

# Recognitive Aspects of Moment Invariants

YASER S. ABU-MOSTAFA AND DEMETRI PSALTIS

**Abstract**—Moment invariants are evaluated as a feature space for pattern recognition in terms of discrimination power and noise tolerance. The notion of complex moments is introduced as a simple and straightforward way to derive moment invariants. Through this relation, properties of complex moments are used to characterize moment invariants. Aspects of information loss, suppression, and redundancy encountered in moment invariants are investigated and significant results are derived. The behavior of moment invariants in the presence of additive noise is also described.

**Index Terms**—Circular harmonics, complex moments, feature space, image analysis, information theory, moment invariants, pattern recognition.

## I. INTRODUCTION

THE PURPOSE of this paper is to provide an *analytic* characterization of moment invariants as features for pattern recognition. Moment invariants were introduced by Hu [1] as image recognition features which had the desirable property of being invariant under such variations of the image content as shift, scaling, and rotation. Since then, they have been given considerable attention in the literature and reports of their satisfactory experimental results (e.g., [5]), inaccuracies (e.g., [13]), extensions (e.g., [8]), and variations (e.g., [2]) have made it appropriate to settle the question *how good are moment invariants?* theoretically. Pattern recognition features such as moment invariants are considered “good” if their values are sensitive to the identity of the pattern being recognized (discrimination) but not to the noise encountered (robustness). An analytic framework to predict the discrimination and robustness of moment invariants in any given problem is described here.

In order to analyze moment invariants and derive formulas for different aspects of their performance, we need to define them in an analytically manageable way that will be the basis for further derivations. For this purpose, we introduce the notion of complex moments (CM’s) which are very simple entities and form an intermediate step between ordinary moments and moment invariants. Through the definition of CM’s and a straightforward derivation, the nature of moment invariants and their relation to the circular harmonic expansion [9] of the image become apparent. This leads to several results about the scope of discrimination between patterns provided by moment invariants and *when* they should be expected to perform well. Another formula is derived to relate the second-order statistics of CM’s to those of the image noise and estimate the impact of noise on moment invariants.

Manuscript received January 18, 1983; revised October 3, 1983. This work was supported in part by the IBM Corporation and the U.S. Army Research Office.

The authors are with the Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125.

In Section II, CM’s are discussed and their relation to the circular harmonic expansion is established. The intimate relation between CM’s and moment invariants is established in Section III and used to relate moment invariants to the image function. In Section IV, information properties of moment invariants are investigated. This section accounts for which information is lost and which is emphasized when the image is represented by a few moment invariants. Redundancy in the information provided by various moment invariants is also derived. In Section V, the impact of noise is investigated. The framework for predicting statistical effects of additive noise in a given problem is given and the analysis is carried out for white noise. In the conclusion, the question *how good are moment invariants?* is addressed through the results of Sections IV and V.

## II. COMPLEX MOMENTS

In this section, we define the complex moments (CM’s) and derive several relationships that will be essential in evaluating moment invariants as a feature space for pattern recognition in the following sections. The results of Section II-B will be applied in Section IV, and the results of Section II-D will be applied in Section V.

### A. Definitions

We define the complex moment (CM) of order  $(p, q)$  for the two-dimensional function  $g(x, y)$  by the following equation:

$$C_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + iy)^p (x - iy)^q g(x, y) dx dy \quad (2.1)$$

where  $p, q$  are nonnegative integers and  $i = \sqrt{-1}$ . If  $g(x, y)$  is real nonnegative, as the gray level function of an image,  $C_{pp}$  is a real nonnegative number while  $C_{qp}$  is the complex conjugate of  $C_{pq}$ .

The CM of order  $(p, q)$  is a linear combination (with complex coefficients) of the ordinary moments  $M_{rs}$  satisfying

$$r + s = p + q. \quad (2.2)$$

Expansion of CM’s in terms of ordinary moments is used in [1] to derive moment invariants, but no closed form similar to (2.1) is given. In polar coordinates, the CM of order  $(p, q)$  can be written as

$$C_{pq} = \int_0^{2\pi} \int_0^{+\infty} r^{p+q} e^{i(p-q)\vartheta} G(r, \vartheta) r dr d\vartheta \quad (2.3)$$

where  $G(r, \vartheta) = g(r \cos \vartheta, r \sin \vartheta)$ . From this equation, the relation between complex moments and Zernike polynomials

[2], [12] becomes apparent. It can be shown that Zernike polynomials are obtained from the CM kernels  $r^{p+q} e^{i(p-q)\vartheta}$  by applying the Gram-Schmidt orthogonalization procedure [11] for all  $p, q$  having the same value of  $p - q$ . Another related approach for deriving moment invariants are radial and angular moments [6]. However, CM's are both very simple and quite powerful in providing an analytic characterization for moment invariants as will be shown presently. Equation (2.3) also inspires the following definitions. The *degree* of  $C_{pq}$  is given by  $p + q$  while its *repetition* is given by  $p - q$ .

