DIRECTIONAL DESCRIPTORS USING ZERNIKE MOMENT PHASES FOR OBJECT ORIENTATION ESTIMATION IN UNDERWATER SONAR IMAGES

Naveen Kumar[1], *Adam C. Lammert*[1], *Brendan Englot*[2], *Franz S. Hover*[2], *Shrikanth S. Narayanan*[1*]

[1]Signal Analysis and Interpretation Laboratory(SAIL), Department of Electrical Engineering,
University of Southern California, Los Angeles, CA, USA

[2]Hover Research Group, Department of Mechanical Engineering
Massachusetts Institute of Technology, Cambridge, MA, USA

[1]{komathnk,lammert}@usc.edu,shri@sipi.usc.edu
[2]benglot@mit.edu,hover@mit.edu

ABSTRACT

Conventional methods for rotation angle estimation are not very robust to variations in object shape or intensity. However in real object recognition scenarios like in underwater sonar images, the object seldom retains the same appearance in different test cases. Object representation using Zernike moments allows to capture these variabilities in a way that makes it robust in the context of rotation angle estimation. This paper presents a novel way to exploit the phase information of Zernike moments to infer the object orientation. This is achieved via a compact directional representation that describes the variation in object shape along different directions. Results yielded on the DIDSON sonar imageset collected by CSAIL at MIT show that the method can robustly infer the relative orientation between objects.

***Index Terms*—** Zernike Moment Phase, Rotation Angle, Underwater Sonar Image, Directional Descriptors

1. INTRODUCTION

Object recognition applications often require the algorithms to be invariant to rotation, translation or occlusion. These image distortions are often interdependent, although for processing simplicity, they need to be tackled independently of one another, at least initially. Inferring the relative orientation is a principal requirement in such systems. Further more there may be systems, that specifically rely on robust rotation angle estimation for a critical task. For example, this is critical for underwater or aerial vehicles where drastic changes in orientation are generally expected. Indeed, our current target application domain comprises underwater sonar images.

Unlike images acquired by optical cameras, sonar images undergo distortions due to a number of reasons. These could include factors like hardness/texture of the object, characteristics of the medium or even grazing angle of the sonar beam. This makes it challenging to use standard computer vision techniques directly. The clutter around objects for example makes extraction of point features difficult and hence popular local descriptors like SIFT [1] are rendered not effective. It is clear that conventional object descriptors are not appropriate for this imaging modality, and that there is room for future development of object description in the sonar domain.

But inspite of these problems, sonar has certain advantages like low power consumption and low cost which makes it preferable over other imaging methods like laser scanners or optical cameras, and hence the problem of object recognition in these images has become more important. Our specific goal in this paper is to robustly estimate the relative orientation, paving a way for adopting more suitable algorithms for object recognition in future work.

In the past, the method of principal axes has been used for estimating rotation angles. However this method has a drawback in the form of dependence on eigenvalues. Thus it fails for shapes that do not have a well defined direction [2]. Besides in the context of underwater sonar images, it becomes important to use methods that will be able to generalize well over the multiple sources of variabilities that these images exhibit. One well known representation that is promising in this regards relates to Zernike moments. Specifically they use an orthogonal representation where the degree of representation controls the degree of generalizability. Additionally these moments have very convenient rotational properties which made them popular in object recognition. The use of Zernike moment magnitude as a rotationally invariant feature has thus been extensively studied [3],[4],[5]. However using only Zernike moment magnitudes might not be the best way to represent object shapes [6]. Alternatively the relative orientation of the object can also be inferred using the phases of Zernike moments [7],[6],[8]. In this paper we propose a robust framework, to estimate the relative object orientation, suited to the domain of underwater sonar images. The proposed scheme is evaluated on a DIDSON dataset of forward looking sonar images.

2. ZERNIKE MOMENTS

Zernike polynomials were first proposed for diffraction theory in Optics by Zernike [9]. They quickly gained popularity in the image recognition and computer vision communities because of their interesting properties. Zernike moments unlike ordinary centralized moments are orthogonal which make them ideal basis functions for representing images. The orthogonality ensures less redundant representations since the contribution of each moment is independent. This is helpful for our problem because underwater communication

[*]Supported via ONR grant N00014-09-1-0700

channels usually do not offer the most optimal of bandwidths, because of severe attenuation and other channel non linearities.

