On the effects of Sampling Resolution in Improved Fourier Mellin based Registration for Underwater Mapping

Sören Schwertfeger * **Heiko Bülow** * **Andreas Birk** *

* *Jacobs University Bremen, 28759 Bremen, Germany (e-mail: a.birk@jacobs-university.de).*

Abstract: The properties of the fast and robust method for visual odometry based on the Fourier-Mellin Invariant (FMI) descriptor are analyzed here. The algorithm is particularly suited for Autonomous Underwater Vehicles (AUV), namely for the online generation of photo maps that can be the basis of intelligent on-board functionalities. The photo maps can be generated just based on registration, i.e., without any information about the vehicle's pose or its motion. The algorithm makes heavy use of 2D Fourier transformations and its speed thus depends on the resolution of the input images and the size of the intermediate matrices used. Using the ground truth path of generated artificial video streams, the effects of different resolutions and other parameters are compared and evaluated with respect to their speed.

Keywords: Image registration, underwater autonomous vehicles

1. INTRODUCTION

The online generation of maps is one of the core foundations for intelligent cognitive functionalities on-board of Autonomous Underwater Vehicles (AUV)[Birk et al. (2009b)].

Given perfect localization of an AUV and the parameters of a down-looking camera that takes pictures of the sea floor, the generation of a photo map is trivial. But AUV localization is error prone and costly, we hence concentrate here on the case of not having any information about the vehicles absolute pose or motions. The information of how to merge two consecutively acquired images is instead extracted purely out of the images themselves. So, 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 [Fitch et al. (2005); Stricker (2001); Dorai et al. (Jan., 1998); Brown (1992); Alliney and Morandi (1986); Lucas and Kanade (1981); Pratt (1973)]. But the underlying problem for photo mapping is more difficult as the region of overlap is unknown and it has undergone non-trivial transformations due to the AUV's motion between two consecutively acquired images. This is more comparable to image stitching [Lowe (2004)].

The algorithm and its implementation (figure 1) analyzed here are applied not only to the underwater domain [Bulow et al. (2009)] but also to unmanned aerial vehicles [Bulow and Birk (2009)], especially in the domain of safety, security and rescue robotics [Birk et al. (2009a)]. The approach is based on the postulation that the whole information in the images and not only features should be used to minimize uncertainties and ambiguities in registration. Our approach is hence based on a variant of the Fourier Mellin transform for image representation and processing [Chen et al. (1994); Reddy and Chatterji (1996)], which was used before for underwater photo mapping [Pizarro et al. (2001)].

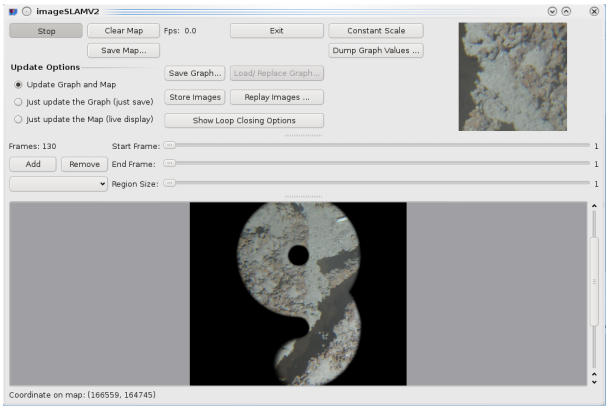

Fig. 1. The GUI of the mapping program used for the experiments of this paper in action.

The rest of this paper is structured as follows. In section 2, the improved Fourier Mellin Invariant (iFMI) descriptor is introduced. Section 3 describes how the artificial ground truth data is generated while section 4 introduces the different experiments conducted. In section 5 the results of the experiments are presented and analyzed while the paper is concluded in section 6.

2. IMPROVED FOURIER MELLIN MAPPING

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

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

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 (eq. 2).

2.1 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.

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

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 be registered.

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

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

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

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

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. This is can be done in a different resolution (the *polar-log-resolution*) than the original one (*image-resolution*). For turning a signal scaling into a signal shift several steps are necessary. The following Fourier theorem

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

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:

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

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

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

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

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

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. This intermediate step is called the FMI descriptor.