**B. Rotation Properties**

If the image  $g_1(x, y)$  is rotated through a clockwise angle  $\Phi$  (or the image plane axes rotated through a counterclockwise angle  $\Phi$ ) to become  $g_2(x, y)$ ; more formally if  $g_2(x_2, y_2) = g_1(x_1, y_1)$ , where

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos \Phi & -\sin \Phi \\ \sin \Phi & \cos \Phi \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \tag{2.4}$$

then, from (2.3) the new set of CM's; namely  $\{C'_{pq}\}$ , is related to the old set  $\{C_{pq}\}$  by the relation

$$C'_{pq} = C_{pq} e^{-i(p-q)\Phi} \tag{2.5}$$

This equation is the basis for derivation of moment invariants from CM's. In the terminology of algebraic invariants [1], it says that CM's are *relative* invariants with respect to the transformation (2.4). If we cancel out the exponential factor in (2.5), we get *absolute* invariants. For instance  $|C'_{pq}| = |C_{pq}|$  or  $C'_{pq} C'_{qp} = C_{pq} C_{qp}$ .

Some basic properties of CM's can be derived by observing that (2.3) can be reduced to a simple form by using the *circular harmonic expansion* of  $G(r, \vartheta)$ . This expansion has been used in several areas as an analysis tool of two-dimensional data that possesses circular symmetry. It can be simply viewed as a Fourier series expansion of  $G(r, \vartheta)$  with respect to  $\vartheta$  at each fixed radius  $r$ . The expansion is given by

$$G(r, \vartheta) = \sum_{n=-\infty}^{+\infty} c_n(r) e^{in\vartheta} \tag{2.6}$$

The coefficient  $c_n(r)$  is a measure of the angular variation of frequency  $n$  cycles/cycle at radius  $r$ . Substituting (2.6) into (2.3), we get the following expression for CM's:

$$C_{pq} = 2\pi \int_0^{+\infty} r^{p+q+1} c_{q-p}(r) dr \tag{2.7}$$

Equation (2.7) shows that  $C_{pq}$  depends only on one coefficient of the circular harmonic expansion of  $G(r, \vartheta)$ ; namely the function  $c_{q-p}(r)$ . This fact determines specifically which features of the image contribute to the value of each CM. On the other hand, it also identifies the information loss encountered in representing the image by a finite number of CM's. This point is discussed further in Section III-C and will guide the discussion about the quality of the feature space formed by moment invariants in Section IV.

**C. Size Relations**

When several moment invariants are used as the coordinates of a feature space, the relative sizes of their values should be estimated. This estimate will determine the appropriate scale for each coordinate in the discriminant function and in space transformation procedures like the Karhunen-Loeve. Moreover, from a computational point of view, size differences between coordinates by orders of magnitude must be handled carefully to prevent rounding errors from degrading information content. We shall estimate the size of CM's since they are directly related to moment invariants.

Equation (2.7) helps estimate the size of  $C_{pq}$ 's by means of the Cauchy-Schwarz inequality. For this purpose, we take  $G(r, \vartheta)$  to vanish outside the unit disk; an assumption that can be met by scaling down any  $G(r, \vartheta)$  of compact support (bounded domain). In this case, the integration in (2.7) goes from 0 to 1 and the Cauchy-Schwarz inequality reduces to

$$|C_{pq}| \leq 2\pi \left( \frac{1}{2(p+q)+3} \int_0^1 |c_{q-p}(r)|^2 dr \right)^{1/2} \tag{2.8}$$

The Cauchy-Schwarz inequality is not particularly tight in this case because of the nature of the function  $r^{p+q+1}$ . This function becomes very small for small  $r$  as  $p + q$  grows. However, the inequality turns out to be a tight asymptotic estimate in the case of white noise as will be shown in Section II-D.

Another bound which reflects the mutual dependence of  $C_{pq}$ 's of the same degree is

$$|C_{pq}| \leq \int_0^{2\pi} \int_0^{+\infty} r^{p+q} G(r, \vartheta) r dr d\vartheta \tag{2.9}$$

This is true for nonnegative  $G(r, \vartheta)$ . Notice that the right-hand side (RHS) is identically  $C_{mm}$  for  $(p + q)$  even and  $m = (p + q)/2$ . Another relation involving CM's of different degrees for image functions restricted to the unit disk is

$$|C_{(n+m)n}| \leq C_{nn} \quad (m > 0) \tag{2.10}$$

Inequality (2.10) follows from (2.3) by observing that  $G(r, \vartheta)$  is nonnegative. These inequalities are verified in Table I which gives numerical values for the CM's of a typical image (which is an image often used as an example in image processing experiments) with  $p + q$  ranging from 0 to 6. Notice how fast these values decrease as  $p + q$  increases and, more severely, as we move off the main diagonal. If CM's are used as a feature space for pattern classification by the nearest neighbor (NN) rule [10], Euclidean distance would obviously be biased to the diagonal and low degree CM's. Notice the redundancy in this table due to the fact that  $C_{pq}$  is the complex conjugate of  $C_{qp}$ .