The rotational invariance property of Zernike moment magnitudes is a result of how these moments are defined in the polar domain. Zernike basis functions are defined as a complete orthogonal set inside the unit circle : $x^2 + y^2 = 1$. Let the complex Zernike polynomials be represented as $V_{nm}(x, y)$ in the cartesian coordinates. Equivalently in (ρ, θ) polar coordinates they can be decomposed into the product of a radial polynomial $R_{nm}(\rho)$ and a phase term depending on θ.

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho)e^{-jm\theta} \tag{1}$$

The index n is constrained to be a positive number or zero, while m can take positive or negative integer values subject to the constraints that $n- \mid m \mid$ is even and $\mid m \mid \leq n$.The radial polynomial can be written as

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!}\rho^{n-2s} \tag{2}$$

Once the Zernike basis functions have been computed, the moments can be simply computed by projecting the image function onto them. Particularly a moment of order n, m can be computed as

$$A_{nm} = \frac{n+1}{\pi}\sum_x\sum_y f(x, y)V^*_{nm}(\rho, \theta), x^2 + y^2 \leq 1 \tag{3}$$

To analyze the effect of rotation we observe equations (1) and (3). A rotation by an angle α manifests itself as an extra phase in the Zernike moments. If A^{α}_{nm} are the Zernike moments after a counterclockwise rotation by α then,

$$A^{\alpha}_{nm} = A_{nm}e^{-jm\alpha} \tag{4}$$

Since there is only a difference in phase due to rotation of the pattern, the magnitude of the Zernike moments remains invariant to rotation and is popularly used as an object descriptor.

$$||A^{\alpha}_{nm}|| = ||A_{nm}|| \tag{5}$$

For selecting the highest order of Zernike moments, we must consider the fact that our goal is robustness. It can be seen through reconstruction [3] that low order Zernike moments contain coarser shape descriptions which would generalize well to noise or deformations present in the sonar images. Hence it is desirable to choose a low order Zernike description. Through similar reconstruction experiments Zernike moments of order 10 were found to be sufficiently informative for representing objects in the underwater sonar images in our study.

3. ZERNIKE PHASE AND DIRECTIONAL INFORMATION

Directional information in the image is important for many applications like textural analysis, image coding or compression. The technique proposed here uses this information for rotation angle estimation. We use the phases of Zernike moments to compute a directional representation for the object. This directional description can then be used to infer the relative orientation between two objects.

It was shown in the previous section that the magnitudes of Zernike moments are invariant to changes in rotation. Hence the rotation is encoded in the phases of Zernike moments. Specifically, if ϕ_{nm} indicate the phase/argument associated with the complex Zernike moment A_{nm} and ϕ^{α}_{nm} be the phase after a counter clock wise rotation by angle α, then it follows from equation (4) that

$$\phi^{\alpha}_{nm} = \phi_{nm} - m\alpha \tag{6}$$

From Eq.(6) the problem of finding α seems trivial. However, in a sonar image the pattern will usually be embedded in considerable noise. Computing α then becomes an estimation problem. Kim and Kim [7] proposed a probabilistic approach, to resolve this ambiguity. In the proposed approach we encode the Zernike moment phases into a compact representation which can capture the directional variation in object shape.

4. DESIGN OF THE DIRECTIONAL DESCRIPTOR

We first consider the phase function under rotation. Equation (6) comprises two phase components: a) ϕ_{nm}, the observed phase from the reference image and b) $m\alpha$, the shift in phase due to rotation. Before designing a descriptor using these it is important to realize that ϕ_{nm} is a wrapped phase whereas $m\alpha$ is an unwrapped phase term. ϕ^{α}_{nm} will also be observed as a wrapped phase from the image.

