On Image Analysis via Orthogonal Moments

Miroslaw Pawlak

Xinmeng Liao

Department of Electrical and Computer Engineering
The University of Manitoba
Winnipeg, Manitoba, Canada R3T 2N2

Abstract

In this paper we study the reconstruction properties of moment descriptors. The set of orthogonal moments based is employed. An asymptotic expansion for the global reconstruction error is established. This reveals mutual relationships between a number of moments, the image smoothness, sampling rate and noise model characteristics.

The problem of an automatic (data-driven) selection of an optimal number of moments is considered.

I. Introduction

Moment descriptors of various forms have been extensively employed as pattern features in scene recognition, registration, object matching as well as data compression, see [1], [4], [5], [9], [10], and references cited therein.

Some examples of moments include: geometric, complex, Fourier-Mellin, radial and orthogonal moments.

Their invariance properties along with numerous applications have been widely studied in the literature. However, the fundamental problems concerning the robustness for noise and digitizing have been rarely addressed [6], [9], [10]. Such issues can be carried out by verifying the accuracy of a reconstruction method derived from moments. That is, how close can we recover the original image from a finite set of moments computed from observed data? Certainly, the higher-order moments suffer greater degradation due to noise. On the other hand, they are able to supply the detail information about the image. These two opposite factors working against one another imply that there exists an optimal number of moments yielding the best possible representation

In this paper the inverse moment problem of reconstruction of the image from a finite set of moments derived from the discrete and noisy data is examined.
We use orthogonal moments since they have better recognition and reconstruction properties [6], [8], [9], [10]. Nevertheless, their geometrical interpretation is less intuitive.

The mean integrated-square reconstruction error between the image and its reconstructed version is evaluated . This reveals how the error, as well as, the optimal number or moments depend on the image smoothness, noise characteristic, and sampling rate. In particular, a relation between a number of moments and sampling rate yielding the convergence property (the error tends to zero) is established. The rate of convergence is computed and the optimal number of moments minimizing the error is determined.

The further problem we address is how the optimal number of moments can be selected directly from the observed data record. That is, a data-driven method of selecting the optimal number of moments is proposed. This is carried out by employing the cross-validation technique [3].

II. Moments and Image Reconstruction

In this section the definition of orthogonal moments is given and corresponding reconstruction algorithm derived from the discrete and noisy data is proposed.

Let us assume that an image f(x,y) is defined over the compact region S of R^2 ..
The moment of order $k + l$ is of the form

$$\lambda_{kl} = \iint \psi_{kl}(x,y)\ f(x,y)\ dx\ dy,$$

$0 \le k, l \le \infty$, where $\{\psi_{kl}(x,y)\}$ is a set of basis functions over S. Here and throughout the paper, if not otherwise stated, all integrals are in S.

The classical geometrical moments correspond to the case when $\psi_{kl}(x,y) = x^k y^l$ [4]. It is known the image reconstruction problem form these moments is strongly ill-posed [2], [8] and computationally very expensive.

Tague [9], see also [6], [10], has introduced the orthogonal moments employing the theory of orthogonal polynomials of two variables to overcome the computational burden associated with classical moments. In particular, the Legendre and Zernike polynomials have been used, however other orthonormal systems can be also utilized [7]. It should be emphasized that the reconstruction procedure based on the orthogonal polynomials does not need regularization [2], it simply adds the individual contribution of each order moment to generate the reconstructed image. The choice of number of moments plays the role of the regularization parameter.

Thus, let throughout the paper $\{\psi_{kl}(x,y)\}$ be an orthonormal complete basis over S, i.e., $\iint \psi_{k'l'}(x,y)\psi_{kl}(x,y)dxdy = 1$ if $(k, l) = (k', l')$ or 0 otherwise. It is well known [7] that $f(x,y)$ can be represented by

$$\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \lambda_{kl}\,\psi_{kl}(x,y) .$$

