Robot Pose Estimation with Omni-iFMI

Kuang Haofei

***Abstract*—To estimate robot pose from image transformation according to camera model is a classical problem in robotics and computer vision domain. Currently, a feature-based visual odometry(VO) algorithm is widely used to solve the problem. These method using a feature-based algorithm to finding a set of correspondences as first step of VO. It could works nicely in most scenarios. But this method will lose its effect in some low feature area, like air or underwater scenario. Extracting and matching features also will spend a lot of time. Here, a spectral method such as the improved Fourier-Mellin Invariant(iFMI) descriptor have proved faster, more robust and accurate than feature-based method. In this project, we will implement a iFMI descriptor to replace feature-based method of VO algorithm to finding a correspondences. Considering this method has not been applied on omnidirectional camera(omni-camera), we will implement the method with it. According to VO algorithm is not much associated with the course, we will concentrate on the theory and show the experiments about image registration with iFMI descriptor and apply it to generate the correspondences of two omni-camera images.**

I. Introduction

THE localization in online is one of the core capability of autonomous mobile robot. The task of robotics localization is estimate the pose of robot in coordinate system. In previous years, a probabilistic-based method is used to achieve the problem [1]. The weakness of these method is that it need expensive ranger sensor and it not suit to some area without rich sets of natural landmark, like underwater environment. In recent year, a visual-based method is proposed to solving the problem. One of novel idea is using image registration method to do it.

For calculating the pose of robot, the transformation of robot is need to compute. Robot transformation could be calculated by using visual odometry(VO) algorithm according to sets of corresponding points of two frame images, it is called correspondences. The mainstream method of finding correspondences is using some feature-based method, e.g. The scale invariant feature transform (SIFT) is a very popular basis for image registration [2]. But these method need extract feature points and matching these points. This step always spend a lots of time and couldn't works in low-feature area, like underwater and air.

In 1994, Chen et al. used Fourier-Mellin Invariant (FMI) descriptor and symmetric phase-only matched filtering to calculate rotation, scaling and translation of two images [3]. Kazik et al. [4] using this method to compute correspondence with a ground-facing camera to calculate a visual odometry. And Thoduka et al. [5] using this method to do motion detection. And it is proved more fast and robust than feature-based method in this domain. On the basis of FMI, Heiko et al. [6] extend the approaches .

The previous works about the method is using FMI to estimate robot pose with general camera model. But the omnidirectional camera has more wild receptive field and iFMI has better performance than FMI. In this project, we want to implement a image registration algorithm with the image which is acquired from a omni-camera to get a accuracy correspondences for robot pose estimation.

This project report is organized in the following way. In the next three sections, we will describe the spectral-based algorithm of image registration. In Section II, we will discuss the classical matched filtering and the phase-only and symmetric phase-only matched filtering firstly. Then the Fourier-Mellin invariant descriptor(FMI) is defined in Section II-B. The improved Fourier-Mellin invariant(iFMI) will be discussed in later. In Section III, we will talk about the implement detail of the image registration algorithm. The we will show our experiment results in Section IV. Finally, our conclusion is given in Section V.

Kuang Haofei was with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China e-mail: kuanghf@shanghaitech.edu.cn. Student ID: 70553890.

II. Methodology

The aim of image registration is to determine the parameters $P = [p_1, p_2, \cdots, p_n]^T$ of a geometric transformation relating two images s and r:

$$s(x, y) = r(x', y') \tag{1}$$

$$\begin{cases} x'(x, y, P) \\ y'(x, y, P) \end{cases} \quad P = \begin{bmatrix} x_0 \\ y_0 \\ \alpha \\ \sigma \end{bmatrix} \tag{2}$$

Considering the application of image registration is robot pose estimation, so the geometric transformation parameter is consisted with four values, including translation(x_0, y_0), rotation(α) and scaling(σ). In this section, we will discuss method which is used to solve the image registration problem in detail. The method is consist of symmetric phase-only matched filtering(POMF) of Fourier-Mellin transforms(FMI) [3] and the improved version of it [6].

