MIGRATING ORTHOGONAL ROTATION-INVARIANT MOMENTS FROM CONTINUOUS TO DISCRETE SPACE

Huibao Lin, Jennie Si

Department of Electrical Engineering
Arizona State University
Tempe, AZ 85287

Glen P. Abousleman

General Dynamics C4 Systems
8201 E. McDowell Road
Scottsdale, AZ 85257

ABSTRACT

Orthogonality and rotation invariance are important feature properties in digital signal processing. Orthogonality enables a target to be represented by a compact number of features, while rotation invariance results in unique features for a target with different orientations. The orthogonal, rotation-invariant moments (ORIMs), such as Zernike, Pseudo-Zernike, and Orthogonal Fourier-Melling moments, are defined in continuous space. These ORIMs have been digitized and have been demonstrated effectively for some digital imagery applications. However, digitization compromises the orthogonality of the moments, and hence, reduces their precision. Therefore, digital ORIMs are incapable of representing the fine details of images. In this paper, we propose a numerical optimization technique to improve the orthogonality of the digital ORIMs. Simulation results show that our optimized digital ORIMs can be used to reproduce subtle details of images.

1. INTRODUCTION

Moments are widely used for signal processing because they capture the global information of the signal. Among various moments that have been proposed, orthogonal, rotation-invariant moments (ORIMs) have received special attention. The most important type of ORIMs, Zernike moments, were proposed by F. Zernike as the eigenfunction of certain second-order partial differential equations [1]. Bhatia and Wolf found that Zernike moments can also be derived from Legendre polynomials with certain constraints. Using the same technology in different settings, they obtained Pseudo-Zernike moments [2]. Recently, Sheng and Shen proposed Orthogonal Fourier-Mellin moments (OFMMs) [3].

Zernike moments, Pseudo-Zernike moments, and OFMMs are developed in continuous space. They have complex coefficients and are defined over the unit disk, i.e., $0 \leq \rho \leq 1, 0 \leq \theta < 2\pi$. The magnitude of the moments is invariant to the rotation of the signal, while the orientation information is conveyed only by the phase. The rotation invariance of the magnitude is useful in pattern recognition applications because it eliminates the difficulty in calculating the orientation of the targets. The coefficients of ORIMs are orthogonal to each other, and this orthogonality has several important consequences. First, the moments are non-redundant, and thus fewer moments are needed to represent a signal than non-orthogonal moments. Second, signal reconstruction is much easier for orthogonal moments. Finally, the more moments that are used, the better the reconstruction will be; and, consequently, every signal can be accurately reconstructed by using a set (possibly infinite) of these moments.

This work was supported by General Dynamics C4 Systems, and in part by the NSF under grant ECS-0002098.

For applications in discrete space, the ORIMs are digitized both geometrically and numerically. Specifically, a set of pixels is selected to resemble the unit disk. For each moment, a coefficient is chosen for each pixel, although originally, different locations within a pixel usually have different coefficients. The direct methods determine the coefficients for a pixel by direct sampling of the ORIMs at one or several locations. For example, the simplest direct method uses ORIM coefficients at the pixel center for that pixel.

The digital ORIMs resulting from simple direct methods have been shown to be effective in some digital image processing applications, such as image indexing [4], patten recognition [5], target orientation estimation, and image compression, and have been proven to be noise resilient [6, 7]. However, the orthogonality of ORIMs is compromised during the digitization process, and hence, limits their precision. For example, if a certain number of moments are used to reconstruct an image, the inclusion of additional moments may degrade the result. Therefore, the digital ORIMs from direct methods cannot represent the fine details of an image, and they cannot be used to distinguish images with subtle difference.

Liao and Pawlak suggested several reasons for the loss of accuracy of digital Zernike moments derived from simplest direct methods [8]. First, the total area of the chosen pixels is different from that of the unit disk (geometric error). Second, the coefficients sampled at the pixel center are different from the average coefficients in that pixel area (numerical error). They proposed the use of a circle smaller than the inscribed circle of the image to select pixels, and the use of numerical integration rather than one-point sampling.

