The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Fast and Robust Photomapping with an Unmanned Aerial Vehicle (UAV)

Heiko Bülow and Andreas Birk

Abstract— **A fast and robust method for visual odometry based on the Fourier-Mellin Invariant (FMI) descriptor is presented. It extends previous FMI based approaches in two ways. 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. It is shown with experiments with an Unmanned Aerial Vehicle that this improved Fourier-Mellin Invariant (iFMI) method is is indeed an advancement and well suited for online visual odometry to generate large photo maps.**

I. Introduction

Photomaps, i.e., image based, metric representations are interesting for getting an overview of the environment where a mobile is operating in. From a simple viewpoint, regions of overlap between two consecutively acquired images have to be found and suitably matched for photomapping, which is related to visual odometry as it also allows to estimate the motion of the vehicle. This process of finding a template in an image is also known as registration [1], [2], [3], [4], [5], [6], [7]. 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 [8], which is for example used to generate panoramic views from several overlapping photographs.

The scale invariant feature transform (SIFT) [8], [9] is at present a very popular basis for image stitching. SIFT delivers point-wise correspondences between distinctive, non-repetitive local 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 [10], [11] and the Kanade-Lucas-Tomasi Feature Tracker (KLT) [12].

We postulate that when using the whole information in the images and not only features, uncertainties and ambiguities are minimized up to a level where they can be completely ignored. We hence base our work presented here on a variant of the Fourier Mellin transform for image representation and processing [13][14], which was for example used in the context of robotics for underwater photomapping [15]. In doing so, we introduce two significant modifications to achieve a very fast and robust method. 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 advantages are demonstrated by experiments with an Unmanned Aerial Vehicle (UAV).

The rest of this paper is structured as follows. In section II, the improved Fourier Mellin Invariant (iFMI) descriptor is introduced. Section III presents experiments with data from an Unmanned Aerial Vehicle (UAV). Section V concludes the paper.

Heiko Bülow and Andreas Birk are with the School of Engineering and Science, Jacobs University, 28759 Bremen, Germany; a.birk@jacobs-university.de, h.buelow@jacobs-university.de; http://robotics.jacobs-university.de.

II. The improved Fourier Mellin Invariant (iFMI) descriptor approach

The classical Matched Filter (MF) of two 2D signals $r*(-x,-y)$ and $s(x,y)$ is defined by:

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

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.

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

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.

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

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):

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

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

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

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

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:

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

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

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

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

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

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

$$\begin{aligned} V^M(f) &= \int_\infty^{-\infty} v(e^{-t}) e^{-t(i2\pi f - 1)} (-e^{-t}) dt \\ &= \int_{-\infty}^\infty v(e^{-t}) e^{-i2\pi f t} dt \end{aligned} \tag{10}$$

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 of the iFMI can be used for photomapping 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.

Fig. 1. 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.

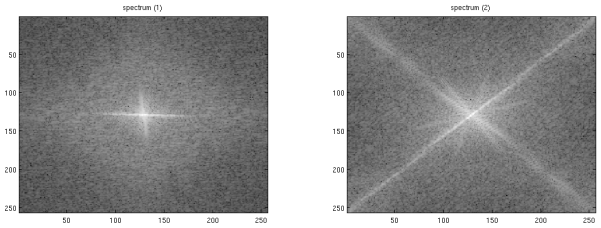

Fig. 3. Magnitudes of two image spectra

III. Experiments and Results

Figure 3 shows the corresponding spectra of an example image pair and figure 4 shows the polar/logarithmic resampled spectra from 0 to 180 degree from figure 3. The periodical shift of the image content on the x axis is clearly visible indicating the underlying rotation. With smaller extent, on the y axis a shift of the image content is also visible indicating the underlying scaling. Figure 5