A. Phase-Only Matched Filter

In this subsection, we consider a simple two-dimensional translation problem with translation offset (x_0, y_0), thus, $x = x - x_0$ and $y = y - y_0$. The Fourier transformation of the two images are defined by

$$\begin{cases} R(u, v) = \mathcal{F}\{r(x, y)\} \\ S(u, v) = \mathcal{F}\{s(x, y)\} \end{cases} \tag{3}$$

The classical matched filter [7] [8], which maximizes the detection signal-to-noise ratio. The out put of the filter is the convolution of $r(-x,-y)$ and $s(x,y)$,

$$q_0 = \int\int_{-\infty}^{\infty} s(a,b)r(a-x,b-y)\,dadb \tag{4}$$

This function has a maximum at (x_0, y_0) that determines the parameters of the translation.

One limitation of the matched filter defined by Equation (4) is that the output of the filter is primarily dependent on the energy of the image rather than 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).

The transfer function of a phase-only matched filter [9] is equal to the spectral phase of the image:

$$H(u,v) = \text{Phase}(R^*(u,v)) = e^{j(-\phi_r(u,v))} \tag{5}$$

$R^*(u,v)$ is the complex conjugate of the Fourier spectrum $R(u,v)$. The detection and location of the maximum is then easier, and the POMF allows a better discrimination between different objects than the matched filtering [10] [11].

A further improvement of the phase-only matched filtering can be achieved by extracting and correlating the phases of both $r(x,y)$ and $s(x,y)$. This is done by calculating the inverse Fourier transform of the cross power spectral of r and s:

$$Q(u,v) = \frac{S(u,v)\circ R^*(u,v)}{|S(u,v)\circ R^*(u,v)|} = e^{j(\phi_s(u,v)-\phi_r(u,v))} \tag{6}$$

where $\phi_s(u,v)$ and $\phi_r(u,v)$ are the spectral phases of $s(x,y)$ and $r(x,y)$, and $\circ$ operator is element-wise product. In the absence of noise, this function reduces to

$$Q(u,v) = e^{-j2\pi(ux_0+vy_0)} \tag{7}$$

So, the $Q(u,v)$ is the shift information within its phase [12]:

$$f(x-x_0, y-y_0) \Leftrightarrow F(u,v)e^{-j2\pi(ux_0+vy_0)} \tag{8}$$

The inverse Fourier transform of Equation (7) is a Dirac function(δ) centered at (x_0, y_0), yielding an even sharper maximum than the phase-only matched filter,

$$\mathcal{F}^{-1}\{Q(u,v)\} = \delta(x-x_0, y-y_0) \tag{9}$$

$$(x_0, y_0) = \arg\max_{x,y} \delta(x-x_0, y-y_0) \tag{10}$$

This technique, referred to as a symmetric phase-only matched filtering (SPOMF), can be seen as a nonlinear two step process, the first step being the extraction of the phases of the input images and the second the phase-only matched filtering [13]. This method also called phase correlation technology [14]

B. The Fourier-Mellin Invariant Descriptor(FMI)

In last subsection, we have discussed how to solve the translation estimation problem through SPOMF. In this subsection, we will talk about the method with estimating rotation and scaling parameters, which is called Fourier-Mellin invariant(FMI) descriptor [3].

Consider an image $s(x,y)$ that is a rotated, scaled, and translated replica of $r(x,y)$,

$$\begin{aligned} s(x,y) = r[&\sigma(x\cos\alpha - y\sin\alpha) - x_0, \\ &\sigma(x\sin\alpha + y\cos\alpha) - y_0] \end{aligned} \tag{11}$$

The Fourier transforms of $s(x,y)$ and $r(x,y)$ are related by,

$$\begin{aligned} S(u,v) =\ & e^{-j\phi_s(u,v)}\sigma^{-2}R[\sigma^{-1}(u\cos\alpha - v\sin\alpha), \\ & \sigma^{-1}(u\sin\alpha + v\cos\alpha)] \end{aligned} \tag{12}$$

where $\phi_s(u,v)$ is the spectral phase of the image $s(x,y)$, This phase depends on the translation, scaling and rotation, but the spectral magnitude of it is like

$$\begin{aligned} |S(u,v)| =\ & \sigma^{-2}|R[\sigma^{-1}(u\cos\alpha - v\sin\alpha), \\ & \sigma^{-1}(u\sin\alpha + v\cos\alpha)]| \end{aligned} \tag{13}$$

