Reconnaissance and Camp Security Missions with an Unmanned Aerial Vehicle (UAV) at the 2009 European Land Robots Trials (ELROB)

Andreas Birk
Jacobs University Bremen
Campus Ring 1
28759 Bremen, Germany
a.birk@jacobs-university.de
http://robotics.jacobs-university.de

Burkhard Wiggerich
AirRobot GmbH & Co. KG
Werler Strasse 4-8
59755 Arnsberg, Germany
info@airrobot.de
http://www.airrobot.de

Heiko Bülow, Max Pfingsthorn, and Sören Schwertfeger
Jacobs University Bremen

Abstract — Results from two response missions at the 2009 European Land Robot Trials (ELROB) are presented. An Unmanned Aerial Vehicle (UAV) in form of an Airrobot AR100-B is used in a reconnaissance and in a camp security scenario. The UAV is capable of autonomous waypoint navigation using onboard GPS processing. A digital video stream from the vehicle is used to create photo maps - also known as mosaicking - in real time at the operator station. This mapping is done using an enhanced version of Fourier Mellin based registration, which turns out to be very fast and robust.

Keywords: *Unmanned Aerial Vehicle (UAV), Autonomy, Registration, Mosaicking, Visual SLAM*

I. Introduction

The 2009 European Land Robot Trials (ELROB) are the fourth event in an annual series of evaluation events for Safety, Security, and Rescue robotics (SSRR) [1]. ELROB took place from 14.-18. June 2009 in Oulu, Finland. This paper describes results from a joined participation of the German company Airrobot with the Jacobs Robotics Group. The team used an Airrobot AR100-B Unmanned Aerial Vehicle (UAV) in two SSRR missions, namely reconnaissance and camp security. This quadrocopter (figure 1) is capable of onboard processing of GPS signals and of stored way points, hence allowing fully autonomous operations. It is also equipped with a high quality digital video link to the operator station.

Data from the video link is used for photo mapping - also known as mosaicking - by registering the video frames with each other and hence providing a larger overview. These photo maps serve two main purposes. First, they help the operator in the mission: they stabilize the video stream, which is quite shaky, and they provide much larger overview than a video. Second, they are an important mission deliverable. Even though there are many Geo Information Systems (GIS) available including widespread popular ones like Google Maps [2] and Microsoft Bing Maps [3], they have severe limitations. First of all, there are still surprisingly many regions that are not or only poorly covered by certain systems. Google Maps has for example only an extremely low resolution coverage of

Fig. 1. The Airrobot AR100-B taking off for a mission at ELROB 2009.

Oulu and especially the competition site (figure 2). Also, GIS may provide inaccurate information (figure 3) and it is even in the best case less precise than what can be achieved with a low flying UAV. But the most important motivation for real time photo mapping is that it provides an actual overview of the mission site at the time of the mission, in contrast to the older data stored in a GIS.

Naively, the generation of a photo map is trivial as long as the exact pose of the vehicle is know. But GPS localization is very imprecise, the 3D orientation of the UAV is hard to determine, and so on. An alternative is to use registration of subsequent frames from a camera on the vehicle. Related work in the areas of visual odometry and visual Simultaneous Localization and Mapping (SLAM) predominantly uses feature based approaches based for example on the scale invariant feature transform (SIFT) [4], [5] or the Kanade-Lucas-Tomasi Feature Tracker (KLT) [6]; this holds in particular for closely related work dealing with UAVs and photo mapping [7][8]. Here, not only local descriptors but the whole information within the images is used, which seems to be very beneficial in the context of photo mapping as described in more detail in [9]. An variant of the Fourier Mellin Invariant (FMI) transform for image representation and processing [10][11]

Fig. 2. The ELROB 2009 competition site is only poorly covered in Google maps; the above image shows the best available resolution of the area.

