Creating Photo Maps
with an Aerial Vehicle in USARsim

Heikow Bülow, Andreas Birk, and Shams Feyzabadi

Jacobs University Bremen*
Campus Ring 1, 28759 Bremen, Germany

Abstract. Photo maps, i.e., 2D grids that provide a large scale bird's eye view of the environment, are of interest for many application scenarios but especially for safety and security missions. We present a very efficient and robust algorithm for this task, which only uses registration between consecutive images, i.e., it does not require any localization. The algorithm is benchmarked in USARsim, where the video stream of a down-looking camera of an aerial vehicle, namely a blimp, is used to generate a large scale photo map.

1 Introduction

Unmanned Aerial Vehicles (UAV) are promising tools for a fast first situation assessment in Safety, Security, and Rescue Robotics (SSRR). They are ideal eyes in the sky that cooperate with Unmanned Ground Vehicles (UGV), which provide the advantages of e.g. higher payloads and of options for mobile manipulation. Figure 1 shows a typical application scenario where a Jacobs land robot and an UAV cooperating at European Land Robotics Trials (ELROB) 2007 in Monte Ceneri, Switzerland. This scenario, in which the Jacobs team won the 1st place, required that the UAV detects hazard spots like seats of fire, which the land robot has to reach [1]. The online generation of a detailed aerial photo map by the UAV is of obvious interest in according SSRR missions.

Precise localization of UAVs is very challenging for several reasons. First of all, they typically only have a limited payload for sensors. Second, commonly found solutions like GPS receivers in combination with compasses only provide coarse pose information, which is not sufficient to fuse a sequence of images into a usable photo map. An alternative approach is to completely omit the problem of localization through additional sensors, and to only use the information in two sequentially acquired images to determine the robots movement between the two frames.

There is hence related work in the computer vision community. Optical flow techniques [2,3,4,5,6,7] are also targeted at motion estimation, but they are best suited for only estimating robot states for control, especially for aerial vehicles [8]. The classic optical flow techniques namely suffer from too large errors when the state is integrated for proper localization. An other line of research is related

---

* Formerly International University Bremen.

Fig. 1. A Jacobs land robot and an aerial robot cooperating in one of the scenarios at the European Land Robot Trials 2007 (ELROB) in Monte Ceneri, Switzerland. The aerial robot has to find and locate hazard spots like seats of fire, which the land robot then has to reach.

to computer vision work on structure from motion [9,10,11,12]. The methods are characterized by various application constraints, especially on the camera motions, as they were developed for specialized 3D model acquisitions. Here, an unconstrained integration of sequential 2D images is investigated.

So, the task here is that regions of overlap between two consecutively acquired images have to be found and suitably matched. This process of finding a template in an image is also known as registration [13,14,15,16,17,18,19]. But the task at hand is more difficult than mere registration as the region of overlap is unknown and it usually has undergone non-trivial transformations due to the robot's movements. This is comparable to image stitching [20], which is for example used to generate panoramic views from several overlapping photographs. We used this idea of employing stitching methods already before for merging 2D occupancy grid maps generated by multiple robots [21]. The scale invariant feature transform (SIFT) introduced by Lowe [20,22] is at present the most popular basis for image stitching. SIFT delivers point-wise correspondences between distinctive, non-repetitive loc al features in the two images. The number of detected features is significantly smaller than the number of pixels in the image. Other methods for identifying features include local image descriptors like intensity patterns [23,24] and the Kanade-Lucas-Tomasi Feature Tracker (KLT) [25]. Based on the Fourier Mellin transform for image representation and processing [26], we have developed an improved version that outperforms SIFT and KLT with respect to processing time and robustness as shown in experiments with real world images including aerial data.

This algorithm is applied here to generating photo maps and benchmarked in the Unified System for Automation and Robotics Simulation (USARsim) [27]. USARsim has the significant advantage that it provides a high fidelity physical simulation, i.e., realistic test data, while ground truth information of the robot is given [28,29,30,31]. This facilitates experimental validations, especially as supplements to field work where ground truth is not known.