The overall algorithm is sketched here. First the calculation of the POMF is shown:

(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 (in *image-resolution*)
(2) calculate the magnitude of the complex spectral data
(3) re-sample the spectra to polar coordinates (in *polar-log-resolution*)
(4) re-sample the radial axes of the polar spectra logarithmically
(5) calculate a POMF on the re-sampled 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 and re-scaled 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.

3. GENERATING GROUND TRUTH DATA

The ground truth data, a simulated dive of an autonomous underwater vehicle (AUV) pointing a camera to the bottom, is generated by taking a high resolution image of the ocean floor and cutting out lower resolution images representing the frames of a video stream. The frames are cut out using the following parameters:

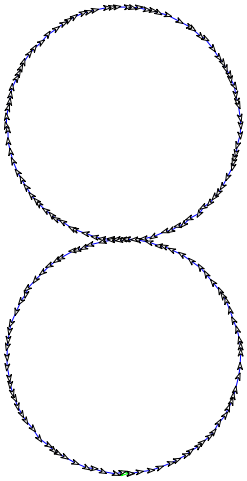

Fig. 2. The figure eight shape. A path with 175 nodes (frames). It is actually path number 1 in table 1.

- Position: the x and y position of the center of the image
- Orientation: the orientation of the frame
- Height: the height corresponds to a scaling of the cut-out image. The minimum height is 1. which corresponds to no scaling, thus avoiding pixel artefacts in the simulated video data.

A program is used to generate paths of different shape and with different parameters like size of the shape, step width (between the frames), orientation, random variance of pose and orientation and possibly changes in height. Figure 2 shows a typical path for the "figure eight" used in all experiments in this paper. Figure 3 shows one of the four images used for cutting out the frames while figure 4 shows a map generated out of those frames (using a 512x512 pixel resolution).

We are well aware that the data generated lacks any errors occurring in reality like noise, dirt, moving objects, perspective errors or pitch and roll. Nevertheless the experiments performed deliver valid data because they always judge the effects investigated by comparing results gained with equally perfect input data.

4. EXPERIMENTS

In the experiments four different images are used with four different paths, leading to 16 different simulated video streams. Two images have few features and relatively uniform terrain while the other two are rich of features. Two of the paths feature no change in height while the other two oscillate twice respectively three times between height 1. and 1.5. The properties of the paths used can be found in table 1.

The smoothness describes how the perfect circles of the figure eight are deteriorated. Smooth paths are not deteriorated - only path number 4 includes additional changes in height. In paths 1 and 3, random changes are added to the poses (including orientation). As a result the maximum step width of those paths are higher than the average step width.

The images are cut out with a resolution of 512x512 pixel. The other resolutions tested in this paper are 448, 384, 320, 256, 192, 128, 92 and 64. In these experiments the *polar-log-* and *image-resolution* are always equal. In order to use a smaller resolution two possible approaches are investigated:

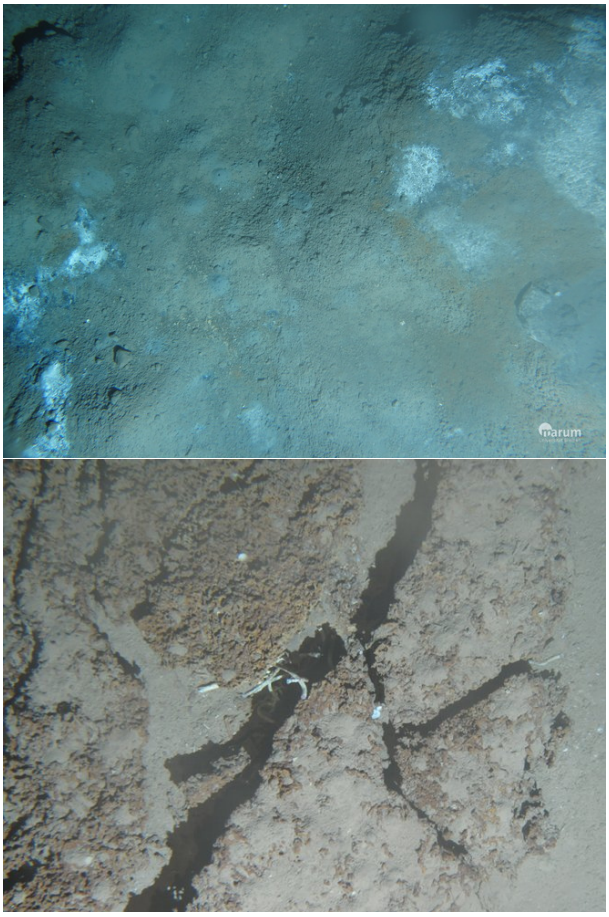

Fig. 3. A uniform (top) and a featue rich (bottom) image used for extracting the simulated video stream (both MARUM).