**D. Noise**

So far, we have considered CM's of a deterministic image  $g(x, y)$ . It is necessary in the noise analysis of moment invariants to study how CM's behave with a stochastic process of images  $\{n(x, y)\}$ . We shall take  $n(x, y)$  to be of zero mean since we can subtract out the deterministic function  $E[n(x, y)]$

TABLE I  
REAL AND IMAGINARY PARTS OF COMPLEX MOMENTS OF DEGREES  
0 THROUGH 6 FOR A SAMPLE IMAGE

q \ p	0	1	2	3	4	5	6
0	10.00185 .00000	.39759 -.15785	-.01626 .00456	-.00640 .21194	-.00084 -.05397	.03240 -.02711	-.02968 .06026
1	.39759 .15785	4.91894 .00000	.26040 -.01555	.03860 .09229	-.05970 .15477	-.01669 -.03399	
2	-.01626 -.00456	.26040 .01555	3.24030 .00000	.19404 .01519	.05067 .00586		
3	-.00640 -.21194	.03860 -.09229	.19404 -.01519	2.40598 .00000			
4	-.00084 .05397	-.05970 -.15477	.05067 -.00586				
5	.03240 .02711	-.01669 .03399					
6	-.02968 -.06026						

and treat it separately. It is immediately obvious in this case that

$$E(C_{pq}) = 0 \quad (2.11)$$

as follows from (2.1) and linearity of the expectation function.

From (2.11), we can write the following expression for the covariance of any two CM's of a process  $\{n(x, y)\}$  in terms of the autocorrelation function of the process

$$\begin{aligned} \text{cov}(C_{pq}, C_{rs}) &= E(C_{pq} C_{rs}^*) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \\ &\cdot (x + iy)^p (x - iy)^q (u - iv)^r (u + iv)^s \\ &\cdot R_{nn}(x, y, u, v) dx dy du dv \end{aligned} \quad (2.12)$$

where \* denotes the complex conjugate and  $R_{nn}(x, y, u, v)$  is the autocorrelation function of the process  $n$ , which is in general nonstationary, evaluated at the image plane points  $(x, y)$  and  $(u, v)$ .

For illustration, let  $\{n(x, y)\}$  be a white noise process with autocorrelation function

$$R_{nn}(x, y, u, v) = N\delta(x - u, y - v) \quad (2.13)$$

where  $\delta(x - u, y - v)$  is the two-dimensional Delta function and  $N$  is the two-dimensional spectral density of the process. Suppose also that we restrict ourselves to the unit disk. Substituting (2.13) into (2.12) and integrating over the unit disk yields the following expression for the covariance of CM's of a white noise process:

$$\text{cov}(C_{pq}, C_{rs}) = \frac{2\pi N}{p + q + r + s + 2} \delta_{(p-q)(r-s)} \quad (2.14)$$

where  $\delta_{(p-q)(r-s)}$  is the Kronecker delta function which is one when  $p - q = r - s$  and zero otherwise. In terms of the repetition of  $C_{pq}$  defined in Section II-A as  $p - q$ , it follows that in the case of white noise, only moments with the same repetition are correlated. Observe also that the RHS of (2.14)

is real and positive, which means that CM's which are correlated are statistically in phase. Notice that the experimental data of Table I reflect this correlation between CM's with the same repetition.

From (2.14), it follows that the variance of  $C_{pq}$  is given by

$$\text{var}(C_{pq}) = \frac{\pi N}{p + q + 1} \quad (2.15)$$

which shows that the variance of CM's of white noise goes like the inverse of their degrees. Variances of real and imaginary parts of  $C_{pq}$  are both equal to one-half of the RHS of (2.15).

Comparison of (2.15) and the square of (2.8) shows that the statistical behavior of white noise parallels the bound given by Cauchy-Schwarz inequality.

### III. MOMENT INVARIANTS

In this section, we shall give a brief account of moment invariants and relate them to complex moments defined in the previous section. This relation is used to define specifically the image features which affect the values of moment invariants and establish the basis for their evaluation as pattern recognition features in the following sections.

#### A. Review

Moment invariants are features of the image which are calculated in terms of ordinary moments and have the property that they retain their exact values (i.e., they are invariant) when the image is shifted, scaled, or rotated. Moment invariants were originally established from the relation between moments and the mathematically developed algebraic invariants [1]. Through this relation, moment invariants can be found up to any order.

Another equivalent set of invariants defined in terms of moments is the Zernike moment invariants [2] based on Zernike polynomials [12]. These polynomials obey an equation similar to (2.5), and thus they are also easier to derive than ordinary moment invariants.