The rest of this paper is structured as follows. A Fourier Mellin based approach to image registration is presented. It can be used to generate photo maps in a fast and robust way. In section 3, the approach is tested through experiments with a aerial vehicle in USARsim. Due to the availability of ground truth data, a quantitative assessment of strengths and weaknesses is possible. Section 4 concludes the paper.

2 Photo Mapping with Improved Fourier Mellin Registration

First, some basic terms and concepts are introduced on which our algorithm for photo mapping is based. The classical Matched Filter (MF) of two 2D signals $r * (-x, -y)$ and $s(x, y)$ is defined by:

$$q(x, y) = \int \int_{-\infty}^{\infty} s(a, b) r * (a - x, b - y) da db \qquad (1)$$

This function has a maximum at (x0,y0) that determines the parameters of a translation. One limitation of the MF is that the output of the filter primarily depends on the energy of the image rather that on its spatial structures. Furthermore, depending on the image structures the resulting correlation peak can be relatively broad. This problem can be solved by using 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.2). Furthermore the POMF calculation is much faster than the MF because if a signal frame of size 2^N is used, the advantages of the Fast Fourier Transform (FFT) can be exploited.

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

$$f(t - a) \circ - \bullet F(\omega) e^{i\omega a} \qquad (2)$$

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 registered.

$$d(t - a) \circ - \bullet 1 e^{i\omega a} \qquad (3)$$

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 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 5 rotates exactly the same way as the signal in the time domain itself (equ.4):

$$s(x, y) = r[(x \cos(\alpha) + y \sin(\alpha)), (-x \sin(\alpha) + y \cos(\alpha))] \qquad (4)$$

$$| S(u, v) |=| R[(u\cos(\alpha) + v\sin(\alpha)), (-u\sin(\alpha) + v\cos(\alpha))] | \qquad (5)$$

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

$$f(\frac{t}{a}) \circ - \bullet aF(a\omega) \qquad (6)$$

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:

$$V^M(f) = \int_0^\infty v(z)z^{i2\pi f-1}dz \qquad (7)$$

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

$$V_2^M(f) = \int_0^\infty v1(az)z^{i2\pi f-1}dz$$
$$= \int_0^\infty v1(\tau)(\frac{\tau}{a})^{i2\pi f-1}d\tau$$
$$= a^{-i2\pi f}V_1^M(f) \qquad (8)$$

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

$$z = e^{-t}, ln(z) = -t, dz = -e^{-t}dt,$$
$$z \to 0 \longrightarrow t \to \infty, z \to \infty \longrightarrow t \to -\infty \qquad (9)$$

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

$$V^M(f) = \int_\infty^{-\infty} v(e^{-t})e^{-t(i2\pi f-1)}(-e^{-t})dt$$
$$= \int_{-\infty}^\infty v(e^{-t})e^{-i2\pi ft}dt \qquad (10)$$

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.

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 photo 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.

3   Experiments and Results

Figures 2 and 3 show results of using the algorithm with real world UAV data for photo mapping. No localization information for the UAV is given, not even GPS data, only the raw video data is used. This real world data is well suited for a comparison with alternative approaches like SIFT, but only on the basis of a comparison of the performance between two consecutive images, i.e., single registration steps. In doing so, it can be shown that SIFT performs poorly on scenes with few distinct features and that it is in addition computationally expensive.

The real world data is much less suited for assessing the quality of the photo mapping approach in a quantitative way as ground truth information is not available. The significant advantage of the following experiments within USARsim is that this restriction does not hold there. The presented results were processed by

Fig. 2. An image map generated with iFMI in real-time from about 300 images acquired with an UAV. The scene involves several challenges, especially large featureless areas.

a MATLAB implementation. A first C/C++ implementation were tested with 4 different frame sizes: (192x192) = 150msec, (256x256) = 230 msec, (320x320) = 260 msec and (480x480) = 430 msec on a standard PC with a 1.7 GHz CPU. The processing times already include data acquisition and overview display using the INTEL OpenCV library.