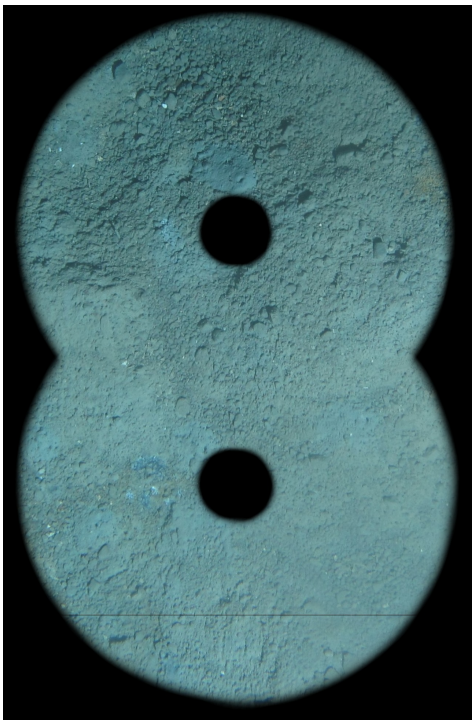

Fig. 4. A map generated out of the simulated video stream using a resolution of 512 and the figure eight path shown in figure 2.

Table 1. Properties of the paths used.

	# frames	Average step width	Maximum step width	Height	Height Frequency	Smoothness
1	175	25	41	1	-	random
2	211	30	30	1	-	smooth
3	129	30	52	1 - 1.5	2	random
4	211	30	30	1 - 1.5	3	smooth

Table 2. The speed on an Intel Core 2 2.8 GHz CPU. The registration itself did not make use of multithreading.

Image & polar-log resolution	Registration		GUI	
	gsl fps	fftw fps	gsl fps	fftw fps
64	640.8	1168.4	111.6	128.9
96	281.0	412.8	67.7	93.3
128	159.1	272.5	46.4	53.6
192	67.7	117.6	23.0	28.0
256	34.8	66.3	13.2	16.9
320	21.8	40.6	10.0	14.4
384	14.9	27.4	5.8	7.7
448	10.2	19.3	4.9	7.1
512	7.00	14.3	2.9	4.3

- Resize: The original 512 pixel image is resized. Thus the overlap stays the same but the details vanish.
- Cut-out: The image registration only takes the center of the original frame. The overlap decreases, possibly too much, but a high accuracy is possible if enough overlap is present.

All 16 video streams are run with both approaches and all nine resolutions, leading to a total number of 272 mapping runs - for the 512x512 resolution the resize approach and the cut-out approach deliver the same result (since the resolution is not changed). The resulting camera path generated during the mapping is then compared to the ground truth path. The errors in the position and the orientation are determined together with the standard deviations of those errors.

4.1 Runtimes

The runtime of the FMI algorithm does not depend on the content of the image and is thus constant for a given pair of resolutions (image and polar-log). The runtime for the pure image registration depends to a great extend on the specific FFT library used. In table 2, first the speed of the pure registration is given, which scales with the resolution. The last values show the speed of the GUI which incorporates various tasks like loading, converting, and undistorting of the images, the actual registration, as well as merging the frames according to the transformation parameters together in one map and displaying it. A screenshot of the GUI is given in figure 1.