In our investigation, we found that no matter what size of circle is used, nor what type of numerical integration formulas are used, the digital ORIMs derived from direct methods are different from those in continuous space, and this difference prevents the digital ORIMs from accurately reconstructing the signal.

In this paper, we use the coefficients derived from direct methods as an initial estimate, and we then use numerical optimization to improve the orthogonality, while preserving the rotation invariance. We use the limited-memory BFGS algorithm [9] for optimization because of its fast convergence rate and low computational complexity. Experiments show that our optimized digital ORIMs have better orthogonality, and hence, better image reconstruction performance.

2. ORTHOGONAL ROTATION-INVARIANT MOMENTS IN CONTINUOUS SPACE

The ORIMs are developed in continuous space over the unit disk, D, i.e., $D = \{(\rho, \theta) \mid 0 \le \rho \le 1,\ 0 \le \theta < 2\pi\}$. Bhatia and Wolf [2] proved that the coefficient of every ORIM must be the product of a certain polynomial and a phase component, i.e.,

$$v_{nm}(\rho, \theta) = r_{nm}(\rho) p_{nm}(\theta), \tag{1}$$

where m is an integer, $p_{nm}(\theta) = e^{jm\theta}$, and $r_{nm}(\rho)$ is a polynomial in ρ of degree n, which contains no power of ρ lower than $|m|$. n is the "order" and m is the "repetition" of the moment.

The coefficients of the ORIMs are orthogonal to each other, i.e.,

$$\int\int_{(\rho,\theta)\in D} v_{nm}^*(\rho, \theta) v_{pq}(\rho, \theta) \rho d\rho d\theta = \frac{\pi}{n+1} \delta_{np} \delta_{mq}, \tag{2}$$

where $\delta_{np} = 1$ if $n = p$, and 0 otherwise.

The moment of order n, with repetition m, for a continuous 2-D signal, $f(\rho, \theta)$, that vanishes outside the unit disk is

$$a_{nm} = \frac{n+1}{\pi} \int\int_{(\rho,\theta)\in D} f(\rho, \theta) v_{nm}^*(\rho, \theta) \rho d\rho d\theta. \tag{3}$$

Due to the orthogonality of the coefficients, $v_{nm}(\rho, \theta)$, $f(\rho, \theta)$ can be easily reconstructed by

$$\hat{f}(\rho, \theta) = \sum_n \sum_m a_{nm} v_{nm}(\rho, \theta). \tag{4}$$

The more moments that are used in the reconstruction, the closer $\hat{f}(\rho, \theta)$ becomes to $f(\rho, \theta)$. Furthermore, every signal can be accurately reconstructed by a set of (possibly infinite) moments.

Rotating the signal, $f(\rho, \theta)$, does not change the magnitude of the ORIMs. For example, let the rotated signal be $f(\rho, \theta - \phi)$, then the corresponding moments will be $a_{nm}^{\phi} = a_{nm} e^{-jm\phi}$.

Zernike moments, Pseudo-Zernike moments, and OFMMs differ in the polynomials, $r_{nm}(\rho)$, and in the constraints between n and m. In particular, the polynomials of OFMMs are the same as those of Pseudo-Zernike moments when $m = 0$. Refer to [1, 2, 3] for detailed formulas.

3. DIGITAL ORIMS BY DIRECT METHODS

The digital ORIMs derived from direct methods are produced by digitizing ORIMs both geometrically and numerically. To illustrate, the unit disk, $D = \{(\rho, \theta) \mid 0 \le \rho \le 1,\ 0 \le \theta < 2\pi\}$, is digitized to $\hat{D}$, which consists of a set of pixels. For each moment, the coefficient values, $v_{nm}(\rho, \theta)$, over the area of a pixel are quantized into one $v_{nm,ij}$. This procedure is illustrated in Fig. 1.

