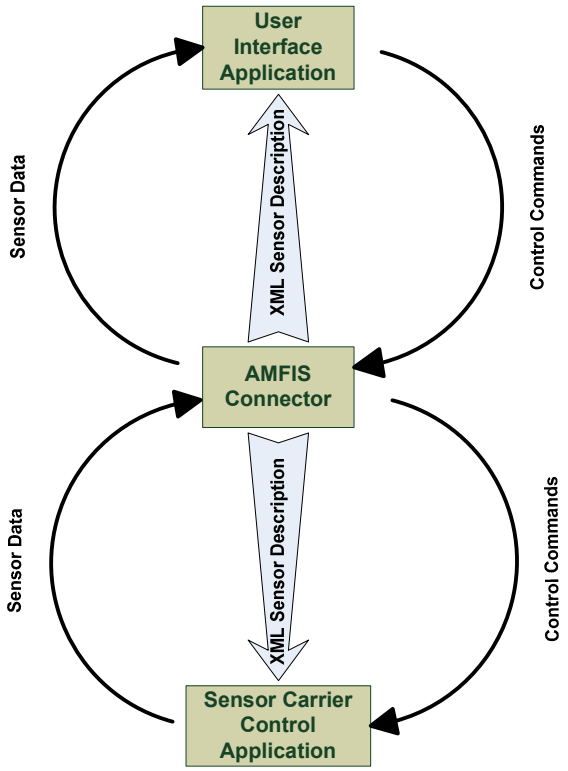

Figure 1. Software architecture of the AMFIS ground control station.

message protocol and may represent both sensor measurements as well as sensor carrier control commands. Since the biggest amount of data transmitted during a typical scenario is video (thus binary) data, the connector is tightly coupled with a second server application, the Videoserver. It is responsible for storing and distributing video streams, serving the dual purpose of providing time shifting capabilities to the network as well as reducing the load on the usually wireless links between sensor carriers and the ground control station. Since time shifting or archiving is not always required, this functionality was not integrated in the Connector in order to keep it as light-weight as possible.

Upon connection, each client application first receives an XML document describing the various sensor carriers currently active within the ground control station's network along with a unique ID used to control metadata flow. A communication library (AmfisCom) builds an object tree from the XML data, providing the application developer with a type-safe, object-oriented view of the network of sensors and sensor carriers.

A client application in this context is any application that either includes one of the numerous AmfisCom implementations (.NET, Qt, Java) or implements the AMFIS message protocol directly. Prominent examples are the GUI (Graphical User Interface) applications (analyst's interface, pilot's interface, and situation overview) or the photo flight or various transcoder applications responsible for translating metadata between the AMFIS message protocol and a

proprietary protocol used, for example, on a low-bandwidth radio data link to a distant sensor node.

After successful establishment of a connection, the Connector supplies the application with a constant stream of live sensor data. Optionally, the Videoserver provides time shifted video and metadata streams, in case an operator requires reviewing a critical situation.

V. PHOTO FLIGHT

To map an area of interest by multiple drones, the polygon defining that area must be divided into several subareas, which can then be assigned to the individual UAVs. It is important that each of the branches is economically optimized for its appropriate drone. UAVs with longer endurance or higher sensor payload can clear up larger areas and should therefore receive longer flight plans than systems with a lower performance.

The flight routes must also consider the behavior of the drones at the single photo points and their flight characteristics. Tests with the multicopter systems have shown that an optimum picture result can be achieved if the system stops at each photo point for two to four seconds to stabilize. If proceeding precisely in this manner, no special flight behavior is necessary, because the drones show identical flight characteristics in every flight direction due to their construction. Indeed, this behavior also decisively affects the operation range, because such stops reduce the efficiency of the drones. To solve this problem, a stabilized camera platform was developed at Fraunhofer IOSB, which compensates the roll, pitch and yaw angles. Nevertheless, if the photo points are flown by without a stop, an enlargement of the flight radii must be considered at turning points. As fixed-wing aircrafts may be used in future versions, more attention must be paid to the fact that the calculated flight paths can also be optimized for systems with different flight characteristics.

The algorithm developed from these demands consists of two main steps. The first part is to break down the given polygon of the search area in suitable partial polygons (A). Secondly, the optimum flight route per partial polygon is searched for each individual drone (B).