5. RESULTS

In the discussion about the results of the experiments two different error measurements are of relevance. The first is the error of the orientation as absolute value. As stated in section 2, the orientation (and also the scale, or height) is determined using polar-logarithmic re-sampling in the *polar-log-resolution*. The absolute difference (in degrees) between the calculated orientation of two successive frames and the actual change in orientation as given in the ground truth path

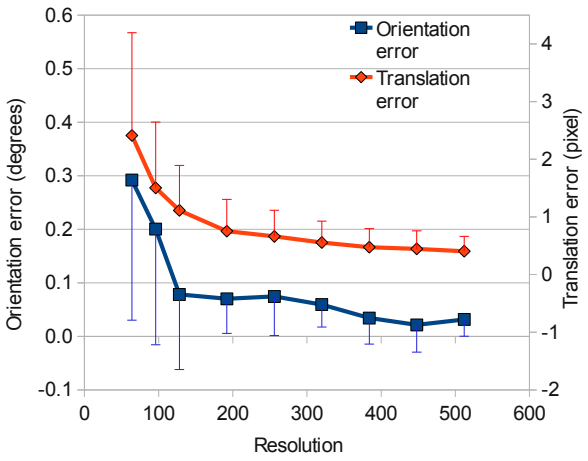

Fig. 5. Average over the errors of all video streams of a given resolution. The standard deviation of the errors is also shown. It is symmetrical but one side is omitted here to ease readability. Please also note the scale and range of both y-axis in this and further diagrams.

is averaged for every video stream of a given resolution and represents the error of this step.

The second error is the error in translation. After finding the rotation and scaling one of the input frames is transformed according to those values such that afterwards there is only a translational difference between the transformed and the other frame. Thus this step is vulnerable to errors occurring already during the determination of the scale and orientation. In this paper the translation error was not assessed independend of the *polar-log-resolution*. Again, the geometirc distance (an absolute value with the unit 'pixel') between the calculated translation and the actual translation defined in the path is averaged for every video stream of a given resolution. This is the translation error.

Looking at figure 5, we see the two results for the average error at a given resolution averaged over all 16 video streams. The values are the average inaccuracy between two successive frames. Those errors accumulate over time. One could say that in general down to a value of 192 or even 128 for both resolutions (*polar-log-* and *image-resolution*) the algorithm performs quite well.

For the following figures the standard deviations are omnitted either because those are hard to interpret on a logarithmic scale (figures 6 and 7) or because the graphs would be hardly readalbe anymore (firugres 8, 9 and 10). Those deviations are anyways very similar to those in figure 5 and scale with the value of the error.

5.1 Resize vs cut-out

In section 4 two approaches to lower the resolution of the input video stream are presented: resize and cut-out. Figures 6 and 7 compare their performance for different resolutions. The graphs can be explained as follows: For the orientation both approaches have nearly the same errors down to 192x192 pixel. The resized frames perform that well since even small rotations lead to big changes in the outer parts of the image such that the higher details of cut-out lead to no advantage. In lower resolutions cut-out is performing much worse than resize. This

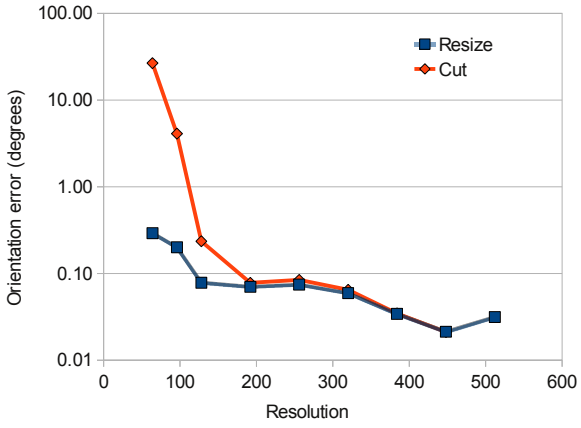

Fig. 6. Comparison in the accuracy of the orientation between the resize and cut-out approach.

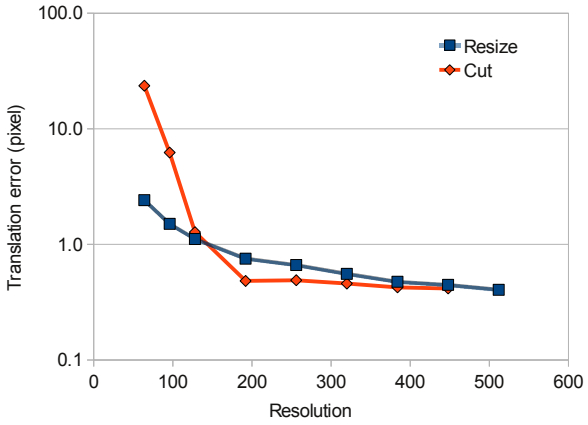

Fig. 7. Comparison in the accuracy of the translation between the resize and cut-out approach.

is because here the overlap is decreasing too much for cut-out while it does not change at all for resize.