The simplest direct method works as follows. Consider an image of $M \times M$ pixels. The coordinates of the pixel center, (x_i, y_j), can be calculated by assuming that the image is sampled from the square within $[-1, 1]^2$. Definition domain $\hat{D}$ is chosen to contain the pixels with centers falling on the unit circle, i.e.,

$$\hat{D} = \left\{ (i, j) | \sqrt{x_i^2 + y_j^2} \le \gamma \right\}, \tag{5}$$

where $\gamma = 1$ is the radius of the circle, and (i, j) represents the pixel at the ith row and the jth column. The coefficients are simply the ORIM values at the pixel centers. Specifically, $v_{nm,ij} =$

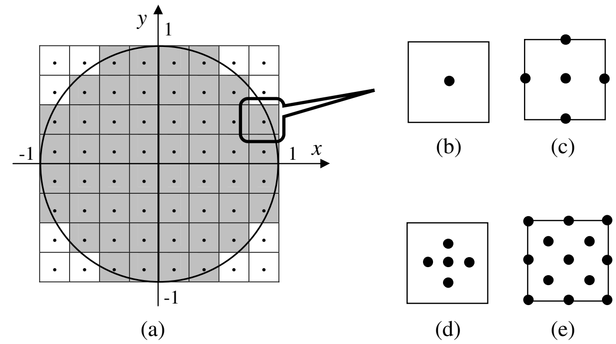


Fig. 1. Direct methods for converting ORIMs from continuous to discrete space. Each square in (a) represents a pixel, and each black dot represents the center of a pixel. The circle is the original definition domain, D, and is digitized into the shaded pixels, $\hat{D}$. The coefficients for each pixel are derived by sampling ORIMs at one or multiple locations within the pixel. (b), (c), (d), and (e) are the sampling locations for formulas 1-D, 5-D(I), 5-D(II), and 13-D(I & II), respectively.

$v_{nm}(x_i, y_j)$. Since a single sampling is used to calculate the coefficients, this method is referred to as the "1-D" formula.

The moments and signal reconstruction in discrete space are

$$a_{nm} = \frac{n+1}{\pi} \sum_{(i,j)\in\hat{D}} f_{ij}\, v_{nm,ij}^*, \tag{6}$$

and

$$\hat{f}_{ij} = \sum_n \sum_m a_{nm}\, v_{nm,ij}. \tag{7}$$

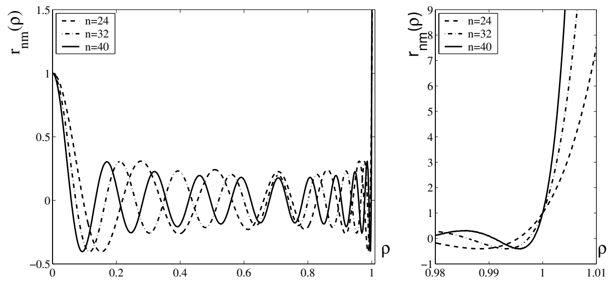
It can be seen from Figs. 1 and 2 that the selected pixels in $\hat{D}$ cannot exactly match the unit disk, D, and that the ORIM coefficients at the pixel center are usually different from the average coefficients over the entire pixel area. Hence, geometric error and numerical error are expected.

Liao and Pawlak [8] proposed improvements to the accuracy of Zernike moments by using a smaller circle for determining $\hat{D}$, and multiple point numerical integration formulas. The proposed radius for the circle is

$$\gamma = \sqrt{1 - 1/M - 0.0001}. \tag{8}$$

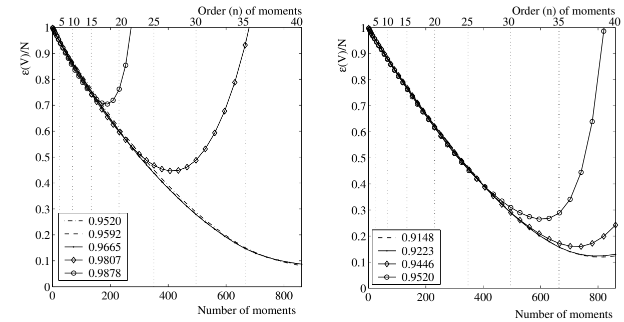
The proposed formulas are 5-D(I), 5-D(II), 13-D(I), and 13-D(II). The sampling locations of these formulas are illustrated in Fig. 1, with details provided in [8].