Hence, the reconstruction algorithm employing $(N+1)^2$ moments $\{\lambda_{kl}\}$ is of the form

$$\sum_{k=0}^{N}\sum_{l=0}^{N} \lambda_{kl}\,\psi_{kl}(x,y) .$$

Clearly, the square reconstruction error

$$I(N) = \iint [f_N(x,y) - f(x,y)]^2\, dx\, dy \qquad (1)$$

goes to zero as $N \to \infty$, i.e., by including higher order moments one can make the error arbitrarily small. This has been a commonly recommended prescription in the image processing literature. Nevertheless, this scheme breaks down if the image is contaminated by noise or/and only a finite number of observations is available and moments must be determined from imperfect data. Indeed, in that case an optimal number of moments exists, giving the minimal value of the reconstruction error, i.e., the error is initially decreasing (not necessarily in monotonic way) up to a certain number of moments and then it increases as $N \to \infty$. In order to verify such properties one has to assume some models for the underlying image as well as the distortion process.

Let us adapt the following simple image observation model

$$g(x, y) = f(x, y) + z(x, y),$$

for $(x,y), (x',y') \in \{(x_i,y_j): 1 \le i \le n, 1 \le j \le m\}$ - the set of nxm nodes of the grid. Here $z(x, y)$ is random process with zero mean and finite variance σ^2. $z(x,y)$ is spatially uncorrelated process. See [6] for more general observation models.

Suppose now that one measures moments of $g(x,y)$ up to the order 2N, i.e. $\bar{\lambda}_{kl} = \iint \psi_{kl}(x, y)\, g(x,y)\, dx\, dy$, $0 \le k, l \le N$.

Then, one can reconstruct $f(x,y)$ by using

$$\bar{f}_N(x,y) = \sum_{k=0}^{N}\sum_{l=0}^{N} \bar{\lambda}_{kl}\,\psi_{kl}(x, y) .$$

Let us observe that $E\bar{\lambda}_{kl} = \lambda_{kl}$ and consequently $E\bar{f}_N(x,y) = f_N(x,y)$. However, since only a finite number of observations is given, one has to use a certain discrete version of $\bar{\lambda}_{kl}$.

Let

$$\hat{\lambda}_{kl} = \sum_{i=1}^{n}\sum_{j=1}^{m} h_{kl}(x_i, y_j)\, g(x_i, y_i) \qquad (2)$$

be a discrete approximation of $\bar{\lambda}_{kl}$, where $\{h_{kl}(x,y)\}$ is a weight sequence (see for examples below).

That is, a linear weighted combination of $\{g(x_i, y_j), 1 \le i \le n, 1 \le j \le m\}$ can serve as an estimate of λ_{kl}.

Many prescriptions for $\{h_{kl}(x,y)\}$ are possible. One used in this paper has the form

$$h_{kl}(x_i, y_j) = \int_{x_i-\frac{\Delta x}{2}}^{x_i+\frac{\Delta x}{2}} \int_{y_j-\frac{\Delta y}{2}}^{y_j+\frac{\Delta y}{2}} \psi_{kl}(x, y)\, dx\, dy , \qquad (3)$$

where $\Delta x = x_i - x_{i-1}$, $\Delta y = y_j - y_{j-1}$ are sampling intervals in the x and y directions, respectively, and $x_1 - \frac{\Delta x}{2} = y_1 - \frac{\Delta y}{2} = -1$, $x_n + \frac{\Delta x}{2} = y_m + \frac{\Delta y}{2} = 1$. The coefficient $h_{kl}(x_i, y_j)$ represents the integration of $\psi_{kl}(x, y)$ over the pixel $\left[x_i - \frac{\Delta x}{2}, x_i + \frac{\Delta x}{2}\right] \times \left[y_j - \frac{\Delta y}{2}, y_j + \frac{\Delta y}{2}\right]$.

It is worth to note that the integral in (3) can be approximated very accurately using well-known techniques for numerical integration.

These considerations yield the following reconstruction algorithm

$$\hat{f}_N(x,y) = \sum_{k=0}^{N} \sum_{l=0}^{N} \hat{\lambda}_{kl} \psi_{kl}(x, y) \quad , \qquad (4)$$

where $\hat{\lambda}_{kl}$ is defined by (2) and (3). As a measure of reconstruction performance we choose the integrated-square error

$$J(N) = \iint [\hat{f}_N(x,y) - f(x,y)]^2 \, dx \, dy.$$

In the next section we give the error decomposition and its asymptotic evaluation. The problem of data-driven selection of N minimizing J(N) is studied in Section 4.