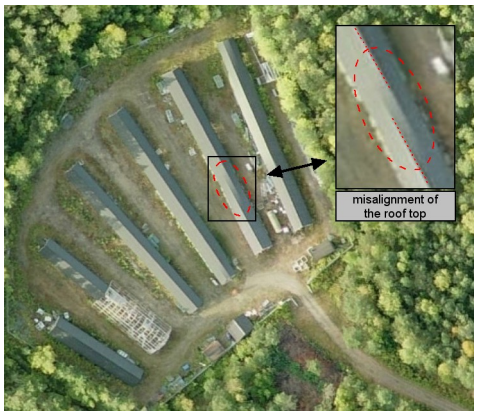

Fig. 3. Even professional map services are far from perfect. The above image shows an obvious bug in the coverage of the ELROB reconnaissance site by Microsoft Bing Maps.

is used with two significant modifications. First, a logarithmic representation of the spectral magnitude of the FMI descriptor is used. Second, a filter on the frequency where the shift is supposed to appear is applied. An analysis of the benefits of this improved FMI (iFMI) is provided in [9]. The iFMI provides 2D translations, rotations, and scaling, i.e., registrations when the UAV is moving horizontally, rotating its yaw, and changing its altitude. In addition, it requires in theory a strictly down looking camera and a flat world hypothesis is assumed. But as shown in the experiments, these constraints do not have to be strictly fulfilled to achieve meaningful results.

II. Improved Fourier Mellin Mapping

The basis for the 2D signal registration in this approach is a Phase-Only Matched Filter (POMF). This correlation approach makes use of the fact that two shifted signals having the same spectrum magnitude are carrying the shift information within its phase (equ.1).

A. Principles of the Fourier Mellin transform

The principle of phase matching is now extended to additionally determine affine parameters like rotation, scaling and afterward translation.

$$f(t-a) \circ\!-\!\bullet\, F(\omega)e^{i\omega a} \tag{1}$$

When both signals are periodically shifted the resulting inverse Fourier transformation of the phase difference of both spectra is actually an ideal Dirac pulse. This Dirac pulse indicates the underlying shift of both signals which have to be registered.

$$d(t-a) \circ\!-\!\bullet\, 1e^{i\omega a} \tag{2}$$

The resulting shifted Dirac pulse deteriorates with changing signal content of both signals. As long as the inverse transformation yields a clear detectable maximum this method can be used for matching two signals. This relation of the two signals phases is used for calculating the Fourier Mellin Invariant Descriptor (FMI). The next step for calculating the desired rotation parameter exploits the fact that the 2D spectrum (equ.4) rotates exactly the same way as the signal in the time domain itself (equ.3):

$$s(x,y) = r[(x\cos(\alpha) + y\sin(\alpha)), (-x\sin(\alpha) + y\cos(\alpha))] \tag{3}$$

$$\mid S(u,v) \mid = \mid R[(u\cos(\alpha)+v\sin(\alpha)), (-u\sin(\alpha)+v\cos(\alpha))] \mid \tag{4}$$

where α is the corresponding rotation angle.

For turning this rotation into a signal shift the magnitude of the signals spectrum is simply re-sampled into polar coordinates. For turning a signal scaling into a signal shift several steps are necessary. The following Fourier theorem

$$f(\frac{t}{a}) \circ\!-\!\bullet\, aF(a\omega) \tag{5}$$

shows the relations between a signal scaling and its spectrum. This relation can be utilized in combination with another transform called Mellin transform which is generally used for calculations of moments:

$$V^M(f) = \int_0^\infty v(z) z^{i2\pi f-1} dz \tag{6}$$

Having two functions $v1(z)$ and $v2(z) = v1(az)$ differing only by a dilation the resulting Mellin transform with substitution $az = \tau$ is:

$$\begin{aligned} V_2^M(f) &= \int_0^\infty v1(az) z^{i2\pi f-1} dz \\ &= \int_0^\infty v1(\tau)(\frac{\tau}{a})^{i2\pi f-1}\frac{1}{a}d\tau \\ &= a^{-i2\pi f} V_1^M(f) \end{aligned} \tag{7}$$