To resolve this incompatibility between input and reference phases we use trigonometric functions of the phases. This would effectively wrap both the input and the reference phases making them compatible. Particularly sin function was selected to ensure bounded valued features. For a given orientation of the object we compute the following

$$J(\alpha) = \sum_{n=0}^{N}\sum_{m} sin^2\left(\frac{\phi_{nm} - m\alpha}{2}\right) \tag{7}$$

where the indices n and m are constrained as before

For a given object the $J(\alpha)$ so obtained is the aforementioned directional descriptor for the object. This concise representation effectively captures the variation in the object shape at different angles of rotation, where the appearance of the object at some specific angle θ is given by a single number $J(\theta)$ lying between 0 and z, where z is the total number of Zernike moments of all orders. Given this formulation, we now try to examine the effect of rotation on the directional descriptors. The directional descriptors now obtained would be a circularly shifted version of the original one (Fig. 1). The problem of

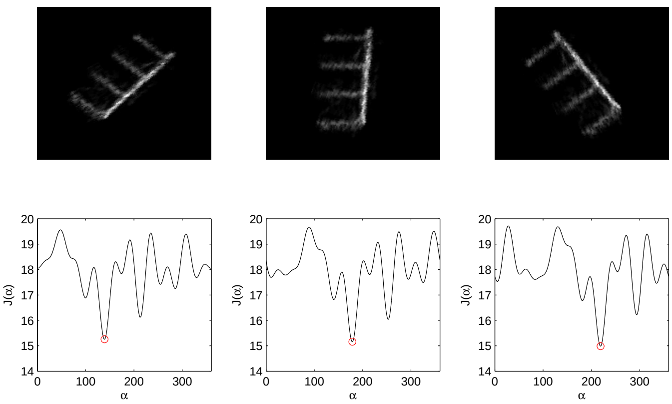
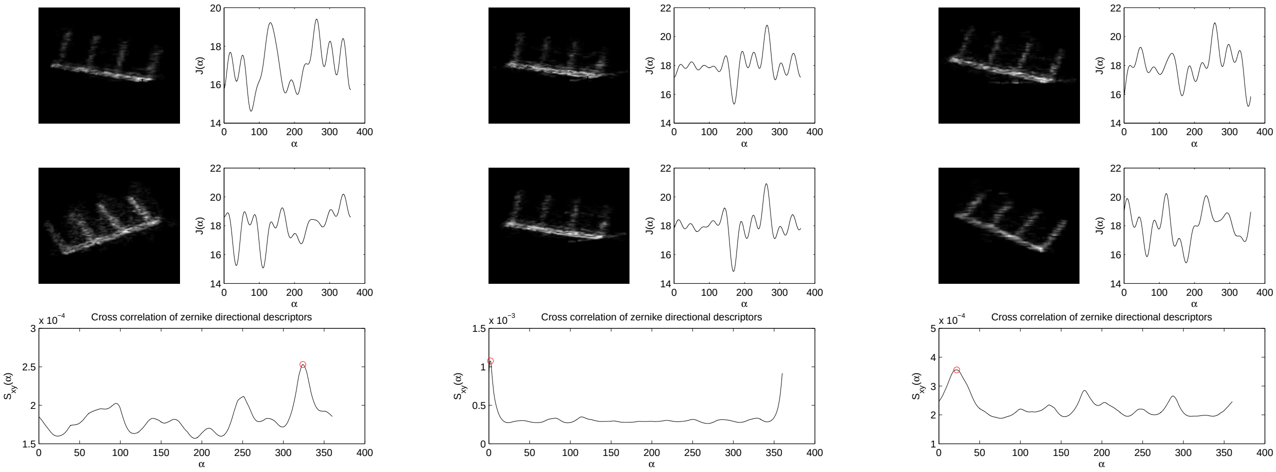


Fig. 1. An object synthetically rotated at angles of 30°, 70° and 110° and their corresponding directional descriptors. Note the shift in the directional descriptor due to rotation. *The circled point indicates a common point in each curve.*



(a) Clockwise rotation (b) No rotation (c) Counter-clockwise rotation

Fig. 2. Rotation angle estimation examples on objects selected from the sonar image dataset. *Peak of the cross correlation of descriptors corresponds to the retrieved rotation angle.*

rotation angle estimation has thus now been reduced to a problem of 1-D shift estimation.

It is worth noting here that to generate a complete directional map of the object, Zernike moments in only one initial orientation need to be computed. Because of the rotational property of Zernike moments we can analytically generate $J(\alpha)$ for the entire range. Also verify that $J(\alpha)$ is periodic with period 2π i.e. $J(\alpha + 2\pi) = J(\alpha)$. This is consistent with the fact that rotation by an angle $2\pi + \theta$ is the same as rotation by an angle θ.