A. Calculation of the partial polygons

Different attempts for decomposing the whole polygon into single sub cells were investigated.

An elegant method to divide an area into subareas is the so-called “Delaunay-Triangulation” [16]. Unfortunately it proved to be very difficult to divide a polygon in such a way that the resulting partial polygons correspond to a certain percentage of the whole area.

In addition to the basic triangulation, the single polygon had to be checked for their neighborhood relations in order to recompose them accordingly. The originating branches would hardly correspond to the targeted area size so that additional procedures would have to be used.

As an alternative, the possibility to divide a polygon by using an approximation procedure and surface balance calculation to get partial polygons was investigated. With this variation, the polygon is disassembled first into two incomparably large parts by using predefined angles through the surface balance point. It is irrelevant whether the balance point lies outside or within the body. According to the desired size, the algorithm can select the bigger or the smaller partial polygon as a source area for any further decomposition. Afterwards, the calculated partial polygon is divided again by the surface balance point. This process continues recursively until the requested area size is reached. On this occasion, an approximation procedure could be used to calculate a solution as quickly as possible. However, this segmentation method only works with convex polygons. For concave polygons, it is necessary to prevent the area being divided into more than two parts.

A quicker and mathematically less complicated variation to split a polygon is the scanning procedure (see Figure 3). The method is equal to what is called rendering or scan conversion in 2D computer graphics and converts the polygon into a grid of cells. That implies that a higher resolution (i.e., a smaller cell size) will result in a more accurate match of the grid with the originally defined area. To be able to divide the grid afterwards, the number of required cells is calculated from the desired area size. With this information and by using a suitable growth algorithm, which extends from any start cell within the grid as long as enough cells have melted, one single continuous area of the desired size can be calculated. This technique resembles the flood-fill algorithm [17] also known from computer graphics. Likewise in this case, it is important to know the neighborhood relationship of the cells. Nonetheless, this is quite simple because in contrast to the same problems with the triangulation each of these cells is commensurate and is therefore easy to assign to the co-ordinate system.

To receive a very simple and steady grid polygon, different growth algorithms were compared to each other. A straight growing algorithm turned out to be the most efficient because the results showed more straight edges than other algorithms. This means a significant reduction of the required rotary and turn maneuvers of the UAV, which leads to a better cost-value ratio if using UAVs with a limited turning rate. The generated grid polygons are recalculated into partial polygons just to disassemble them once more into a grid. This time the grid size corresponds to the calculated dimension of the footprint, which depends on the camera specification (focal length and picture sensor) in combination with the desired flight altitude of the drone.

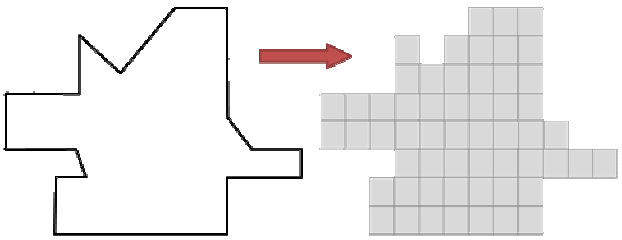

Figure 3. Scanning procedure.

B. Calculating the flightpath

To receive an efficient and economically reasonable flight route, it is important to find the shortest path that includes all way points and in addition contains the smallest possible number in turn maneuvers.

The best flight path solution can only be calculated by using a highly complex algorithm and even then, an optimal result cannot be achieved in reasonable time (see the problem of the travelling salesman [18]).

To achieve acceptable results under the constraint to keep expenditures as low as possible, different variations were checked concurrently.

As mentioned earlier, a very steady flight route with as few as possible direction changes offers huge economic advantages, a method was developed, which processes the polygon according to its expansion in columns or line-by-line similar to a typewriter. The resulting flight route shows a clearer construction in particular with bigger areas.

Afterwards the calculated flight route is complemented with safe approach and departure air corridors to avoid collisions between the swarm members.

C. Transferring the images

In order to mosaic and geo-reference the data accumulated by the drones, the high-resolution images have to be transferred to the ground control station. The most elegant way to transfer the images is to use the downlink of the drone. This requires that the UAV has an interface, which can be used to feed the data into the downlink of the system. If such an interface is not available, other procedures have to be found. During the development of the photo flight, different technologies were tested and evaluated.