The factor $a^{-i2\pi f} = e^{-i2\pi f ln(a)}$ is complex which means that with the following substitutions

$$\begin{aligned} z = e^{-t}, ln(z) = -t, dz = -e^{-t}dt, \\ z \to 0 \Rightarrow t \to \infty, z \to \infty \Rightarrow t \to -\infty \end{aligned} \tag{8}$$

the Mellin transform can be calculated by the Fourier transform with logarithmically deformed time axis:

$$\begin{aligned} V^M(f) &= \int_\infty^{-\infty} v(e^{-t}) e^{-t(i2\pi f-1)}(-e^{-t})dt \\ &= \int_{-\infty}^{\infty} v(e^{-t}) e^{-i2\pi f t} dt \end{aligned} \tag{9}$$

Now the scaling of a function/signal using a logarithmically deformed axis can be transfered into a shift of its spectrum. Finally, the spectrum's magnitude is logarithmically re-sampled on its radial axis and concurrently the spectrum is arranged in polar coordinates exploiting the rotational properties of a 2D Fourier transform as described before. Scaling and rotation of an image frame are then transformed into a 2D signal shift where the 2D signal is actually the corresponding spectrum magnitude of the image frame. This intermediate step is called the FMI descriptor.

Here, a sketch of the overall algorithm. The POMF is calculated as follows:

1) calculate the spectra of two corresponding image frames
2) calculate the phase difference of both spectra
3) apply an inverse Fourier transform of this phase difference

The following steps are taken for a full determination of the rotation, scaling and translation parameters:

1) calculate the spectra of two corresponding image frames
2) calculate the magnitude of the complex spectral data
3) resample the spectra to polar coordinates
4) resample the radial axes of the spectra logarithmically
5) calculate a POMF on the resampled magnitude spectra
6) determine the corresponding rotation/scaling parameters from the Dirac pulse
7) re-size and re-rotate the corresponding image frame to its reference counterpart
8) calculate a POMF between the reference and re-rotated/scaled replica image
9) determine the corresponding x,y translation parameters from the Dirac pulse

The steps are used in the Fourier Mellin based Mapping in a straightforward way. A first reference image I_0 is acquired or provided to define the reference frame F and the initial robot pose p_0. Then, a sequence of images I_k is acquired. Image I_1 is processed with the above calculations to determine the transformations T_0^M between I_0 and I_1 and hence the motion of the robot. The robot pose is updated to p_1 and I_1 is transformed by according operations T_0^F to an image I_1' in reference frame F. The transformed image I_1' is then added to the photo map. From then on, the image I_n', i.e., the representation of the previous image in the photo map, is used to determine the motion-transformations T_n^M in the subsequent image I_{n+1}, which is used to update the pose p_{n+1} and the new part I_{n+1}' for the photo map.

B. Improvements for more registration stability

In order to make the registration more stable the search area within the FMI descriptor the search area for scaling and rotation is restricted to reasonable parameters between two image frames. As a second measure the FMI descriptor is in addition to a standard spectral window processed by a logarithmic function in order to suppress low frequencies which often lead to misregistrations. Furthermore the domain after the inverse Fourier transform is processed by a FIR interpolation filter to suppress interference from aliasing image content and lift the energy from a correct registration peak which can be distributed around several frequency cells. A more detailed discussion of the approach can be found in [9].

scenario	autonomy (mm:ss)	overall (mm:ss)
reconnaissance	11:00	22:00
camp security	20:00	57:30

TABLE I

THE AUTONOMOUS FLIGHT TIMES OF THE UAV. PLEASE NOTE THAT THE OVERALL MISSION TIMES INCLUDE THE SET UP AND THE TIME FOR THE PROVISION OF THE MISSION DELIVERABLES LIKE MAPS, PHOTOS OF THE INTRUDERS, ETC. TO THE ORGANIZERS.