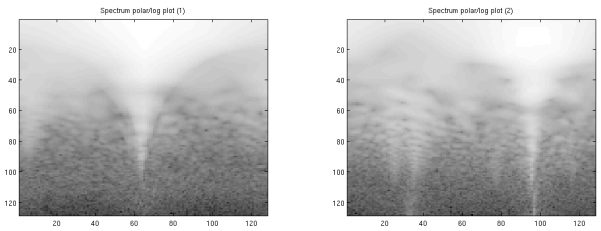

Fig. 4. Polar/logarithmic re-sampled spectra

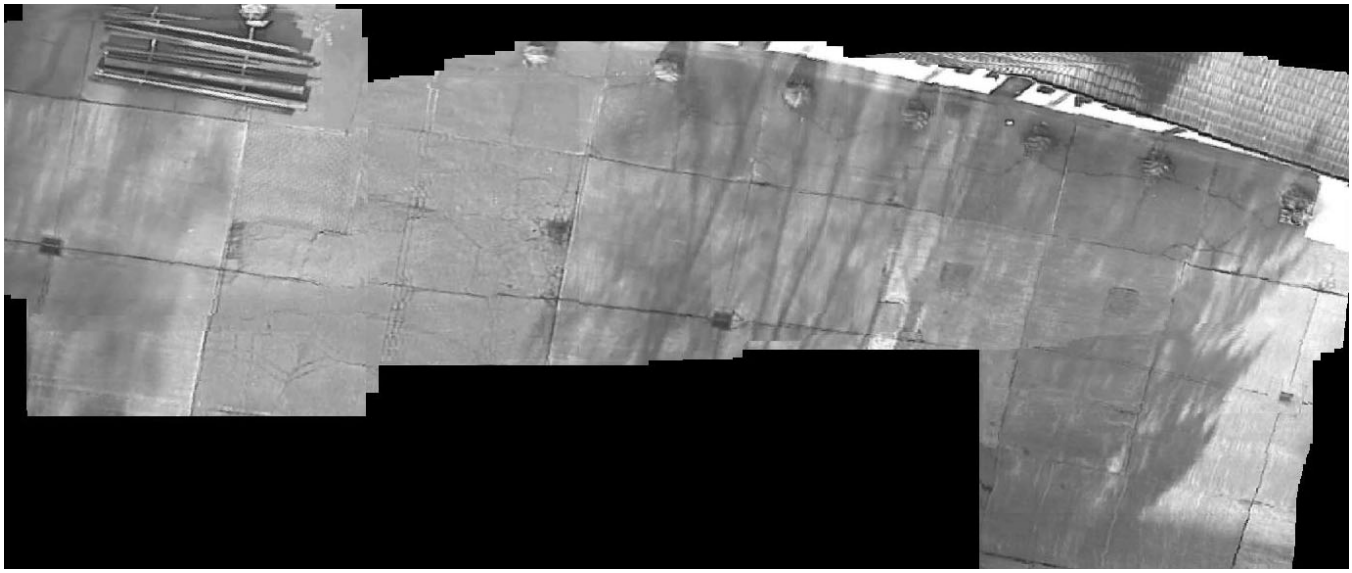

Fig. 2. About 600 areal images from an UAV are combined by iFMI in real-time into an image map.

shows the resulting 2D Dirac pulse from the POMF applied to the re-sampled spectrum magnitudes. Note the clearly distinguished peak, which is also postulated by equation 3. Rotation and scaling imply always change of the signal content in the sense that old image content is lost and that new content moves in, i.e., previously seen parts of the environment are not seen anymore and vice versa. The theorem 3 theoretically only holds for a complete signal which is periodically shifted to obtain an ideal Dirac pulse [16]. But as will be shown, the method robustly works with small overlaps in consecutive images.

The following results shows the progression of the resulting Dirac amplitude while shifting an image frame and its counterpart against each other. As mentioned, figure 5 shows the amplitude of the resulting Dirac progression for the determination of FMI parameter. Figure 6 shows the same pulse, but for a case where there is a significant translation and more than 30% of the original image content are "lost" in the new image. The amplitude of the corresponding Dirac pulse decreases from 0.15 to 0.087, but it is still clearly identifiable.