The discrimination provided by moments as well as their vulnerability to noise are crucial factors to the design of moment-based pattern recognition systems. However, the analytic treatment of these factors has been given less attention in the literature than other practical considerations. The justification for using moment invariants has been loosely based on the uniqueness theorem of moments and some experimental results.

We shall concentrate here on rotational invariance. Other transformations of the image, like scale and shift, can be readily dealt with through the use of normalized central moments [8]. This amounts to centralizing and scaling the image such that  $C_{00}$  is a specified constant, and  $C_{01} = C_{10} = 0$ . The behavior of moment invariants upon reflection of the image is adequately treated in [1], [2] where "skew-invariants" are used.

*B. Relation to Complex Moments*

Upon rotation of the image plane axes through a counter-clockwise angle  $\Phi$ , the new set of CM's is related to the old set by relation (2.5) which we rewrite here for convenience:

$$C'_{pq} = C_{pq} e^{-i(p-q)\Phi}.$$

This means that rotation of the image affects the phase but not the magnitude of CM's. Hence the absolute values of CM's are indeed invariant under rotation and can be taken as our moment invariants. However, since  $C_{pq}$  is the complex conjugate of  $C_{qp}$ , we get only  $[n/2] + 1$  independent invariants out of  $n + 1$  CM's of degree  $n$ . This is a consequence of the loss of information encountered in taking the absolute value of a complex number.

Therefore, we should expect to get further invariant functions using the phase information of CM's. Equation (2.5) suggests that the relative phases of CM's, after we cancel out the exponential factor in (2.5), will provide invariant functions. For example,  $C_{21}^2 C_{24}$  is a complex number with invariant magnitude and phase under rotation.

Using this method, we can get as many as  $(N + 1)(N + 2)/2 - 1$  independent invariants out of the  $(N + 1)(N + 2)/2$  independent CM's of orders 0 through  $N$ . The difference of 1 is inevitable because we lose a degree of freedom by requiring rotation invariance. A specific method for getting this maximum number of independent invariants is described in [2]. Such invariants are by no means unique.

For the purpose of our analysis, we consider the so far most commonly used forms of invariants which can be written in terms of CM's as

$$(C_{pq} C_{qp}) \text{ or } (C_{rs}^k C_{tu}^k + C_{sr}^k C_{ut}^k) \tag{3.1}$$

where the relation between  $r, s, t, u, k$  is

$$(r - s) + k(t - u) = 0. \tag{3.2}$$

Notice that the expressions of (3.1) under condition (3.2) are designed to cancel out the exponential factor of (2.5) and also be real valued.

In conclusion, we see that moment invariants are simple algebraic combinations of CM's. Therefore, we shall now analyze