	reconnaissance			camp security	
fig.	6	7	8	11	12
#frames	31	52	87	337	784

TABLE II

NUMBER OF FRAMES IN THE EXAMPLE MAPS

III. RESULTS

As discussed in a bit more detail later on, the UAV performed significant amounts of time in autonomous mode (table I) despite challenging weather conditions. The main limitation in the photo mapping are error frames in the video (figure 4), which could not be registered. When an error frame occurred, the generation of a new photo map was started. We are currently working on an automatic detection of error frames to be able to ignore them and to try to register the first new correct frame with the last known good frame. The photo mapping took place on a Laptop with a Core-2 Duo 2.0 GHz CPU running Ubuntu 9.04 Linux.

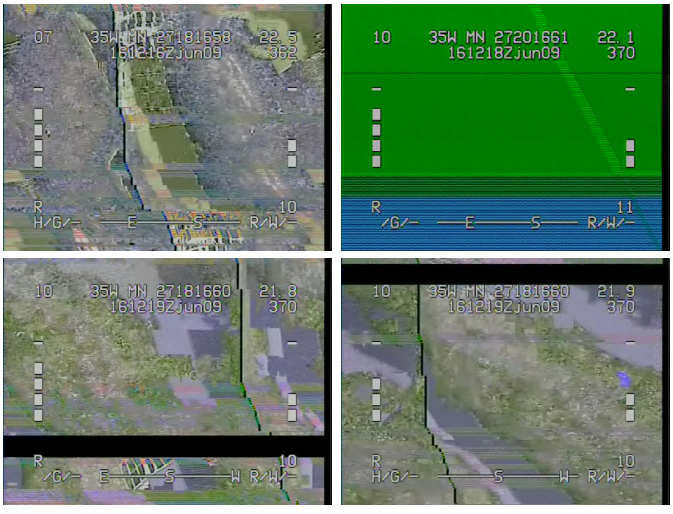

Fig. 4. The digital video link from the UAV had to cover up to 300 m through forest under humid weather conditions. There were occasional video errors despite using a high quality, military grade RF connection. The above images show some typical examples.

A. Reconnaissance

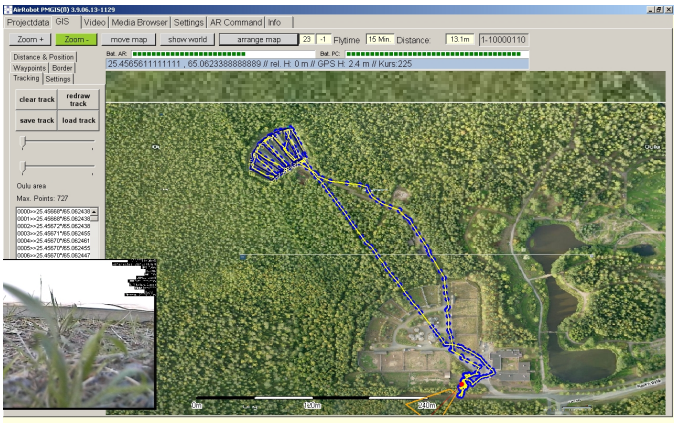

Fig. 5. The waypoints in the reconnaissance path

The mission task for the reconnaissance mission was to detect hazardous spots in an Nuclear-Biological-Chemical (NBC) incident scenario. The hazardous spots were marked by Emergency Response Intervention Cards or short ERI-cards. ERI-cards are rectangular plates of about 40 cm $\times$ 30 cm, with black number codes for different hazardous materials on orange background.

The main site for the reconnaissance operations was about 360 m away from the operator's station. The site itself spans about 60 m by 60 m (figure 5). The UAV used GPS way points to autonomously reach the site and to explore it.