To keep the system as simple as possible, the best solution would be to select a communication device that has a great acceptance and is widely used. Therefore the first drafts where done by Wi-Fi. To build such an additional communication line between the UAVs and the ground station, a small secure digital memory card was used. This SD card fits perfectly well into the payloads and is able to establish a Wi-Fi connection and to transmit the captured images automatically. The disadvantage of this solution is that the frequencies for the digital video downlink of the drones are in the 2.4GHz band, which is also mostly used by Wi-Fi for broadcasting. For this reason, it can be assumed that at least the Wi-Fi transmission will experience heavy interferences. The best solution for this problem is to move either the digital video downlink or the Wi-Fi to the 5GHz band. Unfortunately in the current system stage the video downlink is fixed and the used Wi-Fi SD card is not capable of using the 5GHz band.

For now, the images have to be transferred manually to the ground station.

VI. GEO-REFERENCED MOSAICKING AND APPLICATIONS

To benefit fully from the advantages of the photo flight, the images taken must be merged to an overall situation picture. In addition to the offline mosaicking based on high resolution still images, there is also the possibility to create a near real-time mosaic using the live video stream as described below.

A. 2D geo-referencing of video frames and aplications

We start a description of our geo-referencing algorithm of a video taken by a straight-line-preserving camera for the case of negligibly small depth of the scene compared to the sensor's altitude. This makes a 2D homography (plane perspective transformation, see [19]) a suitable model for registration. We denote the (global) homography between the frame I_t captured at time t of the sequence and the orthophoto by H_t, while the local transformation between the frames I_t and I_j is denoted by $H_{j,t}$. Interest points in the frame I_t will be denoted by x_t and in the orthophoto by X.

As previously discussed, the availability of GPS/INS is not constantly needed but useful for the initialization. It can be assumed that the geo-coordinates of the AMFIS base-station as well as a coarse initial value of the sensor altitude and orientation are given. As a consequence, the value of the initial homography H_1 is assigned.

The process works in the same way as described in our previous work [5] and visualized in Figure 4. The intrasensorial registration of neighboring frames of the video sequence – also known as mosaicking – is performed via KLT-tracking algorithm [3] supported by a robust method of outlier rejection (in our case, it is the RANSAC [20]

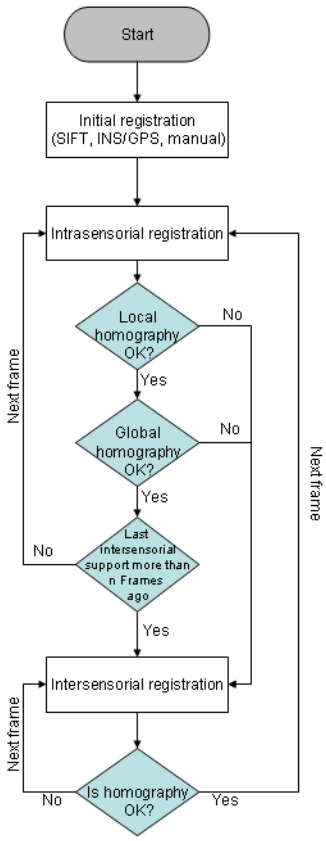

Figure 4. Flow-chart of the Image-based Georeferencing Process, see also [5].

algorithm accelerated by a $T_{1,1}$-test [21]). The quality of the local homography $H_{t-1,t}$ between two subsequent frames I_{t-1} and I_t is estimated by the number and distribution of inliers of $H_{t-1,t}$. The quality of the global homography is given by the average value of the reprojection errors between X and $H_t x_t$. As soon as the quality of either local or global homography is below a threshold, it is rejected and the intersensorial registration procedure between the current frame and the orthophoto has to be performed again.

This intersensorial registration comprises the crucial part of our procedure. The task of automatic matching images having different radiometry and also taken from slightly different viewing directions is known to be challenging. Among numerous methods [1][4][22][2] that we tested for registration, there were two candidates that most closely met our expectations.

First, the SIFT method [4] is invariant against rotations and scaling differences (moderate, until approximately 1:3). These properties, as well as the fact that SIFT is easily parallelizable and implementable on GPUs [23], make it one of the most popular state-on-the-art methods for registration, especially, when combined with a sophisticated software architecture as illustrated in Figure 4.