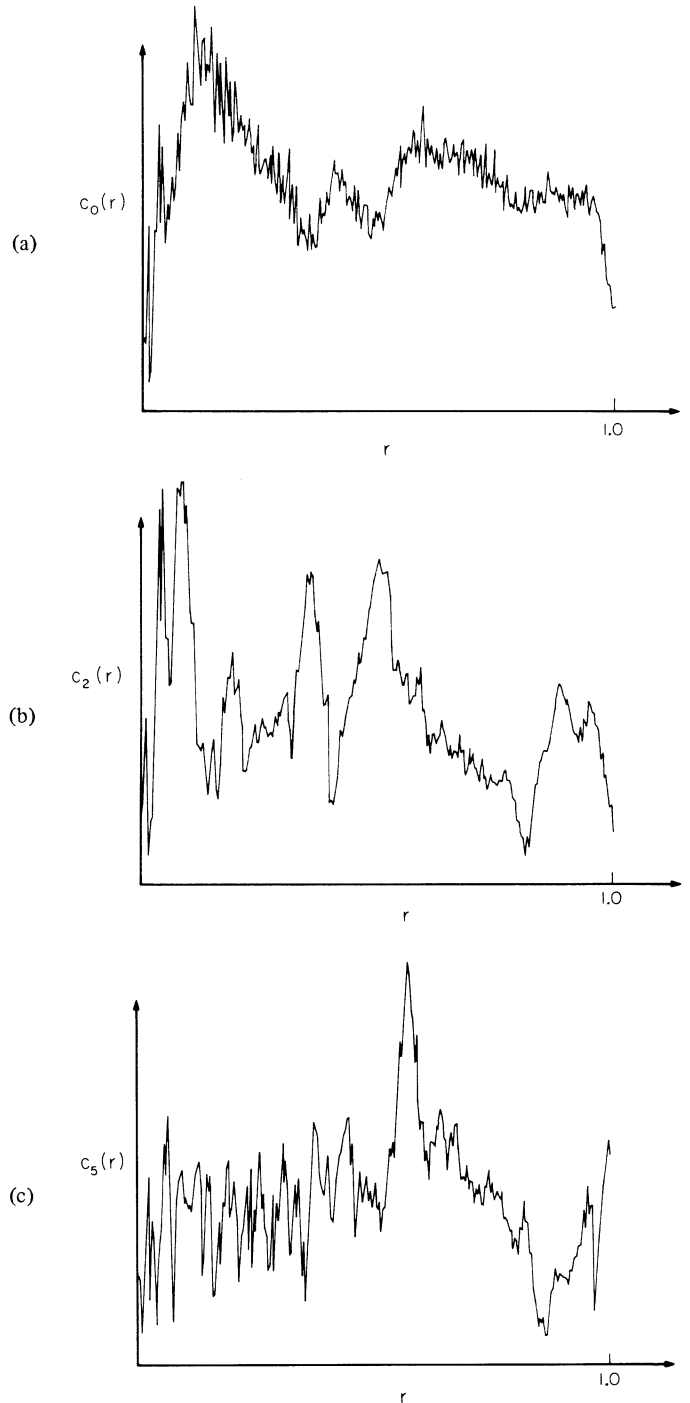


Fig. 1. (a) The zeroth circular harmonic coefficient function of a sample image (absolute value). (b) The second circular harmonic coefficient function of a sample image (absolute value). (c) The fifth circular harmonic coefficient function of a sample image (absolute value).

CM's in terms of which image features they are affected by, and these results carry on immediately to moment invariants.

*C. Image Features*

If moment invariants are to be used to discriminate between images, the vital question to ask is which features of the image contribute to their values. Once these features are determined, questions regarding the discrimination power and the robustness of moment invariants can be readily answered.

Equation (2.7) is the basis of this analysis. It defines  $C_{pq}$  in terms of the  $(q-p)$ th coefficient  $c_{q-p}(r)$  of the circular harmonic expansion  $\{c_n(r)\}$  of the image function  $G(r, \vartheta)$ . This shows that each CM depends on a single coefficient of this expansion and is independent of the rest. Circular harmonic expansion of images has been used in rotation-invariant pattern recognition [7] independent of its relation to moment invariants. Fig. 1(a)-(c) show three circular harmonic functions of an actual image.

Further inspection of (2.7) shows that even the circular harmonic coefficient contributing to the value of a certain CM does so in a special way. This follows from the well-known relation between moments of a function and its Fourier transform:

$$F(\omega) = \sum_{n=0}^{+\infty} (-i)^n M_n \frac{\omega^n}{n!} \quad (3.3)$$

where  $M_n$  is the  $n$ th ordinary moment of the function whose Fourier transform is  $F(\omega)$ . Applying (3.3) to the Fourier transform of the  $n$ th circular harmonic coefficient  $c_n(r)$  of the image  $g(x, y)$ , it follows from (2.7) that  $C_{pq}$  is affected, directly and solely, by the  $(p+q+1)$ th coefficient of the Taylor series expansion of the Fourier transform of  $c_{q-p}(r)$ . Hence, ordinary moment invariants as well as Zernike moment invariants can be now written as a finite expression with terms coming from these coefficients.

It is interesting to observe that although CM's form a complete basis for expanding the image function  $g(x, y)$  it seems at first glance that one cannot reconstruct the circular harmonic expansion of  $g(x, y)$  from them because they do not cover all the coefficients in (3.3). To see this, we observe that the Taylor series of the Fourier transform of  $c_n(r)$  could be reconstructed from CM's of the form  $C_{m(m+n)}$  only. Letting  $m$  go from 0 to  $\infty$ , we cover the Taylor coefficients  $M_{n+1}$ ,  $M_{n+3}$ ,  $M_{n+5}$ ,  $\dots$ . This leaves out all the  $M_i$ 's up to  $M_n$  as well as  $M_{n+2}$ ,  $M_{n+4}$ ,  $M_{n+6}$ ,  $\dots$  which are more than half the coefficients. However, since  $c_n(r)$  is an even function for  $n$  even and an odd function for  $n$  odd (because  $c_n(r)e^{in(\vartheta+\pi)}$  must equal  $c_n(-r)e^{in\vartheta}$ ), and by the limiting behavior of the Fourier transform at infinity, these coefficients can be determined.

#### IV. INFORMATION PROPERTIES

Feature selection for discrimination between images is a dimensionality reduction process which, for pattern recognition purposes, should retain a sufficient amount of information to discriminate between the patterns in question. The word "information" used here can be taken as the intuitive meaning of what we think of as information for discrimination between patterns. However, throughout this section, we shall use formulas which were actually based on Shannon's information and are directly related to the probability of error in pattern classification [4].

CM's play the central role in this section. On the one hand, they provide a very simple way to define moment invariants. On the other hand, the information properties are established through their direct relation to the circular harmonic expansion

of the image function. Since moment invariants are actually computed in terms of CM's, their information is a processed version of the information in CM's and therefore will not provide more discrimination information than CM's themselves. We shall be concerned mainly with ordinary moment invariants. The same approach used here can be adapted to evaluate other forms of moment invariants as well.