The UAV operated fully autonomously for 11 minutes in total out of 22 minutes mission time (table I). Please note that the mission time includes set up and the provision of mission deliverables to the organizers. There were only about 4 minutes of manual intervention during the actual flight time of about 15 minutes. The manual operations included start and landing as well as the upload of a second set of preplanned GPS way points to pursue a different search strategy. The vehicle was capable of autonomous operation under difficult wind conditions; wind gusts in the air exceeded 12 m/sec. The UAV did not detect any of the ERI-cards. But also 5 out of the 6 other teams - all of them using land robots - did not find any ERI-cards either. Only 1 team out of the 7 teams managed to find 2 ERI-cards.

The mission scenario involved significant communication challenges for the video transmission. The maximum distance was about 400 m with significant amounts of vegetation, namely forest in humid weather conditions, between the vehicle and the operator's station. There were hence quite often error frames in the video data stream. As mentioned, we are currently working on an automatic detection of error frames to prevent a disruption of the map generation. Typical error-free video sequences consisted of about 75 frames. Figures 6 and 7 show two examples with 31, respectively 52 frames. It has to be noted that the approach also gives reasonable results when the camera is tilted though it should fail in theory under such conditions. Figure 8 shows an according example based on 87

Fig. 6. An example of a photo map in the reconnaissance scenario (31 frames).

frames. The maps were generated in real time at the operator's station with about 50 msec processing time for registration and visualization, i.e., 20 Hertz update rate.

B. Camp Security

The mission task for the camp security was to detect intruders. The operator's station was located about 30 m from the camp site, which spans about 90 m by 90 m (figure 9). The UAV mainly autonomously hovered over the site and intruders were "chased" in manual mode by the operator when being detected. The overall mission time was 60 minutes including set up and the provision of mission deliverables in the end.

In total, 6 intruders out of 8 possible ones were detected and "caught" by the UAV. Two examples of detected intruders are shown in figure 10. Only a 2nd team out of the 7 teams in total - all other 6 teams using land robots - was also able to catch 6 intruders. One other team caught 2, three teams caught 1, and one team caught 0 intruders. The 2 intruders that were missed entered the camp while the UAV was for a 5 minute stop at the operator's station for swapping the batteries - as mentioned before, the maximum operation time of the UAV is about 25 to 30 minutes, i.e., batteries had to be changed once during this mission.

As shown in table I, the UAV operated fully autonomously for 20 minutes in total out of 57.5 minutes flight time, i.e., for about 34.8% of the time. The wind conditions were also quite challenging during this mission, the wind speed on ground was

Fig. 7. A second example of a photo map in the reconnaissance scenario (52 frames).

Fig. 8. An example of a photo map of the way of the UAV to the reconnaissance site (87 frames). Please note that there is a significant tilt of the camera and hence skew in the images. Though the algorithm is in theory not capable of registering under these conditions, the result is nevertheless quite usable of a rough orientation.

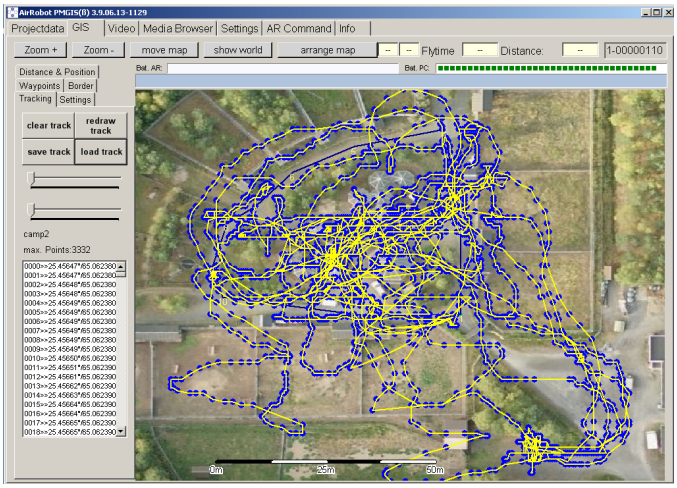

Fig. 9. The actual path of the UAV in the camp security scenario based on GPS track points.