However, standard descriptors like SIFT do not perform well if corresponding images have a completely different radiometry. Also, when differences in resolution are insuperable for SIFT-descriptors, we follow a different approach. We use Self-Similarity descriptors [2] in combination with an Implicit Shape Model (ISM) in order to find matches between local image regions. We additionally filter the correspondences based on a self-similarity distances densely computed across both images.

Our current implementation robustly handles very large radiometric changes but has a limited scale and rotation invariance. Due to these restrictions and the involved high computational cost, we use this method after the creation of large and accurate mosaics (e.g., in the case of synthetic images described in the following section), when rotation and scaling parameters are approximately known. This also avoids unnecessary registrations.

Several applications of the registration procedure were discussed in [5]. For motion detection, weighted image differences of video frames can be computed and a threshold decision can deliver possible alarms. The advantages of the geo-referencing consists, first of all, of an elegant and efficient possibility to remove false alarms (in the areas, which are known to be parts of buildings, there can be no motion, so the false alarms in these areas are probably 3D-influenced). Secondly, trajectories of different objects are mostly easy to recognize, if they are referenced on a map. For an object reported by different sources, e.g., two UAVs of the swarm, or an UAV and a stationary camera looking from a different direction, the operator has less difficulties deciding whether all reports refer to the same object or several different objects in the case of given geo-coordinates and geo-referenced object trajectories. Finally, additional information, such as the speed of vehicles is easily calculable from their tracks on the maps and the camera frame rate. Results of geo-referenced motion detection are presented in Figure 5.

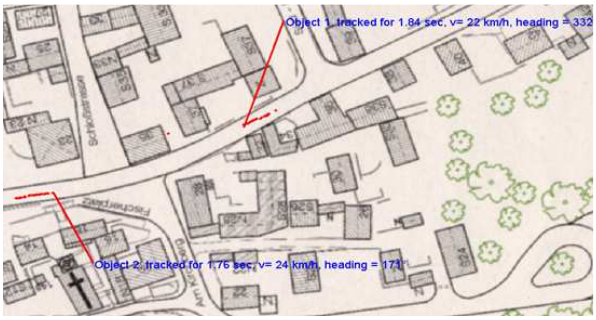

Figure 5. Two trucks have been detected and tracked on a UAV flight. Their velocities and headings can be computed, as in [5].

The other important application is given by annotation of objects of interest into the video by means of the inverse homographies. These objects can be loaded, if needed, from a data-base.

B. 3D reconstruction and geo-referencing of 3D mosaics

The assumption of a negligible spatial depth can either be made if the sensor's altitude is low, if the field of view is extremely narrow (very short focal length) or if the terrain is approximately planar. In other words, in order to achieve a satisfactory resolution of the acquired images and videos in urban terrain, the presence of buildings and vegetation can no longer be interpreted as a disturbing factor for the upcoming computations, even for Nadir views. As a consequence, the 3D structure of the scene must be taken into account for geo-referencing.

For 3D reconstruction, one of numerous approaches for structure-from-motion (SfM) or simultaneous localization and mapping (SLAM) available in the literature can be used. We use an approach of [6] because no additional information is needed here except the video stream itself. Characteristic features are detected [24] and tracked [3] from image to image. Then, fundamental matrices can be retrieved from pairs of images. As a result, a sparse set of 3D points and camera matrices in a projective coordinate system are computed. A step of self-calibration is needed to obtain an (angle- and ratio-preserving) Euclidean reconstruction, illustrated in Figure 6. In order to minimize the geometric error and thus improve reconstruction results, bundle adjustment over camera parameters and coordinates of 3D points can be activated. If the application is time-critical, dense reconstruction can be performed by creating triangular networks [16] from already available points and triangular interpolation [6]. Otherwise, a robust and accurate multi-camera approach [7] recently developed can be applied to extract the depth information from (almost every pixel) of several frames of the sequence.