#### A. Information Loss

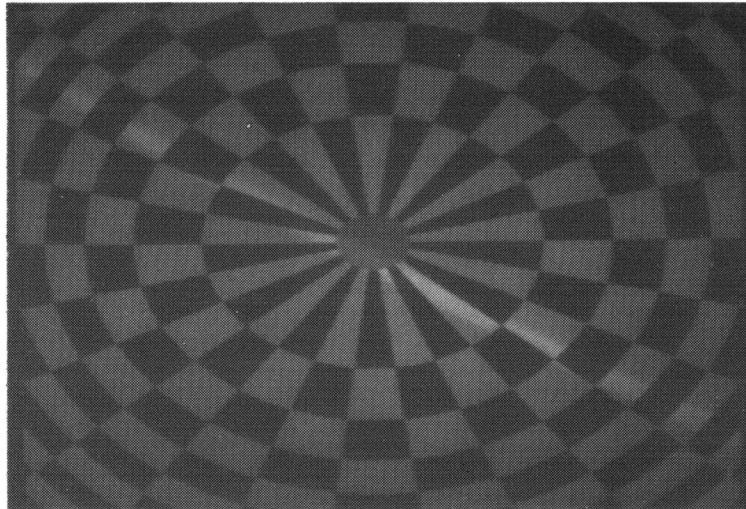
Since we are representing an entire image by a few numbers (of finite accuracy), it is inevitable that some information will be lost. We now consider *which* information is lost in the case of moment invariants and what is the impact of this loss on discrimination between patterns.

Our finite feature space includes moment invariants through a certain degree  $N$ . In fact, from the point of view of computation efficiency, once we go to degree  $N$ , we might as well compute *all* moment invariants through this degree since their components will be available. However, from the relation between moment invariants and the circular harmonic expansion of the image discussed in the previous section, such a feature space will depend only on the circular harmonic coefficients  $c_{-N}(r)$  through  $c_N(r)$  of the image function  $g(x, y)$ . Features of the image corresponding to angular variations of  $N+1$  cycles/cycle or more will be completely lost although the number of features considered is of the order  $(N+1)(N+2)/2$ . This loss is encountered when either Zernike moment invariants or ordinary moment invariants are used.

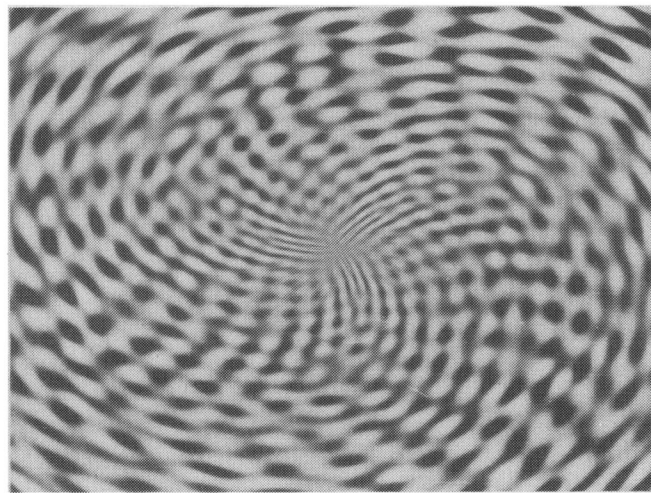
To illustrate the impact of this loss, Fig. 2(a) and (b) show two images (dc suppressed) for which *all* moment invariants through the tenth degree (65 features) are zeros (computed within a unit circle). The images are synthesized by setting to zero all circular harmonic coefficients which appear in (2.7) for  $p, q$  ranging from 0 to 10 and choosing all other coefficients arbitrarily to construct two images which are "different."

It is essential to observe that the zeroes assumed by moment invariants in this case are not *informative* zeros, i.e., they do not correspond to an accumulation of independent informative values which happen to cancel out and give the total of zero. Each pattern of angular variation  $N+1$  cycles/cycle or more is simply not sensed at all by any feature in the feature space used. Insensitivity to specific details of the pattern is *not* an inherent property of dimensionality reduction. As an example, we observe that normalized template matching is sensitive to every detail of the image although it reduces the whole image to a single number. Indeed, normalized template matching gives the value 1 if, and only if, every single pixel of the pattern checks.

The described loss will evidently degrade a general pattern recognition problem where the images have useful information in the high frequency part of their spectra. On the other hand, moment invariants' independence of certain angular variations can be used in a favorable way if the expected noise has a known distribution of angular variation frequencies. As an example, if the images of Fig. 2(a) and (b) are regarded as additive noise, moment invariants of degrees 0 through 10 will be optimal in the sense of noise independence.



(a)



(b)

Fig. 2. (a) An image (dc suppressed) whose moment invariants of orders 0 through 10 (65 features) are all zero. (b) Another image (dc suppressed) whose moment invariants of orders 0 through 10 (65 features) are all zero.

*B. Information Suppression*

We have already used (2.7) to show that even the circular harmonic coefficients which do affect certain CM's are not fully represented by a finite number of CM's. We now reach the same conclusion with more insight into the nature of information suppression encountered.