In the following experiment, x and y coordinates are shifted in a systematic way to illustrate the robustness but also the limits of our method. Synthetic image data showing White Gaussian noise is used for a worst case test, e.g., when the robot moves over a homogeneous concrete floor or a lawn. Note that the popular SIFT is known to perform very poorly in these cases as it requires distinctive, non-repetitive local features. A base image I_G is registered with different test images I_k. All images are 256x256 pixel. We consider here the worst case scenario of moving both in x- and y-direction at the same time. Each I_k is shifted one pixel in x and y further away from the image shown in I_G. The method works robustly up to the translation of 78 pixel in both x and y. The FMI parameters are always correctly determined. Figure 7 shows that the subsequent determination of the translation by the POMF computation also fails when exceeding this translation of 78 pixels in both directions. Experiments with real world data show that the detection of the parameters is possible with larger translational motions when there are highly distinguishable features in the image content. Nevertheless, it is obvious that the method

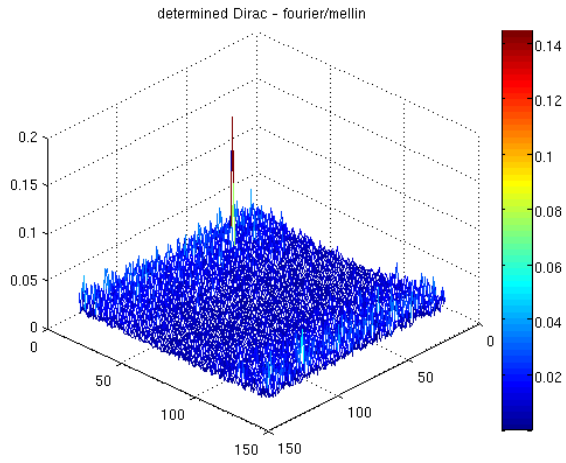

Fig. 5. The 2D Dirac pulse. As can be seen, the parameters of the transformations and hence robot movements are clearly identified.

fails when there is not sufficient overlap in consecutive images. When this worst case happens, the robot can start a new partial map, which can later on be combined with the first one through map merging [17]. But as shown in remainder of this section, this can be avoided with a decent image acquisition rate.

As default, we assume in the following an aerial robot with a down-looking camera. Horizontal translations lead to shifts in x, respectively y direction in subsequent images. A change in altitude causes a change in scaling between two frames. Roll and pitch are very well stabilized by gyroscopes and currently not taken into account; if data is to be gathered in extreme flight maneuvers, the gyroscope readings can be used to unwarp any shear from the images. The yaw is determined by image rotations. In the following, ground truth comparisons are presented. The top of figure 8 shows a sequence of vastly changing yaw values of the robot. The ground truth orientations are manually determined based on the image sequence. This extreme case is used to validate the

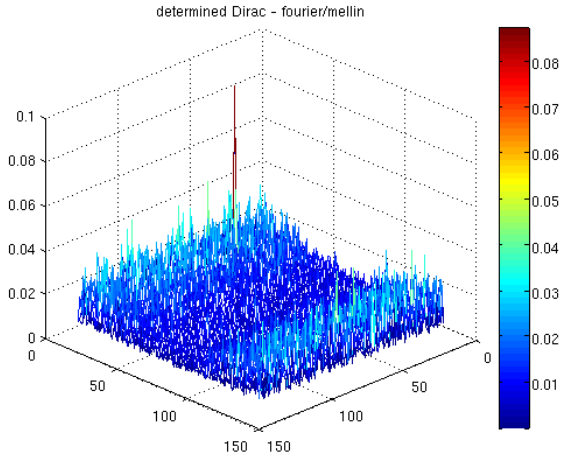

Fig. 6. The 2D Dirac pulse after rotation/scaling and a significant translation where more than 30% of the previous image content are lost. The amplitude of the pluse descreases, but it is still more than sufficient to determine the parameters of the underlying transforms.