As one can see in figure 7, the approach works also well in USARsim to generate interesting overviews of a scenario - like a disaster scene - from the video stream of an aerial vehicle. The visual quality of the photo map is fully sufficient for mission planning and other qualitative tasks. The next question is how the approach performs in quantitative terms. Figures 4, 5, and 6 show a direct comparison of the estimated versus the ground truth positions. The x-axis shows time in terms of steps for 180 consecutively acquired images. The y-axis shows absolute coordinates in meter, respectively pixels for global x- and y-coordinates. The scaling shown in figure 6 shows the estimated, respectively ground truth scaling effects of the elevation of the aerial vehicle above ground; the higher the vehicle flies above ground the smaller the images its camera delivers. As mentioned, the generated photo map has a sufficient quality for performing high level tasks with it. But the quantitative analysis with USARsim allows to study the effects of the cumulative errors in the localization. Concretely, it can be shown that there is a severe drift in the real versus the estimated position, i.e., that the approach has its limits in the current form for localization. This is of interest for further work
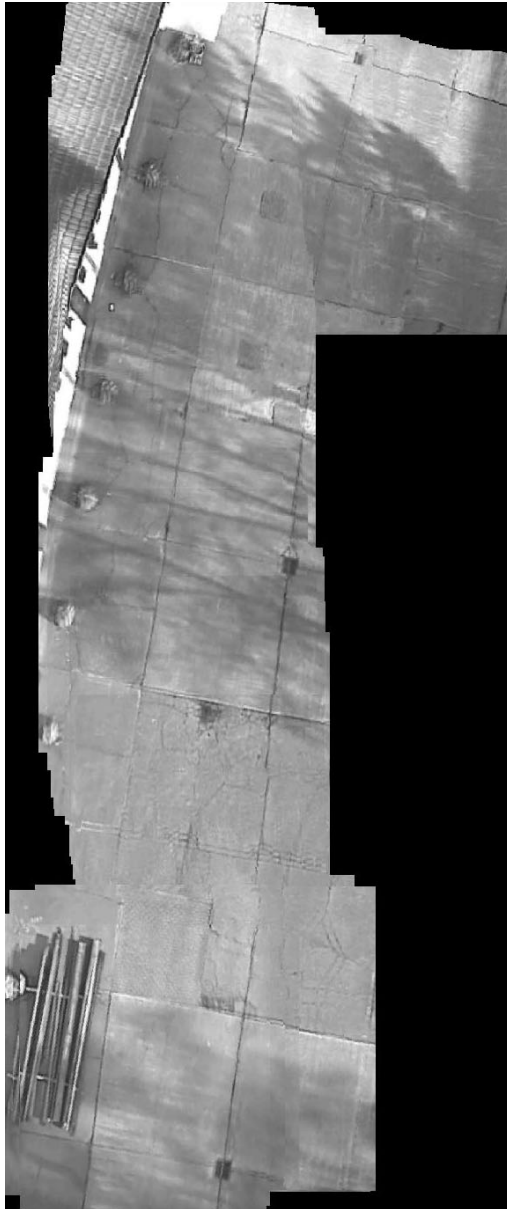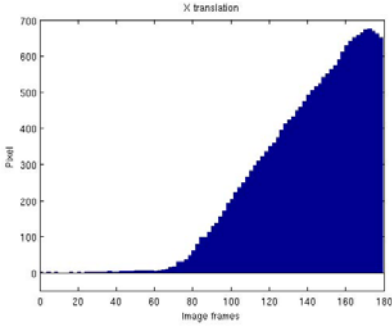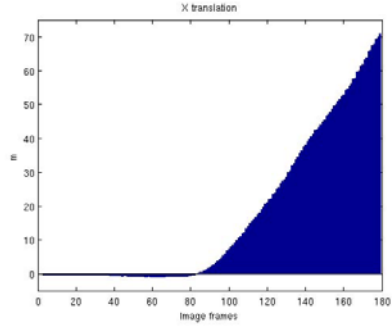
Fig. 3. About 600 areal images from an UAV are combined by iFMI in real-time into an image map
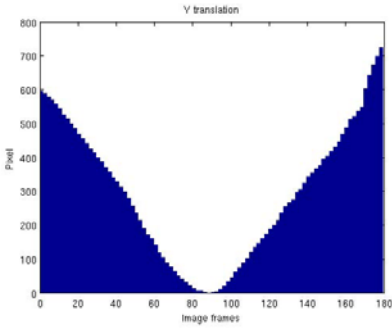
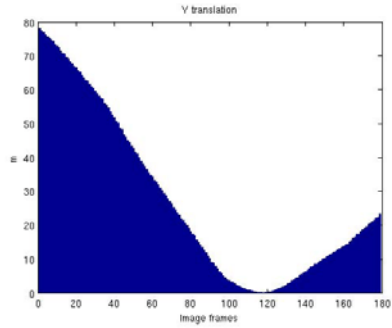(a) FM transform   (b) ground truth