If we consider 192x192 as still acceptably performing for cut-out (which can also be seen in the graph for the translation in figure 7) we can calculate the needed overlap, knowing that the average step width is 30 (see table 1): $1 - 30/192 \approx 85\%$. With less than this overlap it is likely that other peaks in the Dirac-domain have higher energy than the "correct" one and the determined result is not just inaccurate but completely wrong. If, in real world applications, further error sources appear (see section 1) more overlap is strongly recommended.

The interesting graph is in figure 7 for the translation. As long as there is enough overlap it is preferable to cut out. This is due to the effect that nearly no information is lost in cut-out in resolutions down to 192 since the overlap is always sufficient. Resize, on the other hand, constantly looses accuracy due to the scaling and the interpolation. Even at 128 both approaches perform nearly equally.

5.2 Further results

The experiments for this paper were deliberately planned such that the other properties of the iFMI algorithm can be shown. Figure 8 is comparing the performance of the image registration between the video streams made from the feature rich and more uniform pictures. It can clearly be seen that both error developments are very similar, leading to the conclusion that the content or the feature richness of the scene does not affect the registration accuracy (as long as there is any non-ambiguous structure).

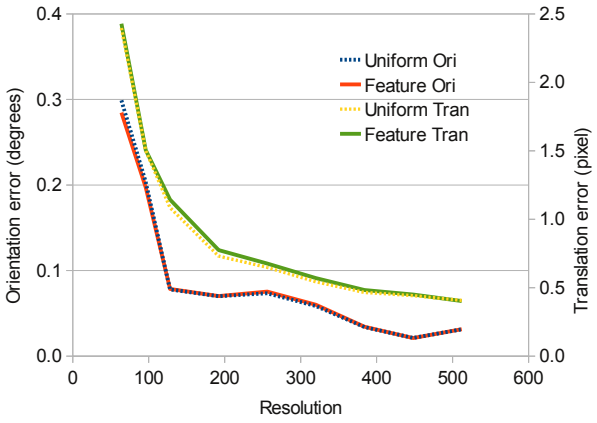

Fig. 8. Comparison of the results of feature rich and uniform images.

Figure 9 compares the algorithm's performance between the flat paths and those which oscillate in height. It can be seen the the determination of the orientation (and thus also the scale as it is done in the same step) is not effected by the height changes. The accuracy of the translation suffers from the scaling and interpolation needed in the oscillating paths.

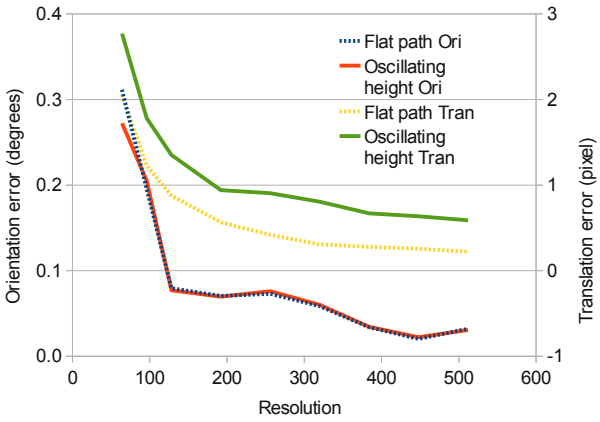

Fig. 9. Comparison of the results of the flat and height oscillating paths.

The most surprising result was that shown in figure 10. Here the effects of the smoothness property of the paths as described in section 4 and table 1 are investigated. Since the algorithm can handle any combination of rotations and translations (including height changes) it was expected to see no differences. Alas, the orientation of the smooth path is showing a strange pattern. It was also checked in the raw data and found that this is not due to one outstanding run but consistent with all smooth runs. It is believed that this is an effect of the circle radius and step width (leading to a certain, constant orientation change in each step) interfering the the chosen resolution. This could lead to rotations being directly on one pixel (matrix element) in the Dirac-domain in some cases (leading to better results than

normal) and in other combinations laying more or less between two (or four) pixel, thus performing worse than average.