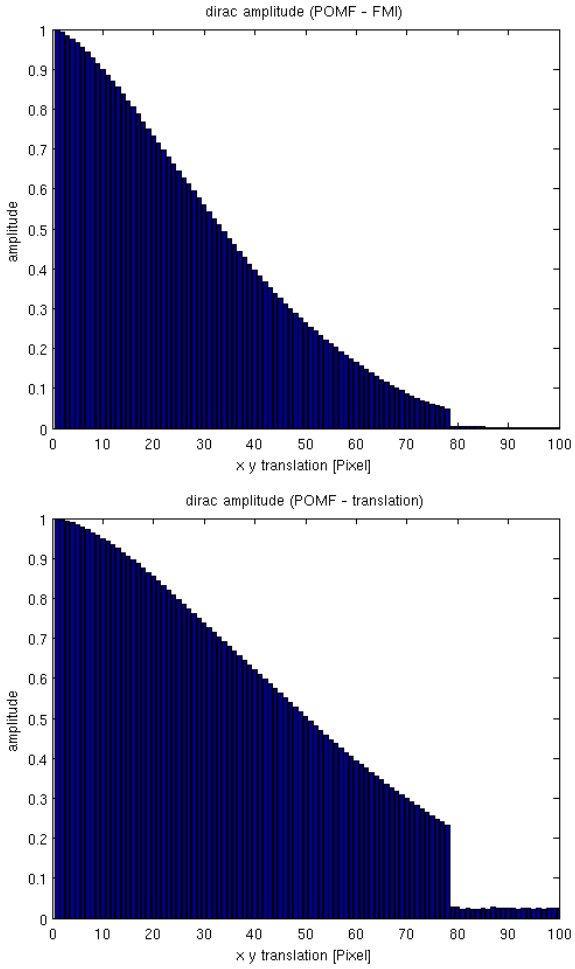

Fig. 7. The Dirac amplitudes when the overlap between registered images I_B and I_k is systematically decreased by shifting each I_k in x and y by one pixel. The top graph shows the amplitude for the FMI. The bottom graph shows the pulse for the subsequent determination of the translation. As can be seen, the method fails when the translation in both x- and y-direction exceeds 78 pixel. Note that this is a worst case analysis based on synthetic data.

robustness against rotations of the proposed algorithm. The bottom part of figure 8 shows the determined rotation parameters between subsequent image frames. Figure 9 shows the deviations between the true and the FM-SLAM orientation of the robot. As can be seen, the errors are even in this extreme case on average well below one degree.

A second ground truth experiment illustrates the performance on fast changing altitude, i.e., scale. Figure 10 shows the true change of scale between two image frames. Zooming into the scene (*scaling* > 1) means moving down in altitude and a *scaling* < 1 means moving up in altitude. Figure 11 shows the corresponding errors between the true and the determined parameters. The errors are in the order of at most 2% in the worst case and below 1% on average. If the robot is moving in about 2.5 m above ground, these values accordingly coorespond to 5 cm, respectively 2.5 cm.

Figures 12 and 13 show in addition to the ground truth evaluations some qualitative results.

The first image I_0 taken by the robot is the so to say seed of the map. When a new image I_n is taken, it is used to determine the underlying transformations and the new pose of the robot. This new pose is in turn used to transform the image I_n into an image I'_n such that it can be fused into the map. The qualitative result, namely the continuous image without any visible disruptions, nicely illustrates the quality of iFMI. An image map from a front or side looking camera is for example interesting when flying along the facade of a building.

Two other more typical examples from a down looking camera are shown in figures 12 and 13. Figure 12 shows an image map generated in a sequence including translations, rotations and the change in altitude. In figure 13, the robustness against altitude changes is illustrated in a qualitative way. The changes in scale between two subsequent images were in this sequence up to 25%.

Our C/C++ implementation of the approach is laid out for different frame sizes that allow to trade resolution for processing speed: (256x256) = 170 msec, (350x350) = 260 msec and (480x480) = 430 msec on a low-end PC with a single-core 1.7 GHz CPU. The processing times already include data acquisition and overview display using the INTEL OpenCV library. The field of view of the down-looking camera is about 90^o; one image can hence cover about 20 m wide ground when the robot is in a typical altitude of 10 m for aerial maps; even when the robot moves very fast with 1 m/s, an image acquisition rate of 0.2 Hz - allowing more than sufficient 5 sec processing time - still leads to consecutive images with significant overlap of 15 m or 75%.