Fig. 4. Comparison of the estimated and the real translations along the x-axis

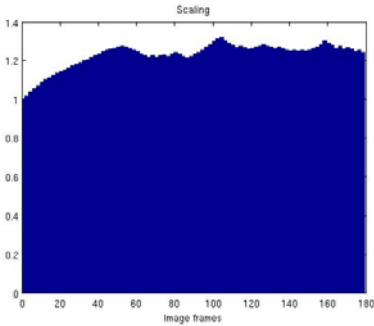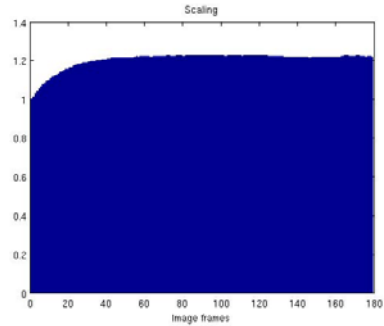(a) FM transform   (b) ground truth

Fig. 5. Comparison of the estimated and the real translations along the y-axis

(a) FM transform   (b) ground truth

Fig. 6. Comparison of the estimated and the real scaling, i.e., the effects of the elevation of the blimp above ground

Fig. 7. Photo map M_1 generated from a sequence of 180 images in USARsim from a blimp without known poses

along this line of research, especially for incorporating this new visual odometry approach into a Simultaneous Localization and Mapping (SLAM) framework.

4   Conclusion

A Fourier Mellin Transform based approach to photo mapping was presented. It is benchmarked with experiments with videos from a blimp in USARsim. This has the advantage over field experiments that the data is physically realistic while ground truth is known for a detailed analysis. The experiments show that the approach is robust and fast. But the ground truth comparisons in USARsim also reveal limits of the approach, especially with respect to robot localization. While the photo maps are usable and provide an interesting overview of for example an incident scenario, the ground truth analysis shows that there is a clear drift in the localization estimation.

References

1. ELROB: European Land-Robot Trial, ELROB (2007),
   `http://www.elrob2007.org`
2. Bruhn, A., Weickert, J., Feddern, C., Kohlberger, T., Schnoerr, C.: Variational optical flow computation in real time. IEEE Transactions on Image Processing 14, 608–615 (2005)
3. Kohlberger, T., Schnoerr, C., Bruhn, A., Weickert, J.: Domain decomposition for variational optical-flow computation. IEEE Transactions on Image Processing 14, 1125–1137 (2005)
4. Liu, H., Hong, T., Herman, M., Camus, T., Chellappa, R.: Accuracy vs efficiency trade-offs in optical flow algorithms. Computer Vision and Image Understanding 72, 271–286 (1998)
5. McCane, B., Galvin, K.N.: On the evaluation of optical flow algorithms. In: Proceedings of the 5th International Conference On Control, Automation, Robotics and Vision, Singapore, pp. 1563–1567 (1998)
6. Beauchemin, S., Barron, J.: The computation of optical flow. ACM Computing Surveys (CSUR) 27, 433–466 (1995)
7. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence 17, 185–203 (1981)
8. Kehoe, J., Causey, R., Arvai, A., Lind, R.: Partial aircraft state estimation from optical flow using non-model-based optimization. In: American Control Conference, invited paper, Minneapolis, MN, pp. 2868–2873 (2006)
9. Fitzgibbon, A., Zisserman, A.: Automatic camera recovery for closed or open image sequences. In: Burkhardt, H.-J., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, pp. 311–318. Springer, Heidelberg (1998)
10. Azarbayejani, A., Pentland, A.P.: Recursive estimation of motion, structure, and focal length. IEEE Transactions on Pattern Analysis and Machine Intelligence 17, 562–575 (1995)
11. Faugeras, O.: Three-dimensional computer vision – a geometric viewpoint (1993)
12. Adiv, G.: Inherent ambiguities in recovering 3-d motion and structure from a noisy flow field. IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 477–489 (1989)