III. Assessing the Performance of the Reconstruction Algorithm

It is clear that by virtue of Parseval's formula we have

$$J(N) = \sum_{k=0}^{N} \sum_{l=0}^{N} \left(\hat{\lambda}_{kl} - \lambda_{kl}\right)^2 + I(N), \qquad (5)$$

where I(N) is defined in (1), i.e., it is the reconstruction error of the noise and sampling free image. The first term on the right-hand side of (5) can be viewed as a matching measure between $\{\lambda_{kl}\}$ and $\{\hat{\lambda}_{kl}\}$ based on the total $(N+1)^2$ moments. Clearly, this term either converges to a finite number or to infinity as $N \to \infty$. Its asymptotic behavior in the mean-sense is given in Theorem 1 below. Our results model the performance of the reconstruction procedure on grids which become increasingly fine, i.e., as $\Delta x \Delta y \to 0$.

Furthermore, it is plain that

$$E\left(\hat{\lambda}_{kl} - \lambda_{kl}\right)^2 = \mathrm{var}\left(\hat{\lambda}_{kl}\right) + \left(E\hat{\lambda}_{kl} - \lambda_{kl}\right)^2. \qquad (6)$$

Here, the first term is due to the noise, whereas the second one is caused by discrete approximation of λ_{kl}.

The asymptotic behavior of these terms is described by the following theorem.

Theorem 1. Suppose that f is a function of bounded variation on S with the total variation V(f). Then

$$EJ(N) \approx \sigma^2 (N + 1)^2 \Delta x \, \Delta y + MV(f) \, (N + 1)^2 \, \Delta x \, \Delta y + I(N), \qquad (8)$$

where $M = \max_{(x, y) \in S} f(x, y)$.

The proof of Theorem 1 is omitted, see [6].

The terms $\mathrm{var}(\hat{\lambda}_{kl})$ and $\left(E\hat{\lambda}_{kl} - \lambda_{kl}\right)^2$ from (6) are represented by the first two leading expressions in Theorem 1, respectively.

A special case resulting from Theorem 1 is of a particular interest.

Remark 1. Let the image f be observed under the noise-free condition, i.e. $g(x_i, y_j) = f(x_i, y_j)$, $1 \le i \le n$, $1 \le j \le m$. Then under the conditions of Theorem 1 we have

$$J(N) \approx MV(f) \, (N + 1)^2 \, \Delta x \, \Delta y + I(N) \quad .$$

Hence, not only the noise but also the discreteness of data yields the existence of an optimal N which corresponds to trading off between the first term and I(N). Note that $I(N) \to 0$ as $N \to \infty$.

The first two terms in the expansion of EJ(N) are of the same order. If, however, one can assume that the image f(x, y) is smooth then the second term in (8) is a smaller order than the first one.

To understand the rate at which EJ(N) can converge to zero and consequently to determine the moment order minimizing EJ(N) one has to evaluate the rate at which I(N) goes to zero. The latter necessity requires some assumptions about the smoothness of f(x,y). The smoothness of f(x,y) is most naturally expressed in terms of the order of magnitude of moments $\{\lambda_{kl}\}$. In fact we have $|\lambda_{kl}| \to 0$ as $k + l \to \infty$.

Hence, let

$$|\lambda_{kl}| = \sqrt{\eta} \, [(k + 1) \, (l + 1)]^{-\beta}, \qquad (9)$$

where $\eta > 0$, $\beta > 1/2$.

Functions which have "β derivatives" exhibit such a property, e.g., $\beta = 1$ corresponds to functions satisfying the local Lipschitz condition .
Some further considerations yield the following theorem.

Theorem 2. Let all of the conditions of Theorem 1 hold. Let (9) be satisfied. Then for

$$N^* = (\eta/2a)^{1/(2\beta+1)} (\Delta x \Delta y)^{-1/(2\beta+1)}$$

we have

$$EJ(N^*) \approx \Omega\ (\Delta x \Delta y)^{(2\beta-1)/(2\beta+1)},$$

where

$$\Omega = (2\beta+1)(2\beta-1)^{-1}\, 2^{-2/(2\beta+1)}\, (\eta^2 a^{2\beta-1})^{1/(2\beta+1)}$$

$$a = \sigma^2 + MV(f).$$

The N^* given in Theorem 2 is the moment order which minimizes $EJ(N)$, while $EJ(N^*)$ is the corresponding error value.