IV. Extension to other application domains

The presented approach is also of interest for other application domains, especially underwater robotics [18]. Recent experiments include data from a vehicle passing over a cold water coral reef (figure 15)[1] The original aim of this video was to monitor the recovery of cold-water corals from the Tisler Reef. This reef lies along the Norway-Sweden border at a depth of 74 to 155 m. As shown in figure 14, the presented approach is also working well

[1]The video data was provided by Tomas Lundav, University of Gothenburg. This dataset was collected as part of the EU FP6 Hermes Project (GOCE-CT-2005-511234-1).

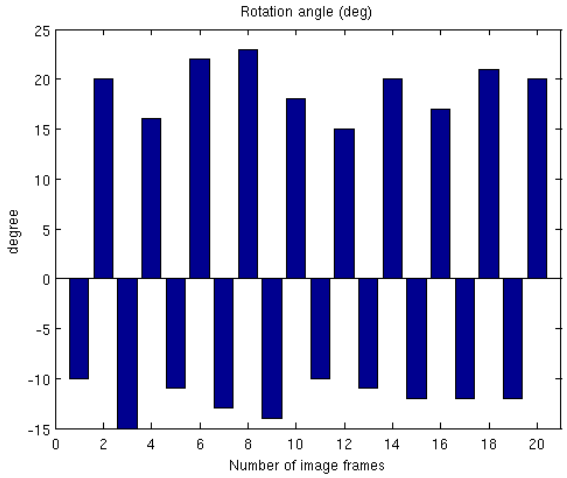

Fig. 8. A sequence of drastically changing yaw values of the robot (top). At each orientation, an image I_n is taken and the underlaying rotation parameter (bottom) is determined with FM-SLAM on I_{n-1}. As shown in figure 9, the errors are even in this extreme case very small.

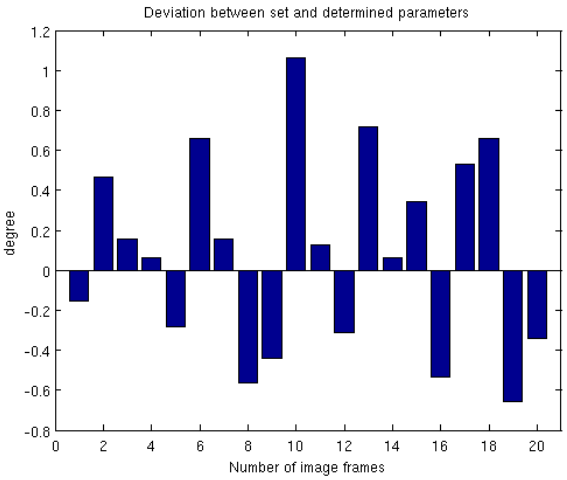

Fig. 9. The errors in absolute orientation when using FM-SLAM in the sequence of drastically changing yaws shown in figure 8. The errors are even in this extreme case on average well below one degree.

in this application scenario and it can be used for mosaicking. A more detailed description can be found in [18].

But visual odometry has its limits when its comes to the generation of maps. The consecutive registration of video frames involves cumulative errors. Especially, errors in orientation build up and lead to significant distortions over time. It is hence of interest to use proper Simultaneous Localization and Mapping (SLAM). An embedding of the presented approach into pose graph SLAM [19], [20] is current work in progress.

V. Conclusion

An improved Fourier-Mellin Invariant (iFMI) descriptor is presented, which extends previous FMI based approaches in two ways. 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 iFMI can be used in a fast and robust manner for visual odometry, as demonstrated with

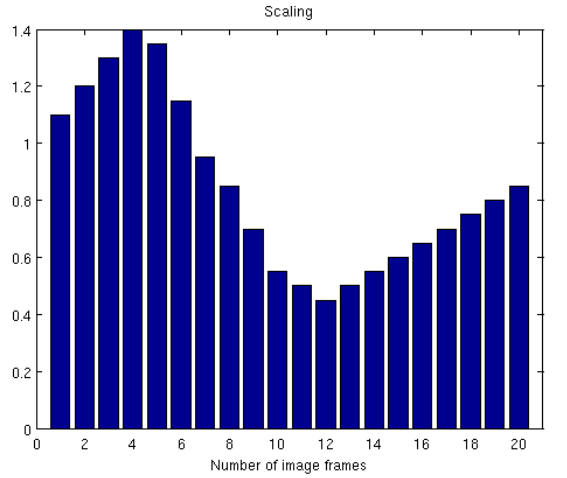

Fig. 10. The ground truth changes in altitude measured by absolute scale changes in subsequent images.