From this point, we can proceed to geo-referencing of the reconstructed scene. Here, we subdivide this task into two subtasks: the first subtask is a quasi-intrasensorial regi-

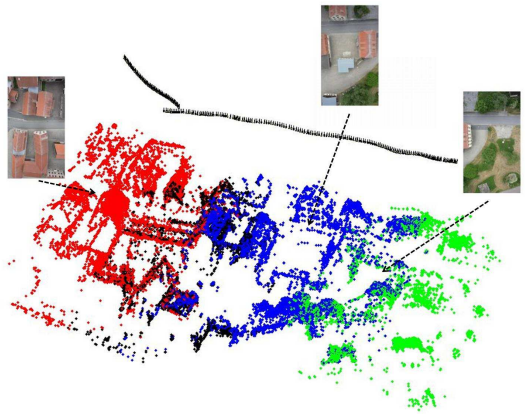

Figure 6. Result of the registration procedure of four sequences (illustrated by different colors) and the camera trajectory depicted by viewing cones. Three example images are shown as well.

stration of different subsequences of the video sequence or different video sequences of a similar appearance. These video sequences can be taken by different UAVs that carry different kinds of cameras and operate, as discussed before, in different parts of the region to be explored. As a consequence, the slightly misleading term of intrasensorial registration is now replaced with quasi-intrasensorial registration. Since the camera trajectory and point coordinates for every reconstruction in some relative, Euclidean coordinate system have different positions, orientations and scales, a registration step is necessary. The second subtask treats registration of the material obtained by videos and our 3D reconstruction procedure onto the orthophoto. A detailed explanation of these subtasks will be provided in the following two subsections.

We assume to be given two sets of camera matrices $\boldsymbol{P}_1$ (= $P_{1,1}$, ..., $P_{1,K}$) and $\boldsymbol{P}_2$ (= $P_{2,1}$, ..., $P_{2,L}$) in homogeneous coordinates that were computed by a SfM approach in different Euclidean coordinate systems (denoted by the first sub-scripts). The desired output is a spatial transformation between these coordinates systems. This transformation H (also called 3D homography) is given by a regular 4×4 matrix in homogeneous coordinates such that the relations

$$P_{1,k} = P_{2,k}H \quad \text{and} \quad P_{1,l}\,H = P_{2,l}$$

hold for $k = 1,\ldots K$ and $l = 1, \ldots, L$. Without loss of generalization, we assume that two images corresponding to cameras $P_{1,K}$ and $P_{2,1}$ cover an overlapping area of the data-set and that point correspondences c_1 and c_2 can be detected in these images by means of a matching operator (e.g. [4]). Given camera matrices $P_{2,1}$, $P_{2,2}$, ..., we now compute, by tracking points c_2 in other images of the second sequence, several 3D points Y in the second coordinate system, see the linear triangulation algorithm [19]. By backward tracking points c_1 in other images of the first workspace and Y, we obtain the set of camera matrices $Q_{1,K}$, $Q_{1,K-1}$,... via camera resection algorithm [19]. For more than one corresponding

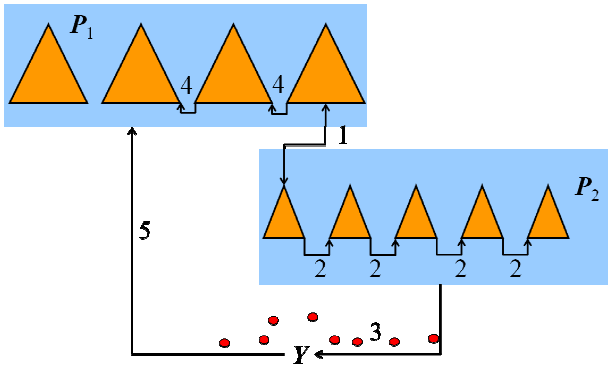

Figure 8. Result of the registration procedure of four sequences (illustrated by different colors) and the camera trajectory depicted by viewing cones. Three example images are shown as well. Registration of two reconstructions: camera locations are depicted by orange viewing cones, 3D points by red circles. Different steps are illustrated by arrows and numbers: 1: Registration of features in two sequences, 2. Tracking of registered features and 3: Triangulation in the second coordinate frame, 4: tracking and 5: obtaining camera parameters of the first reconstruction in the second coordinate frame. The transformation H is obtained from at least two corresponding camera matrices.

cameras $Q_{1,K-n}$ to $P_{1,K-n}$, the initial value of the spatial homography H as a solution of the over-determined system of system up-to-scale

$$\begin{bmatrix} P_{1,K}H \\ P_{1,K-1}H \\ \ldots \end{bmatrix} \cong \begin{bmatrix} Q_{1,K} \\ Q_{1,K-1} \\ \ldots \end{bmatrix}$$

is obtained via Direct Linear Transformation method and refined by means of a geometric error minimization algorithm. The process of the intrasensorial registration is schematically visualized in Figure 8 while Figure 6 shows an example of registering four Euclidean reconstructions for a UAV data-set. From the kink in the camera trajectory in Figure 6, one can see that we are dealing with two different UAV-flights.