The formula given in Theorem 2 reveals that N^* is a decreasing function of both the image smoothness (β) and the sampling rate. The latter denotes that a high-resolution image requires a smaller number of moments for an optimal reconstruction.

In practice, however, the N^* proposed above cannot be used since it involves unknown characteristics of $f(x,y)$ as well as the noise process. The problem of how to choose N directly from the observed data record $\{g(x_i, y_j)\}$ is addressed in the next section.

IV. Data-Driven Selection of N

The problem which we address in this section is how to select a "good" N directly from the available data $\{g(x_i, y_j);\ 1 \le i \le n,\ 1 \le j \le m\}$. Ideally, one wishes to have N_0^* which minimizes the global error $J(N)$. Notice that N_0^* is a function of the data at hand.

This, in turn, is equivalent to taking the minimizer of the following criteria

$$J(N) - \iint f^2(x,y)dxdy = \sum_{k=0}^{N}\sum_{l=0}^{N}\left(\hat{\lambda}_{kl}\right)^2 - 2\sum_{k=0}^{N}\sum_{l=0}^{N}\lambda_{kl} \tag{10}$$

This, however, is infeasible since λ_{kl}'s are unknown. Note also that replacing λ_{kl} with $\hat{\lambda}_{kl}$ in (4.1) yields the unacceptable solution $N = \infty$.

To overcome this difficulty the resampling technique utilizing the cross-validation methodology has been introduced [6]. The asymptotic optimality of such selection has been proved.
Other techniques using discrete measures of the estimate performance are possible. For instance, instead of the $J(N)$ criterion, its discrete approximation

$$JD(N) = \Delta x \Delta y \sum_{i=1}^{n}\sum_{j=1}^{m} (f(x_i, y_j) - \hat{f}_N(x_i, y_j))^2$$

can be used [3].

This measures the discrepancy between $\hat{f}_N(x,y)$ and $f(x,y)$ only on the grid points $\{(x_i, y_j);\ 1 \le i \le n,\ 1 \le j \le m\}$, whereas $J(N)$ is the global criterion.

The empirical selectors corresponding to $JD(N)$ are of the form

$$\widehat{JD}(N) = \Delta x \Delta y \sum_{i=1}^{n}\sum_{j=1}^{m} \left(g(x_i, y_j) - \hat{f}_N(x_i, y_j)\right)^2 \Phi(N)$$

,i.e., it is a penalized version of the residual error $\Delta x \Delta y \sum_{i=1}^{n}\sum_{j=1}^{m} (g(x_i, y_j) - \hat{f}_N(x_i, y_j))^2$, see[3]. Common prescriptions for the penalty factor $\Phi(N)$ are (a) $(1 - L(N)\Delta x \Delta y)^{-2}$, (b) $(1 - L(N)\Delta x \Delta y)^{-1}$, where $L(N)$ is the total number of moments used in $\hat{f}_N$, e.g., $L(N) = (N+1)^2$ for the quadratic summation, or $L(N) = (N+1)(N+2)/2$ for the triangular one .
To illustrate the presented techniques for selection of N let us consider a simple simulation example. Figure 1 shows the original 21x21 binary image (grey levels 17 and 12, respectively) and its noisy version. A white Gaussian noise with variance 2 is added producing a noisy image.

Figure1. The original image and noisy version with white Gaussian noise of variance 2.

The reconstruction process is shown in Fig.2 starting from the noisy image and reconstructed images employing successively moments from order 1 to 23. A triangular summation has been used with the total $(N+1)(N+2)/2$ moments . The Legendre polynomial system has been employed .