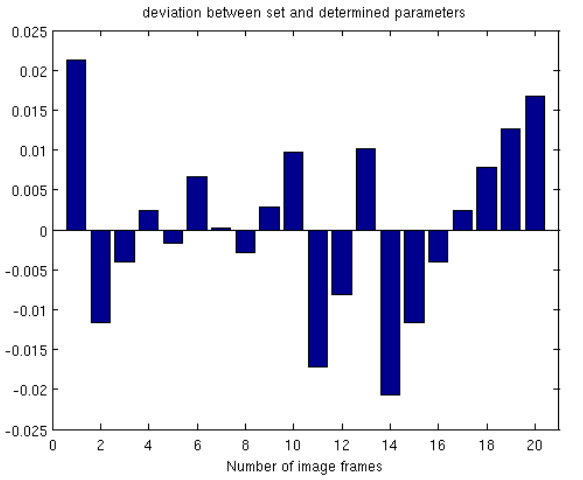

Fig. 11. The errors in the altitude expressed as difference in ground truth versus computed scale factor. Note that the maximum error is about 2%, the average error is below 1%.

experiments on photomapping with an Unmanned Aerial Vehicle (UAV). The related work is also interesting for other application domains, especially underwater robotics.

Acknowledgment

The parts of the research leading to the results presented here has received funding from the European Community's Seventh Framework Programme (EU FP7) under grant agreement n. 231378 "Cooperative Cognitive Control for Autonomous Underwater Vehicles (Co3-AUVs)", `http://www.Co3-AUVs.eu`.

References

[1] A. Fitch, A. Kadyrov, W. Christmas, and J. Kittler, "Fast robust correlation," *IEEE Transactions on Image Processing*, vol. 14, pp. 1063–1073, 2005.

[2] D. Stricker, "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*. ACM Press, 2001, pp. 77–82.

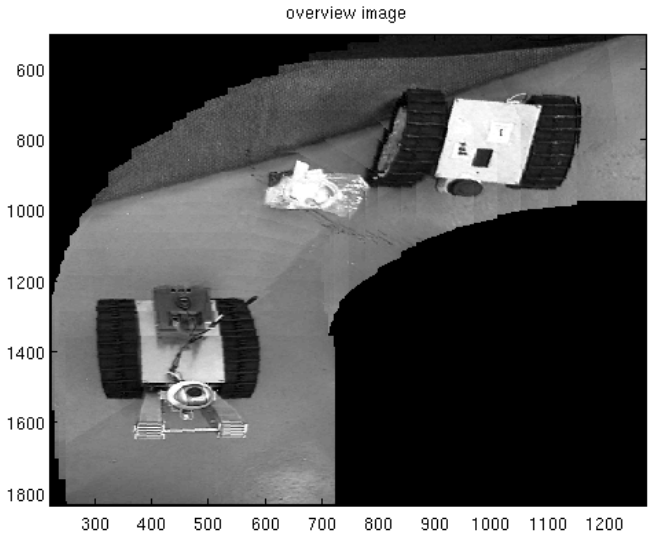

Fig. 12. An image map generated from data of a down-looking camera on a robot passing over two of the Jacobs land robots. The sequence mainly involves translations and rotations as well as some minor amounts of scaling.

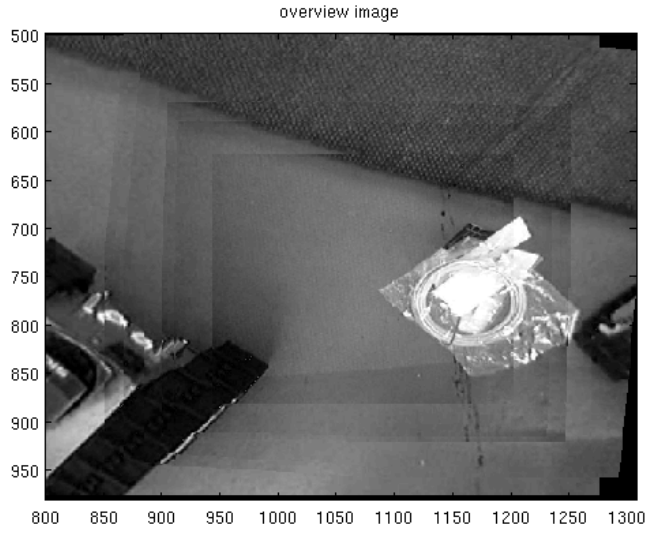

Fig. 13. An image map generated from data of a down-looking camera. The sequence major amounts of scaling and some rotations.