Fig. 10. Two examples of intruders detected during the camp security mission.

Fig. 11. An example of a photo map of the camp generated from 337 video frames in real time.

Fig. 12. A second example of a photo map of the camp generated from 784 video frames in real time.

measured with up to 7.2 m/sec, wind gusts in the air exceeded 12 m/sec.

The mission scenario involved shorter communication ranges, less than 150 m maximum, and fewer vegetation, especially few trees in the line of sight. But the ”camp” - being a part of the former Oulu Zoo - contains a significant amount of metal structures, especially fences and cages with wire meshes. Nevertheless, the video link was much less error prone than for the reconnaissance mission. The video sequences between error frames were hence significantly longer. On average, video sequences with 650 frames were turned into photo maps. Figures 11 and 12 show two examples with 337, respectively 782 frames. The maps were generated in real time at the operator's station with about 50 msec processing time for registration and visualization, i.e., 20 Hertz update rate.

IV. Conclusions

Results from two SSRR missions with an Unmanned Aerial Vehicle (UAV) at the 2009 European Land Robot Trials (ELROB) were presented. An Airrobot AR100-B was used in a reconnaissance and a camp security scenario. The UAV is capable of GPS based way point navigation, which was used for significant amounts of autonomous operations despite very challenging weather conditions with gusts of wind with top speeds of more than 12 m/sec. A digital video link was used for photo mapping in real time. An improved version of Fourier Mellin based registration with two core modifications is used for this purpose. First, a logarithmic representation of the spectral magnitude of the FMI descriptor is used. Second, a filter on the frequency where the shift is supposed to appear is applied. The resulting photo maps provide high resolution, up to date information about the mission site.

References

[1] ELROB, “European Land-Robot Trial (ELROB),” www.elrob2009.org, 2009.

[2] Google, “Google Maps,” http://maps.google.com, 2009.

[3] Microsoft, “Bing Maps,” http://www.bing.com/maps/, 2009.

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

[5] ——, “Object Recognition from Local Scale-Invariant Features,” in *Proceedings of International Conference on Computer Vision*, 1999, pp. 1150–1157.

[6] J. Shi and C. Tomasi, “Good features to track,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR94)*, 1994.

[7] B. Steder, G. Grisetti, C. Stachniss, and W. Burgard, “Visual SLAM for Flying Vehicles,” *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1088–1093, 2008.

[8] A. Angeli, D. Filliat, S. Doncieux, and J. A. Meyer, “2d simultaneous localization and mapping for micro aerial vehicles,” in *Proceedings of the European Micro Aerial Vehicles (EMAV 2006) conference*, 2006.

[9] H. Buelow and A. Birk, “Fast and Robust Photomapping with an Unmanned Aerial Vehicle (UAV),” in *International Conference on Intelligent Robots and Systems (IROS)*. IEEE Press, 2009.

[10] Q.-S. Chen, M. Defrise, and F. Deconinck, “Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 12, pp. 1156–1168, 1994.

[11] B. Reddy and B. Chatterji, “An FFT-based technique for translation, rotation, and scale-invariant image registration,” *Image Processing, IEEE Transactions on*, vol. 5, no. 8, pp. 1266–1271, 1996.

Birk, A., B. Wiggerich, H. Bülow, M. Pfingsthorn, and S. Schwertfeger, "Reconnaissance and Camp Security Missions with an Unmanned Aerial Vehicle (UAV) at the 2009 European Land Robots Trials (ELROB)", IEEE International Workshop on Safety, Security, and Rescue Robotics (SSRR): IEEE Press, 2009.

http://dx.doi.org/10.1109/SSRR.2009.5424163