In our work, we show that finding the best radius for $\hat{D}$ is determined by the numerical formula adopted, and should be reduced for moments of higher order. From Fig. 2 we can see that 1) the polynomial of Zernike moments approaches infinity steeply when ρ is greater than 1, and 2) the polynomial changes more sharply when ρ approaches 1, or when n is larger. Because the reconstruction error is the summation of the geometric and the numerical errors, the best radius should be a tradeoff between the two. If n is small, the numerical error is trivial when the radius is smaller than 1, and the best radius would be close to (but less than) 1, since such radii would minimize the geometric error. In contrast, when n is large, the numerical error is dominant for pixels close to the unit circle, and reducing the radius will yield better performance.



(a) (b)

Fig. 2. Polynomials ($r_{nm}(\rho)$) of Zernike moments for m=0 and n=24, 32, 40, displayed for (a) $0 \leq \rho \leq 1$, and (b) around $\rho = 1$.



(a) (b)

Fig. 3. The normalized reconstruction error, $\epsilon(\mathbf{V})/N$, for digital ORIMs with different radii from (a) 5-D(II) formula, and (b) 13-D(I) formula.

In summary, the best radius is always less than 1, and is smaller when n is larger. Experimental results in Section 5 will verify our claim. The magnitudes of Pseudo-Zernike moments and OFMMs have the same properties as Zernike moments in 1), and less so in 2), but the best radii have similar properties.

Applying numerical integration formulas with more sampling points will reduce the numerical error. However, to avoid points falling outside the unit disk, the radius must be reduced, which consequently increases the geometric error. Hence, there is a limit to the accuracy of digital ORIMs derived from direct methods. As we will illustrate in our experiments, this limit is significant enough to limit the reconstruction accuracy.

4. DIGITAL ORIMS BY NUMERICAL OPTIMIZATION

Because the advantages of ORIMs come from the orthogonality and rotation invariance, the digital ORIMs do not have to be strictly similar to those in continuous space. In this paper, we use numerical optimization to improve the orthogonality of the digital ORIMs derived from direct methods, while preserving their rotation invariance.

For ease of discussion, we arrange the image and ORIMs into vector or matrix forms. Let the number of pixels in $\hat{D}$ be N. A column vector, $F = \{F_k | 1 \leq k \leq N\}$, is used to represent the digital imagery, $\{f_{ij} | (i,j) \in \hat{D}\}$. Let the number of moments be L. A matrix, $\mathbf{P}$, of size $L \times N$, is used to present the phase of the coefficients such that each row corresponds to a moment, and each column corresponds to a pixel. For example, $P_{lk} = p_{nm,ij}$. A matrix, $\mathbf{R}$, is constructed in the same format, but with normalized values,

$$R_{lk} = \sqrt{(n+1)/\pi}\ r_{nm,ij}. \tag{9}$$

The coefficients are

$$V_{lk} = R_{lk} P_{lk}, \ \ 1 \leq l \leq L, 1 \leq k \leq N. \tag{10}$$

If we define operation "$\circ$" to be the dot product between matrices, then

$$\mathbf{V} = \mathbf{R} \circ \mathbf{P}. \tag{11}$$

If the coefficients are orthonormal, then $\mathbf{V}^* \mathbf{V}^T = \mathbf{I}$. Thus, $\kappa(\mathbf{V}) = ||\mathbf{V}^* \mathbf{V}^T - \mathbf{I}||_F^2$ is a measure of the orthogonality of $\mathbf{V}$. Here, $|| \cdot ||_F$ is the Frobenius norm, i.e.,

$$\kappa(\mathbf{V}) = ||\mathbf{V}^* \mathbf{V}^T - \mathbf{I}||_F^2 = \sum_{i=1}^{L} \sum_{j=1}^{L} \left| \sum_{k=1}^{N} V_{ik}^* V_{jk} - \delta_{ij} \right|^2. \tag{12}$$

The moments are then

$$A = \mathbf{V}^* F. \tag{13}$$

The reconstructed image is

$$\hat{F} = \mathbf{V}^T A = \mathbf{V}^T \mathbf{V}^* F, \tag{14}$$

where $\mathbf{V}^*$ is the complex conjugate, and $\mathbf{V}^T$ is the transpose of matrix $\mathbf{V}$. The residue image is

$$\hat{F} - F = (\mathbf{V}^T \mathbf{V}^* - \mathbf{I}) F. \tag{15}$$

Define the reconstruction error for image F as

$$\epsilon(F) = \sum_{k=1}^{N} |\hat{F}_k - F_k|^2 \ / \ \sum_{k=1}^{N} F_k^2. \tag{16}$$