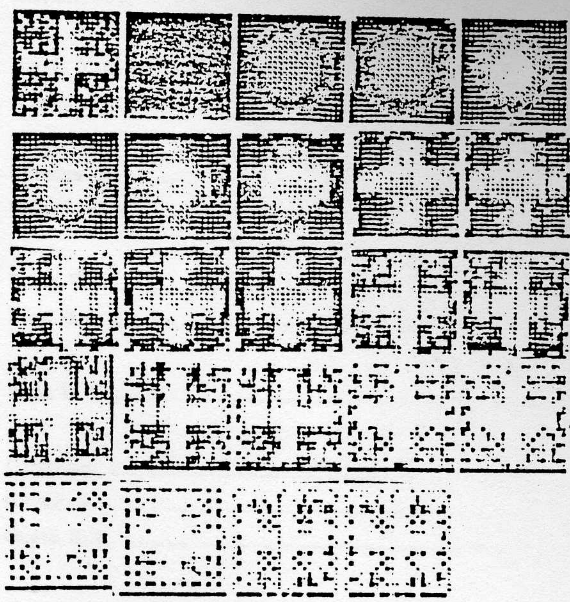

Figure 2. The noisy image and reconstructed images. From top row to bottom row and left to right: noisy image and reconstructed images with up to first order Legendre moment through up to twenty three Legendre order moment.

The corresponding reconstruction error JD(N) (averaged on 20 runs) is depicted in Figure 3. The error slowly decreases, reaches minimum at $N_d^* = 8$, and then it gradually increases up to the value N = 18. From N = 18 the error increases rapidly being , e.g., for N = 20 six times greater than for N = 8.

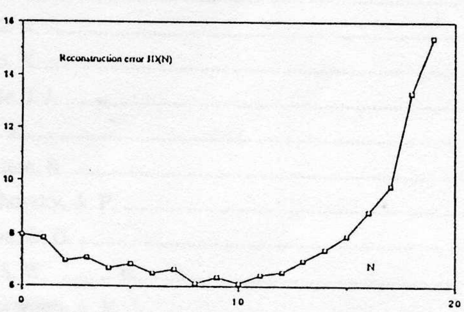

Figure 3 .The discrete reconstruction error versus N

Then, the automatic selectors of N have been calculated giving some estimates of N_d^* .First, the $\widehat{JD}(N)$ criterion with $\Phi(N) = (1 - L(N)\Delta x\Delta y)^{-2}$ has been applied yielding the value $\hat{N}_d = 10$. The relative error $err(N_d^*, \hat{N}_d) = (JD(\hat{N}_d) - JD(N_d^*))/JD(\hat{N}_d)$ of this choice is equal to 0.42%. For the selector $\widehat{JD}(N)$ with $\Phi(N) = (1 - L(N)\Delta x\Delta y)^{-1}$ we get $\hat{N}_d = 15$ with $err(N_d^*, \hat{N}_d) = 22.8\%$. Finally, the selector discussed in [6] has been implemented giving $\hat{N}_0 = 11$, with the relative error 4.89% .

It is worth to note that the visual inspection of reconstructed images (Fig.4) suggests that acceptable solutions lie between N = 8 and N = 13 .

REFERENCES

[1] Y.S. Abu-Mostafa and D. Psaltis, "Recognition aspects of moment invariants", IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-6, pp. 698-706, Nov. 1984.

[2] M. Bertero, T.A. Poggio and V. Torre, "Ill-posed problems in early vision", Proceedings of the IEEE, vol. 76, pp. 865-885, 1988.

[3] R.L. Eubank, **Spline Smoothing and Nonparametric Regression**, Marcel Dekker, New York, 1988.

[4] M.K. Hu, "Visual problem recognition by moment invariants", IRE Trans. Inform. Theory, vol. IT-8, pp. 179-187, Feb. 1962.

[5] A.K. Jain, **Fundamentals of Digital Image Processing**, Prentice-Hall, 1989.

[6] M. Pawlak, " On the reconstruction aspects of moment descriptors", IEEE Trans. Information Theory, 1992, to appear.

[7] G. Szegö, **Orthogonal Polynomials**, vol. 23, 4th Ed., American Mathematical Society Colloquie Publications, Providence, R.I., 1975.

[8] G. Talenti, "Recovering a function from a finite number of moments", Inverse Problems, vol. 3, pp. 501-517, 1987.

[9] M.R. Teague, "Image analysis via the general theory of moments", J. Optical Soc. Am., vol. 70, pp. 920-930, August 1980.

[10] C.H. Teh and R.T. Chin, "On image analysis by the methods of moments", IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-10, pp. 496-512, July 1988.