The Equation (13) is translation invariant. It shows that a rotation of the image rotates the spectral magnitude by the same angle, and that a scaling by σ scales the spectral magnitude by σ^{-1}. Here, we already remove the influence of translation, and it could be use SPOMF to solve the rotation and scaling parameters. But Equation (13) is not clear for SPOMF, we could be helped with polar coordinates to get a clear formation. Rotation and scaling can be decoupled by defining the spectral magnitudes of r and s in the polar coordinates (θ, ρ),

$$\begin{cases} r_p(\theta,\rho) = |R(\rho\cos\theta, \rho\sin\theta)| \\ s_p(\theta,\rho) = |S(\rho\cos\theta, \rho\sin\theta)| \end{cases} \tag{14}$$

Using,

$$\begin{cases} \sigma^{-1}(\rho\cos\theta - \rho\sin\theta) = \frac{\rho}{\sigma}(\cos(\theta-\alpha)) \\ \sigma^{-1}(\rho\cos\theta + \rho\sin\theta) = \frac{\rho}{\sigma}(\sin(\theta-\alpha)) \end{cases} \tag{15}$$

we could get the below equation easily,

$$s_p(\theta,\rho) = \sigma^{-2}r_p(\theta-\alpha, \frac{\rho}{\sigma}) \tag{16}$$

Hence an image rotation shifts the function $s_p(\theta,\rho)$ along the angular axis. The problem is focus on how to explain the scaling as a translation currently. Scaling can be further reduced to a translation by using a logarithmic scale for the radial coordinate. So, we define,

$$r_{pl}(\theta,\lambda) = r_p(\theta, \log\rho) \tag{17}$$

$$s_{pl}(\theta,\lambda) = s_p(\theta,\log\rho) = \sigma^{-2}r_{pl}(\theta-\alpha, \lambda-\kappa) \tag{18}$$

where, $\lambda = \log(\rho)$ and $\kappa = \log(\sigma)$. Now, both rotation and scaling are reduced to translations in this polar-logarithmic representation. By Fourier transforming the polar-logarithmic representations of Equation (17) and (18), we could obtain,

$$S_{pl}(v,\varpi) = \sigma^{-2}e^{-j2\pi(v\kappa+\varpi\alpha)}R_{pl}(v,\varpi) \tag{19}$$

where rotation and scaling are explained as phase shifts now. This technique decouples image rotation, scaling, and translation, and is therefore very efficient numerically. The polar-log mapping of the spectral magnitude corresponds to a physical realizations of the Fourier-Mellin transform [15] [6]. And the $r_{pl}(\theta, \lambda)$ and $s_{pl}(\theta, \lambda)$ are called the Fourier-Mellin invariant(FMI) descriptor of $r(x, y)$ and $s(x, y)$ respectively. According to the result of FMI descriptor, we could using SPOMF to solve the parameter α and κ, then get the ration and scaling parameters.

C. The improved-FMI

Using FMI-SPOMF, we could determine the parameters $P = [x_0, y_0, \alpha, \sigma]^T$ of a geometric transformation relating two images s and$r$. It is a efficient, fast and robust method for image registration problem. For improve the performance of FMI-SPOMF algorithm and makes it suits to some robotics application, Bülow et al. [6] [16] propose two significant modifications makes the algorithm achieve a very fast and robust performance.

- First, a logarithmic representation of the spectral magnitude of the FMI descriptor is used instead of logarithmic representation of the FMI descriptor.
- Second, a filter on the frequency where the shift is supposed to appear is applied.

This method is called improved Fourier-Mellin Invariant(iFMI) descriptor by Bülow et al. And they shows the method could be works with generating photo map in real time with underwater robotics [6] and UAV [16] scenarios.

III. Implementation

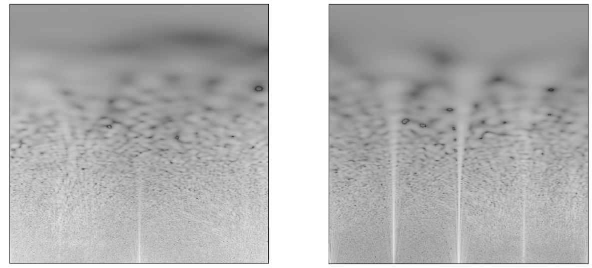
We talk about the theory of FMI-SPOMF algorithm and iFMI descriptor in last section. In this section, we will discuss implement details of FMI-SPOMF algorithm and iFMI algorithm. We first describe a "core" FMI-SPOMF algorithm, and then describe how to modify it using the theory of iFMI.