From Eqns. (15) and (16), we can see that the reconstruction error is determined both by $(\mathbf{V}^T \mathbf{V}^* - \mathbf{I})$ and the original image, F. For a given $\mathbf{V}$, if F happens to be an eigenvector of $(\mathbf{V}^T \mathbf{V}^* - \mathbf{I})$, the reconstruction error is large. On the contrary, if F falls in the null space of $(\mathbf{V}^T \mathbf{V}^* - \mathbf{I})$, the reconstruction error is zero. However, if the Frobenius norm of $(\mathbf{V}^T \mathbf{V}^* - \mathbf{I})$ is small, the reconstruction error will be small no matter what F is.

Define

$$\epsilon(\mathbf{V}) = ||\mathbf{V}^T \mathbf{V}^* - \mathbf{I}||_F^2. \tag{17}$$

Because $\epsilon(\mathbf{V})$ affects the reconstruction for all images, we use it as the measure of the reconstruction accuracy of $\mathbf{V}$. It can be proven that $\epsilon(\mathbf{V})$ and $\kappa(\mathbf{V})$ are different only by a constant quantity. Specifically,

$$\epsilon(\mathbf{V}) = \kappa(\mathbf{V}) + N - L, \text{ if } L \leq N. \tag{18}$$

(Due to page limitations, the proof of Eqn. (18) is omitted.) Hence, improving the reconstruction accuracy will improve the orthogonality, and vice versa. Because $\epsilon(\mathbf{V})$ is directly related to the image reconstruction accuracy, it is chosen as the objective for optimization.

The proposed optimization problem is a large-scale problem. As an example, let $M = 24$, and let the radius be chosen according to Eqn. (8). There would then be $N = 432$ pixels in $\hat{D}$. If $n = 40$, there would be $L = 861$ moments. Optimization of $432 \times 861 = 371,952$ variables is a computationally complex

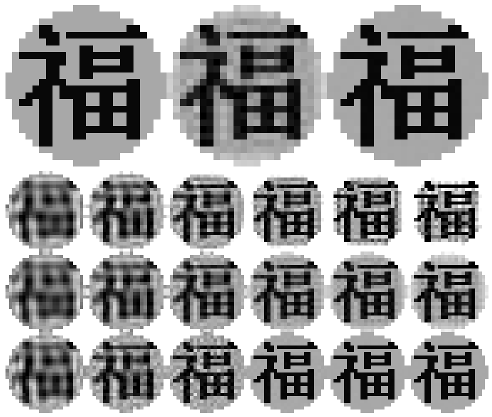


Fig. 4. The Reconstructed images. Top row from left to right: the original image, the best reconstruction (at $n = 33$) for 5-D(II), and the reconstruction (for any $n \geq 33$) by the optimized method, respectively; second to fourth row: the reconstruction for different n's by 1-D, 5-D(II), and the optimized method, respectively. From left to right, the images are reconstructed from moments up to $n =$ 15, 20, 25, 30, 35, and 40, respectively.

task. Accordingly, we chose the limited-memory BFGS (L-BFGS) algorithm, which is inexpensive to implement and fairly robust, but converges rapidly [9].

To avoid the complexity of computing complex numbers in the optimization procedure, we choose to optimize the polynomials, $\mathbf{R}$ rather than the coefficients, $\mathbf{V}$. The polynomials, $\mathbf{R}$, and phases, $\mathbf{P}$, calculated by the direct methods are close to the optimization solution, and are used as initial conditions.