We now proceed to the second part of the algorithm. In the case of nadir views, the results of the registration procedure described above can be rasterized into the xy-plane. To do this, we need both to make the z-axis coincide with the physical vertical direction (there are several simple heuristics to do this job) and to obtain the 3D information for a dense set of points. This was done by our approach [7] for

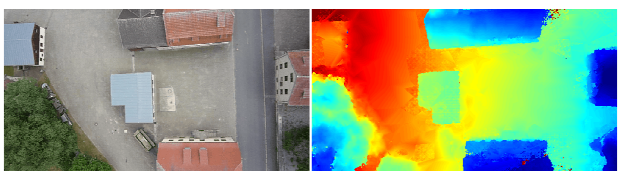

Figure 9. An image and the corresponding depth map obtained with [7]. This is the middle image of Figure 6. Small outliers do not cause any trouble since they can be removed during the rasterization procedure.

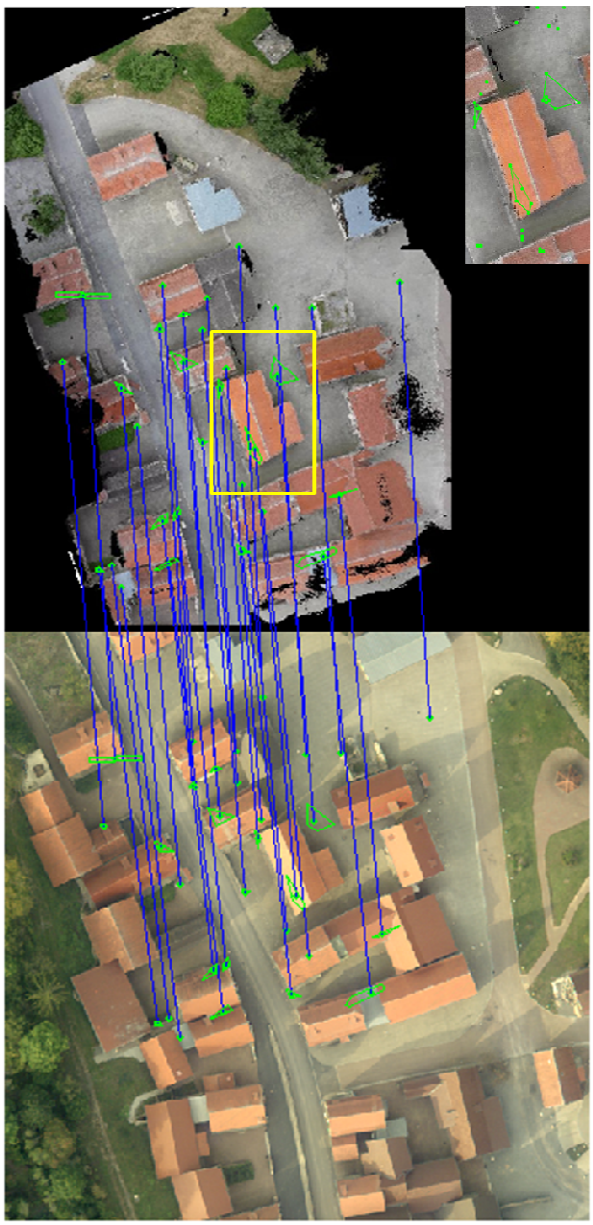

Figure 10. The synthetic image (top) and fragment of the orthophoto (bottom), as well as and clusters (green) of the RANSAC-inliers (connected by blue lines) determined by our extension of the self-similarities algorithm.

computing depth maps. In the future, it will be interesting to investigate to what extent the meshes obtained via triangular interpolation from already available points can compensate for forfeits in the quality of the synthetic image. We show an exemplary depth map in Figure 9 and the synthetic image itself in the top of Figure 10. The result of the registration of the synthetic image to the orthophoto is visualized in Figure 10 and Figure 11. We also mention that the points situated in elevated regions can be optionally identified in the depth maps and excluded from further consideration.