According to the Equation (19), given a reference image $r(x, y)$ and an observed image $s(x, y)$, the output of the FMI-SPOMF is of the form:

$$Q_o(v, \varpi) = \frac{R^*_{pl}(v, \varpi) \circ S^*_{pl}(v, \varpi)}{|R_{pl}(v, \varpi)| \circ |S_{pl}(v, \varpi)|} \quad (20)$$

The core FMI-SPOMF algorithm consists of the following steps

1) Input two image $r(x, y)$ and $s(x, y)$,
2) compute the Fourier transform $R_{pl}(v, \varpi)$ of the FMI descriptor of the reference image $r(x, y)$,
3) extract the phase $\exp[-j\phi_r(v, \varpi)]$ of $R_{pl}(v, \varpi)$,
4) compute the Fourier transform $S_{pl}(v, \varpi)$ of the FMI descriptor of the reference image $s(x, y)$,
5) extract the phase $\exp[-j\phi_s(v, \varpi)]$ of $S_{pl}(v, \varpi)$,
6) compute the output of the SPOMF, the polar-logarithm
$$Q_o(v, \varpi) = e^{j(\phi_s(v,\varpi) - \phi_r(v,\varpi))}$$
7) compute the inverse Fourier transform, the result is like Fig. 2,
$$q_0(\theta, \lambda) = \mathcal{F}^{-1}\{Q_o(u, v)\}$$
8) detect the maximum of the function $q_o(\theta, \lambda)$.
$$(\alpha, \kappa) = \arg\max_{\theta,\lambda} q_0(\theta, \lambda)$$



(a) (b)

Fig. 1: Polar-logarithmic representation of $r(x, y)$ and $s(x, y)$. (a) and (b) are $R_{pl}(v, \varpi)$ and $S_{pl}(v, \varpi)$ respectively. It is clearly to see, the rotation and scaling change to translation in Polar-logarithmic representation.

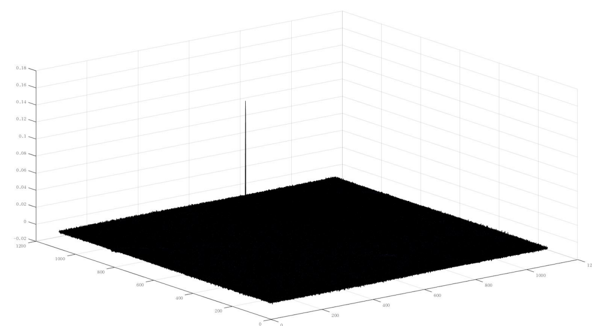


Fig. 2: The inverse Fourier transform of $Q_o(v, \varpi)$ is idea Dirac function, which has a maximum in position (α, κ).

The iFMI algorithm has two modify of the algorithm flow. First, outperform a high pass filter after step 1). Second, compute the spectral magnitude after 2) and 4) of $R_{pl}(v, \varpi)$ and $S_{pl}(v, \varpi)$ instead of iFMI descriptor.

After doing FMI-SPOMF, we could get the rotation and scaling parameters(α and σ). There are last task of estimating translation parameter of these two images. It is easy to outperform. We just need to re-scale and re-rotation image $s(x, y)$, then using SPOMF with $r(x, y)$ to determines the translation (x_0, y_0) as explained in Section II-A. The image registration algorithm is summarized as follows:

1) execute the FMI(iFMI)-SPOMFcore algorithm,
2) compute the parameter (α, κ) of the maximum of $q_o(\theta, \lambda)$,
3) re-scale and re-rotation of image $s(x, y)$
4) calculate the SPOM between $r(x, y)$ and the re-scaled and re-rotated images,
5) locate the position (x_0, y_0) of the maximum,
6) output the parameters of the geometric transformation.

The algorithm flow is shown as Fig. 3 and Fig. 4.