13. Fitch, A., Kadyrov, A., Christmas, W., Kittler, J.: Fast robust correlation. IEEE Transactions on Image Processing 14, 1063–1073 (2005)

14. Stricker, D.: Tracking with reference images: a real-time and markerless tracking solution for out-door augmented reality applications. In: Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage, pp. 77–82. ACM Press, New York (2001)

15. Dorai, C., Wang, G., Jain, A.K., Mercer, C.: Registration and integration of multiple object views for 3d model construction. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(1), 83–89 (1998)

16. Brown, L.G.: A survey of image registration techniques. ACM Comput. Surv. 24(4), 325–376 (1992)

17. Alliney, S., Morandi, C.: Digital image registration using projections. IEEE Trans. Pattern Anal. Mach. Intell. 8(2), 222–233 (1986)

18. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings DARPA Image Understanding Workshop, pp. 121–130 (1981)

19. Pratt, W.: Correlation techniques of image registration. IEEE Transactions on Aerospace and Electronic Systems AES-10, 562–575 (1973)

20. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)

21. Birk, A., Carpin, S.: Merging occupancy grid maps from multiple robots. IEEE Proceedings, special issue on Multi-Robot Systems 94(7), 1384–1397 (2006)

22. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of International Conference on Computer Vision, pp. 1150–1157 (1999)

23. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. In: Proceedings of Computer Vision and Pattern Recognition (June 2003)

24. Gool, L.V., Moons, T., Ungureanu, D.: Affine/photometric invariants for planar intensity patterns. In: Proceedings of European Conference on Computer Vision (1996)

25. Shi, J., Tomasi, C.: Good features to track. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 1994 (1994)

26. Chen, Q., Defrise, M., Deconinck, F.: Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 16, 1156–1168 (1994)

27. USARsim: Unified System for Automation and Robotics Simulator, USARsim (2006), http://usarsim.sourceforge.net/

28. Carpin, S., Lewis, M., Wang, J., Balakirsky, S., Scrapper, C.: Bridging the gap between simulation and reality in urban search and rescue. In: Lakemeyer, G., Sklar, E., Sorrenti, D.G., Takahashi, T. (eds.) RoboCup 2006: Robot Soccer World Cup X. LNCS (LNAI), vol. 4434, pp. 1–12. Springer, Heidelberg (2007)

29. Carpin, S., Lewis, M., Wang, J., Balarkirsky, S., Scrapper, C.: USARSim: a robot simulator for research and education. In: Proc. of the 2007 IEEE Intl. Conf. on Robotics and Automation, ICRA (2007)

30. Carpin, S., Stoyanov, T., Nevatia, Y., Lewis, M., Wang, J.: Quantitative assessments of usarsim accuracy. In: Proceedings of PerMIS (2006)

31. Carpin, S., Birk, A., Lewis, M., Jacoff, A.: High fidelity tools for rescue robotics: results and perspectives. In: Bredenfeld, A., Jacoff, A., Noda, I., Takahashi, Y. (eds.) RoboCup 2005. LNCS (LNAI), vol. 4020, pp. 301–311. Springer, Heidelberg (2006)