Figure 11. The synthetic image registered onto the orthophoto.

VII. CONCLUSIONS

The described algorithms for the photo flight application were implemented as a software library and are integrated into a geographic information system based on ESRI software [25] specially provided for test purposes. The photo flight tool is an independent software module whereas the logic behind it is interchangeable. The results of the algorithm and its ability to adapt to new flight systems with other flight characteristics are currently evaluated. The software was integrated into a three-dimensional simulation tool and the first real test attempts with homogeneous and also with small heterogeneous swarms have taken place.

This research project resulted in a complex prototype system, which is able to form a fully autonomous swarm of UAVs on the basis of several drones and a standard PC or mobile computer at almost any place in very short time that allows acquiring a highly up-to-date aerial image. The sustained data can also make it possible to understand complex blind scenarios quicker. It permits a more exact planning and simplifies the contact with the situation. The deployment of a swarm with a theoretically unlimited number of UAVs thereby means a huge advancement in the field of local just-in-time reconnaissance and geo-referencing.

With respect to the image-based geo-referencing without using internal navigation, we showed a robust and autonomous approach that works both in situations of 2D registration (with applications of motion detection and object annotation) as well as 3D registration of Euclidean reconstructions and matching of a synthetic image thus obtained with an orthophoto, even in the case of different radiometry. To create a synthetic image, the assumption of 2.5D surface (terrain skin $z(x, y)$) must hold; as a consequence, our methods work better for almost nadir-views of video frames. In the case of an oblique view with a non-negligible spatial

Figure 12. Gas sensor to detect inflammable gases, Ammonia, Nitrogen Dioxide, Sulphur Dioxide, Carbon Monoxide and Chlorine.

depth, algorithms of pose estimation can be applied, but they stay beyond the scope of the work presented here.

Detection of motion in 3D scenes is carried out by an accurate occlusion analysis already on the state of multi view dense depth maps computation while annotation can be generalized from the 2D case.

VIII. FUTURE WORK

Parallel to the work on the photo flight algorithms, a small gas sensor, which can also be carried as a payload by a UAV, was developed in cooperation with an industrial partner (see Figure 12). The gas sensor is designed as a very light and compact payload and has been built as a prototype. It can be equipped with up to five different gas sensors and, in addition, contains a photo-ionization detection sensor and a sensor to detect universal inflammable gases. Future versions will also be able to detect temperature and humidity. The selection of the five gas sensors can be changed to fit different applications at any time. A supplementation or a further development of the photo flight, in which at least one UAV is equipped with a gas sensor, is planned. Since the aim of this application differs from the original task of visual reconnaissance, above all, the geometry of the flight routes must be adapted. This can be assumed from the fact that either the propagation of the gases or the concentration at certain places is of interest. That means that a meandering flight path over a relatively small area makes no sense.

To recognize the propagation of gases, certain a priori knowledge like origin, wind force and direction is necessary. With the help of these data, a propagation model can be provided as a basis for the calculation of optimum flight routes to validate the estimated results.

The approaches of geo-referencing are robust, nearly fully-automatic, and real-time oriented; that is, reconstruction goes along with the video sequence from its beginning to its end and is *not* supposed to be performed after the whole movie has been captured. Due to relatively slow matching algorithms [4] and [2], the 3D reconstruction can only be performed offline at the current state of the implementation, but one of our goals for future work consists of creating a real-time interface for estimation of camera trajectory and a sparse point cloud. An obvious hardware-based acceleration of our algorithm consists of intersensorial registration on GPU parallel to the already extremely fast intrasensorial registration. By better exploitation of initial homography (guided matching) and neighborhood relations, the computing time and memory load for point matching can be drastically reduced.

The only interactive part is given by determination of images with covering overlapping parts of the terrain for quasi-intersensorial registration of Section VI.B. However, in the cases of almost nadir views over the urban terrain and, thus a 2.5D representation of the scenery, it will be possible, in the future, to automate the approach by identifying the search space via *xy*-coordinates.

ACKNOWLEDGMENT

The authors would like to thank the following people for their contributions: Sven Müller, Steffen Burger, Thorsten Ochsenreither and Judy Lee-Wing.