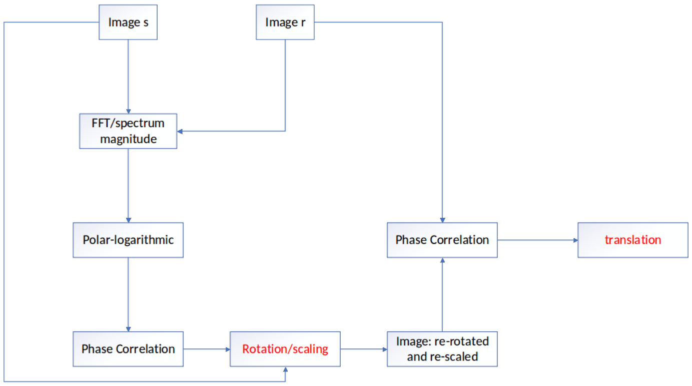


Fig. 3: The work flow of FMI-SPOMF image registration algorithm. The rotation, scaling and translation(with red words) are the final results of the algorithm.

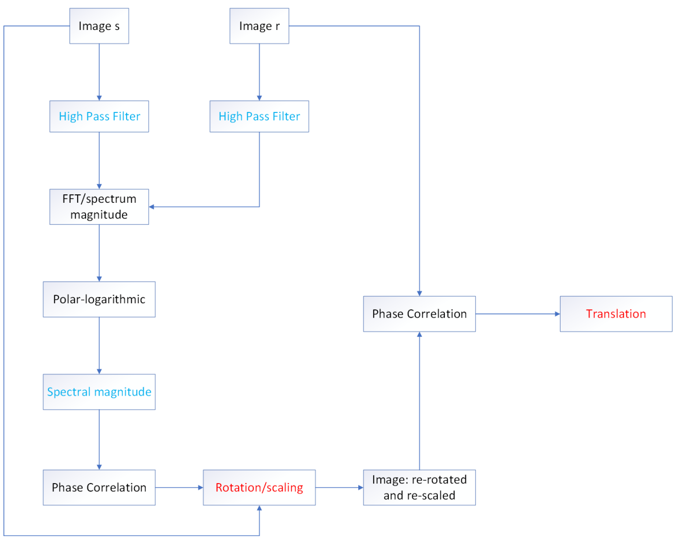


Fig. 4: The work flow of iFMI-SPOMF image registration algorithm. The modification of iFMI is blue items in the algorithm flow.

IV. Experiments and Results

In last section, we describe a image registration which is solved by using FMI(iFMI)-SPOMF algorithm. In this section, we will show our experiments about image registration of two synthetic images(shown as Fig.5a and 5b) and two real images(capture by using phone, shown as Fig.6a and 6b). Then, we will show the final results about registrations of two omni-camera images and the generated correspondences according to the transformation parameters. We implement the algorithm with Python 3.6 and OpenCV 3.4 based on a PC which has 16GB memory and Intel Core i7-6700 CPU with 3.4GHz.

A. Image Registration with FMI(iFMI)-SPOMF

The Fig.5 shows the experiment result of two synthetic images. We use a standard lena gray-level image(because spectral method isn't need to using color information), and translate, rotate and scale it to generate Fig.5b. Then, we use FMI and iFMI algorithm to registration with two images, the results are shown as Fig.5c and Fig.5d. We align the restored images with Fig.5a for convenient comparison results. Both method have a nice registration result, and it is clear to see, the performance of iFMI is better than FMI.

The Fig.6 shows the experiment with two real images. Fig.6c is the registration result of iFMI algorithm. We could see the results is almost restored to reference image.

(a) (b) (c) (d)

Fig. 5: The image registration results of two synthetic images. (a) is reference image $r(x, y)$ and (b) is image $s(x, y)$. (c) and (d) are results of FMI and iFMI algorithm respectively.