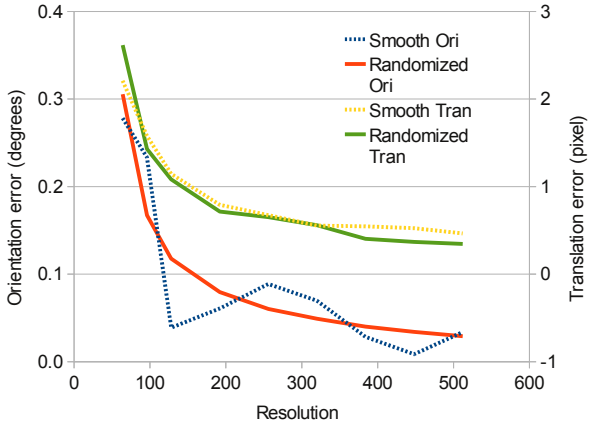

Fig. 10. Comparison of the results of the smooth and the randomized paths.

6. CONCLUSION AND FUTURE WORK

The improved Fourier Mellin Invariant (iFMI) image registration was shortly introduced. The properties of this algorithm, especially with respect to the resolutions used, were analyzed. It was found that resolutions of 192 and up generally perform sufficiently well. It was also concluded that the minimum overlap at this resolution should be at least 85%. With these parameters in mind, the presented program can run at more than 25Hz on a modern laptop, including all tasks like data acquisition, image undistortion, live display, and storage. It should be noted that in this setup only the image registration uses 192 for both resolutions - everything else can be done with this speed in the original size (512x512). Furthermore, it was shown that the algorithm can work with sparse and uniform input data and can also handle changes in height decently well.

Further analysis on this algorithm should involve different step width and less perfect input data with noise and perspective errors. Work on adaptively controlling the resolution, or even changing from cut-out to resize, depending on the speed of the vehicle is another interesting topic.

Generating photo maps just based on registration has the obvious disadvantage that errors accumulate and there hence is a drift. While this is tolerable for short time periods and basic intelligent functions, it is desirable to suppress cumulative errors. The presented approach is therefore currently embedded into Posegraph SLAM. Using error information from the registration when previously seen spots are revisited, relaxation of the spatial transformation estimates in the graph can be used to bound the overall error, i.e., to get proper maps with significantly reduced error. Please note that this still purely relies on registration of consecutively acquired images, i.e., no sensors of pose or motion estimation are needed.

ACKNOWLEDGMENTS

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`. Furthermore we would like to thank the Marum - Center for Marine Environmental Sciences, University of Bremen for providing the underwater images.

REFERENCES

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

Birk, A., Wiggerich, B., Unnithan, V., Bülow, H., Pfingsthorn, M., and Schwertfeger, S. (2009a). Reconnaissance and camp security missions with an unmanned aerial vehicle (uav) at the 2009 european land robots trials (elrob). In *IEEE International Workshop on Safety, Security and Rescue Robotics, SSRR*.

Birk, A., Antonelli, G., Pascoal, A., and Caffaz, A. (2009b). Cooperative Cognitive Control for Autonomous Underwater Vehicles. In *8th International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT)*. Budapest.

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

Bulow, H. and Birk, A. (2009). Fast and robust photomapping with an unmanned aerial vehicle (uav). In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 3368–3373. doi:10.1109/IROS.2009.5354505.

Bulow, H., Birk, A., and Unnithan, V. (2009). Online generation of an underwater photo map with improved fourier mellin based registration. In *OCEANS 2009-EUROPE, 2009. OCEANS '09.*, 1–6. doi:10.1109/OCEANSE.2009.5278193.

Chen, Q.S., Defrise, M., and Deconinck, F. (1994). Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(12), 1156–1168.

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

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

Lowe, D.G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110.

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

Pizarro, O., Singh, H., and Lerner, S. (2001). Towards image-based characterization of acoustic navigation. In *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, volume 3, 1519–1524 vol.3.

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

Reddy, B. and Chatterji, B. (1996). An FFT-based technique for translation, rotation, and scale-invariant image registration. *Image Processing, IEEE Transactions on*, 5(8), 1266–1271.

Stricker, D. (2001). 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*, 77–82. ACM Press.

Schwertfeger, S., H. Bülow, and A. Birk, "On the effects of Sampling Resolution in Improved Fourier Mellin based Registration for Underwater Mapping", 7th Symposium on Intelligent Autonomous Vehicles (IAV), IFAC: IFAC, 2010.