5. ROTATION ANGLE ESTIMATION

Use of the directional descriptors has enabled us to effectively reduce the problem of rotation angle estimation to a 1-D circular shift estimation problem. Inspite of the robustness imparted by the Zernike moments, the directional descriptors will rarely be just pure shifts of each other. Distortions in the sonar images and occlusion of the objects will cause the curves to drift from their ideal shape. It can be seen in Fig. 1 how even for synthetically rotated objects, there is some variation in the descriptors because of interpolation errors in rotation of discrete images, although the overall trend of the curves is similar.

Hence the problem is more of an optimal shift estimation between two given curves, thus optimality has to be defined with respect to some criterion function. We start with the most simple metric: cross correlation function (CCF). Since cross correlation gives a measure of similarity of two signals at different shifts, we would expect to find a maxima at the point of optimal shift. To make a more robust estimation we turn to a method that is used in speech signal processing for pitch detection [10]. Average Magnitude Difference Function (AMDF) is often used as an alternative measure of cross correlation. Here we use a circular variant of the same.

$$D_{xy}(k) = \frac{1}{N} \sum_{n=0}^{N-1} \mid x(mod(n+k, N)) - y(n) \mid \tag{8}$$

where the signals $x(n)$ and $y(n)$ are of length N and $mod(a, b)$ represents the modulo operator.

Note that unlike circular cross correlation, circular AMDF would have a minimum value at the point of optimal shift. AMDF is a popular option when it comes to time delay estimation [11] because in the absence of any multiplications the computational complexity is significantly reduced. To retain the best of both we use the ratio of CCF to AMDF as the metric to estimate the shift. If $C_{xy}(k)$ indicate the circular cross correlation function, then $S_{xy}(k)$ is the similarity metric used

$$S_{xy}(k) = \frac{C_{xy}(k)}{D_{xy}(k)} \tag{9}$$

At the point of optimum shift, the minima in the AMDF would further emphasize the peaks in the CCF, giving sharper peaks.

6. EXPERIMENT AND RESULTS

Fig. 2 shows a complete rotation angle estimation problem. Given input and reference objects, the directional descriptor $J(\alpha)$ is first computed for both of them. The shift can then be estimated using the similarity metric described above. $S_{xy}(\alpha)$ gives a peak corresponding to the relative orientation between the objects. The rotation angle can hence be estimated.

The technique was tested on objects extracted from a dataset of forward looking sonar images collected by a DIDSON [12] sonar at CSAIL, MIT. A fuzzy segmentation is first performed on these images, followed by a connected component analysis. By setting a very low threshold, fuzzy segmentation ensures that for each frame, the object is completely segmented inside the blob. The oversegmentation can be dealt with by using features invariant to it. For example the center of mass or first moment of the blob which is used as the center for computing Zernike moments, is invariant to such oversegmentation, since the intensity of sonar reflected by the clutter around an object is typically much lower as compared to the object itself. The same is true for Zernike moments.

The method was compared both for the cases of a) same object in different frames and b) different but similar objects. An experiment was performed on a set of 81 objects extracted semi automatically from the underwater sonar images. One of the objects was selected from the set as a reference object and its relative orientation with all other objects was estimated and compared against manually tagged rotation angles. For objects that had been well segmented the estimation was fairly robust to changes in intensity or small occlusions, and the estimation error was within $\pm 10^\circ$. However for some objects, the estimation errors were large when the segmented blob also included other artifacts around the object of interest, thereby throwing the estimate of the center off by a large margin. This is infact a drawback of methods that work in the polar domain. The experiment was repeated several times with different reference objects. Fig. 3 shows the distribution of error for one such reference object. The distribution appears to be multimodal with a median absolute error of 6°. In order to better describe this we also fit a 3 component Gaussian mixture model to the histogram. A central Gaussian is observed ($\mu = 1.82^\circ$, $\sigma = 3.61^\circ$) which corresponds to the error for well segmented objects with only moderate occlusions. Another mode is observed (μ=–101.93°, $\sigma = 29.8^\circ$) corresponding to the objects that were heavily occluded. Observing a mode around $\pm 90^\circ$ is not surprising because the objects in this data set have arms at right angles. The third mixture component ($\mu = 40.6^\circ, \sigma = 75.4^\circ$) accounts for the rest of the probability mass. The absolute estimation error was less than 6° for 50% of the objects. Additionally, a z-test for mean was performed with the null hypothesis- $H_0 : \mu = 0$, which yielded a p-value of 0.9762 which is significant at the 5% level. This suggests that there is no sufficient evidence to reject the null hypothesis.