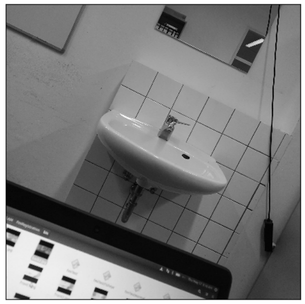
B. Extract Correspondences with Omni-camera Images

In our application, the iFMI image registration algorithm is used to found a set of correspondences with two frame of omni-camera images. We capture these images by using a Omni-Lens with phone, and align the omni images to 2D space(shown as Fig.7a and Fig.7b). The size of these images is 110×1100. The work flow of the method is

1) Input two omni-camera images $r(x, y)$ and $s(x, y)$,
2) divide $r(x, y)$ and $s(x, y)$ into ten equal parts,
3) outperform iFMI-SPOMF of each related parts with $r(x, y)$ and $s(x, y)$, get ten set of transformation parameters,
4) select a point in each part of $r(x, y)$(the center point is used in this experiment), calculate the corresponding point in each part of $s(x, y)$ according to transformation parameters.
5) generated ten set of corresponding points between $r(x, y)$ and $s(x, y)$(called correspondences).

The result of the method is shown as Fig.7. The ten correspondences are found by iFMI algorithm. It is satisfy conditions of computing VO.

(a)



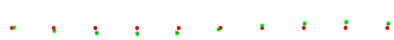
(b)

(c)

Fig. 6: The image registration results of two real images. (a) is reference image $r(x, y)$ and (b) is image $s(x, y)$. (c) iFMI algorithm.

(a)

(b)



(c)

Fig. 7: The image registration results of two omni-camera images. (a) is reference image $r(x, y)$ and (b) is image $s(x, y)$. (c) ten sets of correspondences.

V. Conclusion

In this project, we implement image registration algorithm by using iFMI with omni-camera model to found the correspondences between two continue frame. We compare FMI and iFMI algorithm between two synthetic images and prove the iFMI has better performance. We implement it omni-camera images, and get ten set of accuracy correspondences. The method is very suit to found correspondences of omni-camera image. And it is more fast and robust than feature-based method.

Acknowledgment

This work is a course project of CS270 in ShanghaiTech university. Thanks for our course advisor and every TAs. We also gratefully acknowledge my laboratory colleague provide phone images and omni-camera images for my experiment.

References

[1] S. Thrun, "Probabilistic robotics," *Communications of the ACM*, vol. 45, no. 3, pp. 52–57, 2002.

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

[3] Q.-s. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 12, pp. 1156–1168, 1994.

[4] T. Kazik and A. H. Göktoğan, "Visual odometry based on the fourier-mellin transform for a rover using a monocular ground-facing camera," in *Mechatronics (ICM), 2011 IEEE International Conference on*. IEEE, 2011, pp. 469–474.

[5] S. Thoduka, F. Hegger, G. K. Kraetzschmar, and P. G. Plöger, "Motion detection in the presence of egomotion using the fourier-mellin transform," in *Robot World Cup*. Springer, 2017, pp. 252–264.

[6] H. Bulow, A. Birk, and V. Unnithan, "Online generation of an underwater photo map with improved fourier mellin based registration," in *OCEANS 2009-EUROPE*. IEEE, 2009, pp. 1–6.

[7] A. V. Lugt, "Signal detection by complex spatial filtering," *IEEE Transactions on information theory*, vol. 10, no. 2, pp. 139–145, 1964.

[8] R. N. McDonough and A. D. Whalen, *Detection of signals in noise*. Academic Press, 1995.

[9] J. L. Horner and P. D. Gianino, "Phase-only matched filtering," *Applied optics*, vol. 23, no. 6, pp. 812–816, 1984.

[10] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, 1981.

[11] Q.-S. Chen, "Image registration and its applications in medical imaging," *PhD hesis, Vrije Universiteit Brussel*, 1993.

[12] R. C. Gonzalez, R. E. Woods *et al.*, "Digital image processing," 2002.

[13] E. De Castro and C. Morandi, "Registration of translated and rotated images using finite fourier transforms," *IEEE Transactions on pattern analysis and machine intelligence*, no. 5, pp. 700–703, 1987.

[14] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE transactions on image processing*, vol. 11, no. 3, pp. 188–200, 2002.

[15] Y. Sheng and H. H. Arsenault, "Experiments on pattern recognition using invariant fourier–mellin descriptors," *JOSA A*, vol. 3, no. 6, pp. 771–776, 1986.

[16] H. Bülow and A. Birk, "Fast and robust photomapping with an unmanned aerial vehicle (uav)," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 2009, pp. 3368–3373.