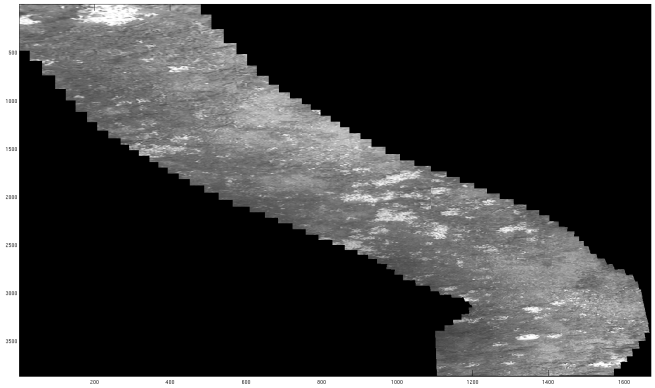

Fig. 14. Overview of the registered underwater scene.

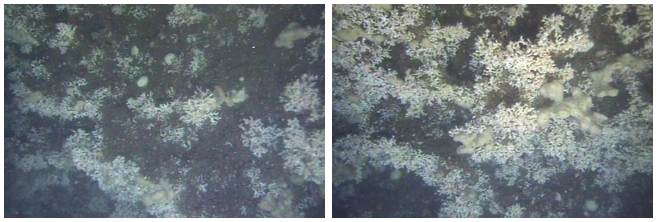

Fig. 15. Example frame from an underwater video sequence.

points," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[3] C. Dorai, G. Wang, A. K. Jain, and C. Mercer, "Registration and integration of multiple object views for 3d model construction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 83–89, Jan., 1998.

[4] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, 1992.

[5] S. Alliney and C. Morandi, "Digital image registration using projections," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 2, pp. 222–233, 1986.

[6] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings DARPA Image Understanding Workshop*, 1981, pp. 121–130.

[7] W. Pratt, "Correlation techniques of image registration," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-10, pp. 562–575, 1973.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] ——, "Object recognition from local scale-invariant features," in *Proceedings of International Conference on Computer Vision*, 1999, pp. 1150–1157.

[10] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *Proceedings of Computer Vision and Pattern Recognition*, June, 2003.

[11] L. V. Gool, T. Moons, and D. Ungureanu, "Affine/photometric invariants for planar intensity patterns," in *Proceedings of European Conference on Computer Vision*, 1996.

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

[13] 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.

[14] 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.

[15] O. Pizarro, H. Singh, and S. Lerner, "Towards image-based characterization of acoustic navigation," in *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, vol. 3, 2001, pp. 1519–1524 vol.3.

[16] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Prentice Hall Signal Processing Series, Englewood Cliffs, 1989.

[17] A. Birk and S. Carpin, "Merging occupancy grid maps from multiple robots," *IEEE Proceedings, special issue on Multi-Robot Systems*, vol. 94, no. 7, pp. 1384–1397, 2006.

[18] H. Buelow, A. Birk, and V. Unnithan, "Online generation of an underwater photo map with improved fourier mellin based registration," in *International OCEANS Conference*. IEEE Press, 2009.

[19] M. Pfingsthorn, B. Slamet, and A. Visser, "A scalable hybrid multi-robot slam method for highly detailed maps," in *RoboCup 2007: Proceedings of the International Symposium*, ser. LNAI. Springer, 2007.

[20] E. Olson, J. Leonard, and S. Teller, "Fast iterative alignment of pose graphs with poor initial estimates," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, J. Leonard, Ed., 2006, pp. 2262–2269.