To speed up the optimization, L-BFGS requires that the derivative of $\epsilon(\mathbf{V})$ be calculated with respect to $\mathbf{R}$, which can be proven (proof omitted due to page limitations) to be

$$\mathbf{G} = 4\Re\left\{\mathbf{V}^*\left(\mathbf{V}^T\mathbf{V}^* - \mathbf{I}\right)\circ\mathbf{P}\right\}, \tag{19}$$

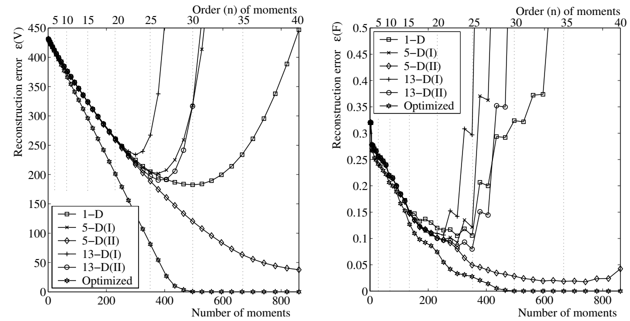
where $\Re\{\cdot\}$ is the real part of the variable.

5. EXPERIMENTAL RESULTS AND CONCLUSIONS

In the experiments conducted, Zernike moments up to $n = 40$ were used. The size of the testing imagery was 24×24, i.e., $M = 24$. For ease of comparison, the radius as defined in Eqn. (8) was used.

Fig. 3 shows the reconstruction error as defined in Eqn. (17), obtained with different radii. Because different radii involve different numbers of pixels in the computation, the reconstruction error is normalized by the number of pixels. Specifically, the curves are $\epsilon(\mathbf{V})/N$. At radius $\gamma = 0.9665$, all sampling points for 5-D(II) are within the unit disk (and $\gamma = 0.9223$ for 13-D(I)). The results show that at these radii, reconstruction accuracies are among the best. However, close inspection reveals that at lower n's, larger radii yield better results.

Fig. 4 shows the reconstructed images obtained by coefficients from 1-D, 5-D(II) (the best formula from [8]), and by our optimized method. The test image was chosen to be a character with uniform background and foreground intensities for easy inspection. From Fig. 4, we can see that the reconstructed images from 1-D and 5-D(II) have significant distortion. On the other hand, the reconstruction with our optimized coefficients is nearly identical to the original when n is larger than 33.



(a) (b)

Fig. 5. (a) The reconstruction accuracy of $\mathbf{V}$, and (b) the reconstruction error for the test image in Fig. 4.

Fig. 5 compares the reconstruction accuracy of $\mathbf{V}$ as defined in Eqn. (17), and the reconstruction error for test images as defined in Eqn. (16), for different formulas. From the figure, we can see that our optimized coefficients yield the best performance. Note also that for our optimized coefficients, the inclusion of more moments produces better reconstruction. In fact, when $n \geq 30$, the reconstruction error is almost zero.

Our experiments also include Pseudo-Zernike moments and OFMMs at different radii and on other images. The results also reveal that our optimized ORIMs are superior to the digital ORIMs from direct methods. However, these results are not be included due to space limitations.

6. REFERENCES

[1] F. Zernike, "Beugungstheorie des schneidenverfahrens und seiner verbesserten form, der phasenkontrastmethode," *Physica*, 1: 689–704, 1934.

[2] A. B. Bhatia and E. Wolf, "On the circle polynomials of zernike and related orthogonal sets," *Proc. Cambridge Philos. Soc.*, 50: 40–48, 1954.

[3] Y. Sheng and L. Shen, "Orthogonal fourier-mellin moments for invariant pattern recognition," *J. Opt. Soc. Am. A*, 11[6]: 1748–1757, 1994.

[4] D. Zhang and G. Lu, "Evaluation of MPEG-7 shape descriptors against other shape descriptors," *Multimedia Systems*, 9: 15-30, 2003.

[5] M. R. Teague, "Image analysis via the general theory of moments," *J. Opt. Soc. Am.*, 70[8]: 920–930, 1980.

[6] C.-H. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Trans. PAMI*, 10[4]: 496–513, 1988.

[7] A. Khotanzad and Y. H. Hong, "Invariant image recognition by zernike moments," *IEEE Trans. PAMI*, 12[5]: 489–497, 1990.

[8] S. X. Liao and M. Pawlak, "On the accuracy of zernike moments for image analysis," *IEEE Trans. PAMI*, 20[12]: 1358–1364, 1998.

[9] D. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical Programming B*, 45: 503–528, 1989.