REFERENCES

[1] F. Segor, A. Bürkle, M. Kollmann, and R. Schönbein, “Instantaneous Autonomous Aerial Reconnaissance for Civil Applications - A UAV based approach to support security and rescue forces,” The 6th International Conference on Systems ICONS 2011, January 23-28, 2011, St. Maarten, The Netherlands Antilles, pp.72-76, 2011.

[2] E. Shechtman and M. Irani, “Matching Local Self-Similarities across Images and Videos,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, pp. 1-8, 2007.

[3] B. Lucas and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision”, Proceedings of 7th International Joint Conference on Artificial Intelligence (IJCAI), pp. 674-679, 1981.

[4] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal on Computer Vision (IJCV), vol. 60, no. 2, pp. 91-110, 2004.

[5] P. Solbrig, D. Bulatov, J. Meidow, P. Wernerus, and U. Thönnessen, “Online Annotation of Airborne Surveillance and Reconnaissance Videos,” The 11th International Conference on Information Fusion, Köln, Germany, pp. 1131-1138, 2008.

[6] D. Bulatov, “Towards Euclidean Reconstruction from Video Sequences,” Int. Conf. Computer Vision Theory and Applications (2), pp. 476-483, 2008.

[7] D. Bulatov, P. Wernerus, and C. Heipke, “Multi-view Dense Matching Supported by Triangular Meshes”, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, no. 6, pp. 907–918, 2011.

[8] Universität der Bundeswehr München, Germany, http://www.unibw.de/lrt13_2/Forschung/Projekte/UAVPF, 18.01.2012.

[9] K. Daniel, B. Dusza, A. Lewandowski, and C. Wietfeld, “AirShield: A System-of-Systems MUAV Remote Sensing Architecture for Disaster Response,” IEEE International Systems Conference (SysCon), Vancouver, pp. 196-200, 2009.

[10] K. Daniel, B. Dusza, and C. Wietfeld, “Mesh Network for CBRNE Reconnaissance with MUAV Swarms,” 4th Conference on Safety and Security Systems in Europe, Potsdam, 2009.

[11] V. Lepetit, F. Moreno-Noguer, and P. Fua: “EPnP: An Accurate O(n) Solution to the PnP Problem”. International Journal of Computer Vision 81(2), pp. 155-166, 2009.

[12] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes: A Comparison of Similarity Measures for Use in 2D-3D Medical Image Registration. IEEE Trans. Med. Imaging 17(4), pp. 586-595, 1998.

[13] D. Scharstein and R. Szeliski. "A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms". International Journal of Computer Vision, 47(1), pp.7-42, 2002.

[14] C. Strecha: "Multi-view Stereo as an Inverse Inference Problem", PhD Dissertation, KU Leuven, Belgium, 2007.

[15] A. Bürkle, F. Segor, and M. Kollmann, "Towards Autonomous Micro UAV Swarms," Proceeding of the International Symposium on Unmanned Aerial Vehicles, Dubai, UAE, 2010.

[16] B. N. Delaunay, "Sur la sphere vide," In: Bulletin of Academy of Sciences of the USSR 7, No 6, pp. 793-800, 1934.

[17] D. Hearn and M. P. Baker, "Computer Graphics, C version, 2nd Ed," Prentice Hall, 1997.

[18] D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook, "The Traveling Salesman Problem. A Computational Study," Princeton University Press, Februar 2007.

[19] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2000.

[20] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Communications of the ACM, vol. 24, no. 6, pp. 381-395, 1981.

[21] J. Matas and O. Chum, "Randomized RANSAC with Td,d –test", Proceedings of the British Machine Vision Conference (BMVA), vol. 2, pp. 448-457, 2002.

[22] P. A. Viola and W. M. Wells, "Alignment by Maximization of Mutual Information," International Journal of Computer Vision (IJCV), 24(2), pp. 137--154, 1999.

[23] C. Wu, "A GPU Implementation of Scale Invariant Feature Transform (SIFT)," http://cs.unc.edu/~ccwu/siftgpu, 18.01.2012.

[24] C. G. Harris and M. J. Stevens, "A Combined Corner and Edge Detector," Proc. of 4th Alvey Vision Conference, pp. 147–151, 1998.

[25] Esri Enterprise, USA, http://www.esri.com, 18.01.2012.