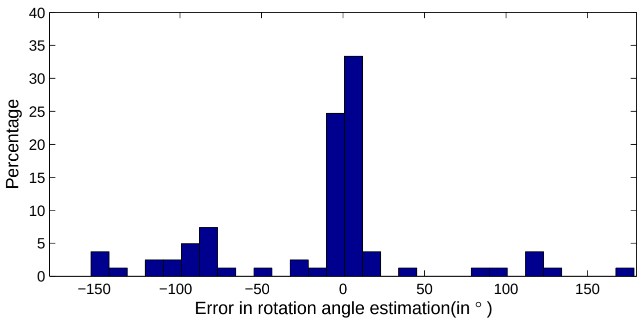


Fig. 3. Histogram of estimation error on the dataset shows that the error distribution is multimodal

7. CONCLUSION

In this paper we proposed directional descriptors for an object as a compact representation of the variation in object shape at different orientations. Apart from the robustness that these descriptors inherit from Zernike moments, their computation is also considerably simplified due to the rotational properties of these moments. The entire directional description can be thus be generated using only a few Zernike moment phases, the number of which we deliberately keep low to decrease sensitivity to inherent noise present in sonar images. It was further shown how this technique reduces the problem of relative orientation estimation to that of a 1-D circular shift estimation. By using correlation techniques an optimal shift can be calculated to estimate a rotation angle in the range [0,2π).

In conclusion this method addresses two of the key issues that are faced in the processing of underwater sonar images. It provides a framework for relative orientation estimation that is insensitive to the noise or deformations in these images. This is partially achieved by using lower order Zernike moments and partly by the simple directional representation which abstracts any unnecessary details that might not generalize well. Secondly, the compaction that is obtained in these representations as a result, can also reduce communication overheads in the underwater acoustic channel.

The method was experimentally tested using a DIDSON dataset of sonar images, and the results are promising. Future work will consider, among other things, the incorporation of these descriptors within an object recognition system.

8. REFERENCES

[1] D.G. Lowe, “Object recognition from local scale-invariant features,” in *ICCV*. Published by the IEEE Computer Society, 1999, p. 1150.

[2] W.H. Tsai and S.L. Chou, “Detection of generalized principal axes in rotationally symmetric shapes,” *Pattern Recognition*, vol. 24, no. 2, pp. 95–104, 1991.

[3] A. Khotanzad and Y.H. Hong, “Invariant image recognition by Zernike moments,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 489–497, 1990.

[4] A. Khotanzad and Y.H. Hong, “Rotation invariant image recognition using features selected via a systematic method,” *Pattern Recognition*, vol. 23, no. 10, pp. 1089–1101, 1990.

[5] O. Pizarro and H. Singh, “Toward large-area mosaicing for underwater scientific applications,” *Oceanic Engineering, IEEE Journal of*, vol. 28, no. 4, pp. 651–672, 2004.

[6] J. Revaud, G. Lavoué, and A. Baskurt, “Improving zernike moments comparison for optimal similarity and rotation angle retrieval,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 4, pp. 627–636, 2009.

[7] W.Y. Kim and Y.S. Kim, “Robust rotation angle estimator,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 8, pp. 768–773, 2002.

[8] Z. Chen and S.K. Sun, “A zernike moment phase-based descriptor for local image representation and matching,” *Image Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 205–219, 2009.

[9] F. Zernike, “Diffraction theory of the cut procedure and its improved form, the phase contrast method,” *Physica*, vol. 1, no. 689-704, pp. 56, 1934.

[10] M. Ross, H. Shaffer, A. Cohen, R. Freudberg, and H. Manley, “Average magnitude difference function pitch extractor,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 22, no. 5, pp. 353–362, 2003.

[11] G. Jacovitti and G. Scarano, “Discrete time techniques for time delay estimation,” *Signal Processing, IEEE Transactions on*, vol. 41, no. 2, pp. 525–533, 2002.

[12] E. Belcher, W. Hanot, and J. Burch, “Dual-frequency identification sonar (DIDSON),” in *Underwater Technology, 2002. Proceedings of the 2002 International Symposium on*. IEEE, 2002, pp. 187–192.