We substitute  $n = p - q$  and  $k = q$  in (2.7) and restrict the image to the unit disk to get the following relation:

$$C_{(n+k)k} = 2\pi \int_0^1 r^{n+2k+1} c_{-n}(r) dr. \tag{4.1}$$

We now fix  $n$  and study the information contained in each  $C_{(n+k)k}$  about  $c_{-n}(r)$  for all  $k \geq 0$ . The kernel of (4.1), namely  $r^{n+2k+1}$ , can be regarded as a weight which extracts features of  $c_{-n}(r)$  depending on where the feature occurs, i.e., depending on  $r$ . The factor of 1 in  $n + 2k + 1$  is fair enough; it gives a weight to  $c_{-n}(r)$ , proportional to  $r$  and we

know that the area of the image affected by  $c_{-n}(r)$  is also proportional to  $r$ . However, as  $n + 2k$  takes the values  $n, n + 2, n + 4, \dots$ , the values of  $c_{-n}(r)$  is *suppressed* for small  $r$  and the portion of  $c_{-n}(r)$  near  $r = 1$  dominates the value of  $C_{(n+k)k}$ .

For example, suppose that the patterns have useful discriminant features in their tenth circular harmonic coefficient. CM's which depend on this coefficient are  $C_{(0)(10)}, C_{(1)(11)}, \dots, C_{(n)(10+n)}$ . The maximum per unit weight given to the region  $r < 0.25$  in the image by any of these CM's is of the order  $10^{-7}$  and is likely to be lost in rounding errors.

Unfair emphasis of extracted features on portions of the image domain has been shown to have a negative impact on discrimination quality [4]. Roughly speaking, there is generally some *essential* discrimination information distributed over the image domain and when a certain portion is suppressed, its essential information is lost. For example, the informative variations in the curves in Fig. 1(a)-(c) are signifi-

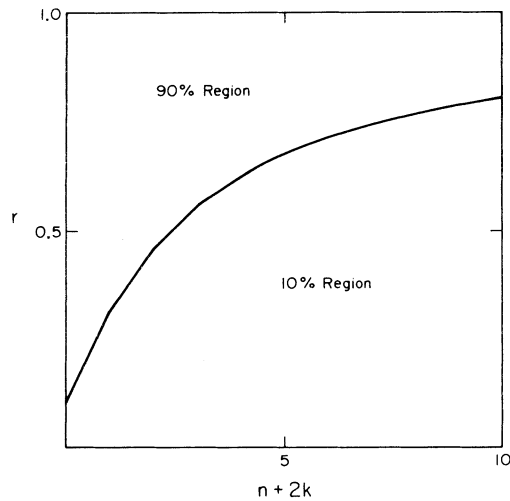


Fig. 3. Illustration of information suppression of moment invariants. The 90 percent region is given 90 percent of the kernel weight and the 10 percent region the remaining 10 percent, which is out of proportion to the region areas.

cant for the whole range of values of  $r$ , and suppressing the part near  $r = 0$  would cause a loss of information. However, in the case where most of the useful information of the image is near its boundary, this factor becomes less significant.

Fig. 3 demonstrates information suppression as follows. The abscissa is  $n + 2k$  varying from 0 to 10 and the ordinate is  $r$  going from 0 to 1. The curve now separates the upper part which is a range of  $r$  given 90 percent of the weight of the kernel (the outer portion of the image) and the lower part which takes the remaining 10 percent (the central portion of the image). It is clear that as the degree of the CM (and hence the moment invariant) goes up, it takes almost all the information from a very small part of the image which is near the image border and severely suppresses central features. We emphasize that the issue here is not that a specific extracted feature suppresses certain information, which is typical of almost any feature space; the problem is that *all* extracted features suppress the *same* information.

We conclude this discussion by noting that Zernike moment invariants do not suffer from radial information suppression as such. However, these invariants should not be obtained by combining *already* computed CM's. Instead, Zernike polynomials should be computed and used directly to extract the invariants from the image function. Otherwise, the finite register limitation will make their values completely dominated by rounding errors.

C. Information Redundancy

When all features suppress and emphasize the same parts, one is inclined to suspect that they are correlated or redundant. It has been shown that information-theoretic redundancy and correlation between kernels are directly related [4]. Considering a fixed circular harmonic coefficient  $c_n(r)$  in (2.7), the CM kernels defined in terms of this coefficient are those CM's of repetition  $-n$  and degrees  $n, n + 2, n + 4, \dots$ . Therefore, we shall consider the correlation between CM kernels having the same repetition but differing by two in

their degrees; namely CM's of the form  $C_{pq}$  and  $C_{(p+1)(q+1)}$ . From (2.7), this is equivalent to considering ordinary (radial) moments whose degrees differ by two. We start with the following simple identity:

$$r^{n+2} = Ar^n + r^n(r^2 - A) \quad \text{for any } A. \tag{4.2}$$

Now we look for the value of  $A$  which makes the two terms of the RHS of (4.2) orthogonal for  $r \in [0, 1]$ , i.e., the value of  $A$  satisfying the equation

$$\int_0^1 (Ar^n)(r^n(r^2 - A)) r dr = 0. \tag{4.3}$$

Excluding the trivial case  $A = 0$ , we get the following solution for (4.3):

$$A = \frac{(n + 1)}{(n + 2)}. \tag{4.4}$$

We have thus decomposed the kernel of the  $(n + 2)$ th moment into a linear combination of two components; one of which coincides with the kernel of the  $n$ th moment and the other orthogonal to it. Ideally, the orthogonal part ought to be dominant, since it is the part that contains new information. The ratio between the weights of the informative part and the redundant part, which we call  $\rho(n)$ , is given by

$$\rho(n) = \left( \frac{\int_0^1 [r^n(r^2 - A)]^2 r dr}{\int_0^1 (Ar^n)^2 r dr} \right)^{1/2} = \left( \frac{1}{(n + 1)(n + 3)} \right)^{1/2}. \tag{4.5}$$

Notice how this ratio becomes very small for large  $n$ . This means that CM's are indeed correlated and that the information in  $C_{(p+1)(q+1)}$  which was already provided by  $C_{pq}$  dominates the new information. In fact, the ratio of the new information to the redundant information goes to zero as  $p + q = n$  goes to infinity.

It can be shown that a recursive attempt to decorrelate CM kernels will yield Zernike polynomials. Therefore, redundancy properties developed here are avoided in Zernike moment invariants.

V. ROBUSTNESS

We now consider how CM's (and hence moment invariants) are affected by the presence of additive noise. We shall first give the general approach to be followed in order to estimate the noise impact in any given problem. We then carry out the analysis in the case of additive white noise and derive the important result that moment invariants of higher degrees are more vulnerable to white noise. This result puts a limitation on the degree of moment invariants that can be useful in the recognition procedure. A conflicting limitation is that only moment invariants of higher degrees carry the essential information of higher order circular harmonics. These two factors



together determine an optimal number of moment invariants in any given problem. Furthermore, the noise impact on  $C_{10}$  and  $C_{00}$ , which affects the values of the scale and shift, should also be taken into consideration since it affects the rest of the CM's through affecting the scaled central moments (or equivalently the normalized image [3]).

A. General Problem

Suppose that we have a given problem where the classes (template images) are  $g_1, g_2, \dots, g_k, \dots, g_n$ . For each class  $g_k$ , the nominal values of the  $(p, q)$  and  $(r, s)$  CM's namely  $C_{pq}(k)$  and  $C_{rs}(k)$ , are given by

$$C_{pq}(k) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + iy)^p (x - iy)^q g_k(x, y) dx dy$$

$$C_{rs}(k) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + iy)^r (x - iy)^s g_k(x, y) dx dy. \quad (5.1)$$

In the presence of a zero-mean independent additive noise process  $\{n(x, y)\}$  with autocorrelation function  $R_{nm}(x, y, u, v)$ , the covariance of  $C_{pq}$  and  $C_{rs}$  is given by (2.12). From (5.1) and (2.12), we can predict the second-order statistical behavior of CM's in the problem in question, and hence the behavior of the moment invariants' feature space.

For example, using (2.12) and (5.1), we can draw a two-dimensional scatter diagram [10]. The axes will be  $C_{pq}$  and  $C_{rs}$  and there will be  $n$  points corresponding to the  $n$  classes  $g_1, g_2, \dots, g_n$ . The coordinates of the point  $g_k$  are  $C_{pq}(k)$  and  $C_{rs}(k)$ . Around each point there is a covariance ellipse whose area, eccentricity and orientation are determined by  $\text{var}(C_{pq})$ ,  $\text{var}(C_{rs})$ , and  $\text{cov}(C_{pq}, C_{rs})$ .

If the *a priori* probabilities of the classes are known, the Gaussian approximation can be used to give an explicit estimate for the probability of misclassification. Higher order statistics can be used to obtain more accurate estimates. Other performance measures like between-class to within-class scatter can be readily computed. Hypothesis testing and other statistical methods can also be used. These are problem-oriented issues, but the tools are available through the analytically attractive definition (2.1) of CM's.

B. White Noise

As an illustration example for the analysis of noise impact on moment invariants, we shall carry out the analysis for a two-class problem with additive white noise.

Suppose that we have two classes (template images)  $g_1$  and  $g_2$ . We fix an integer  $N$  and consider all CM's having repetition  $N$ . Let us define the  $(-N)$ th circular harmonic coefficient of  $g_1$  and  $g_2$  to be  $a(r)$  and  $b(r)$ , respectively. For our purposes, we approximate the difference  $a(r) - b(r)$  by a (finite) polynomial  $P(r)$  for  $r \in [0, 1]$ . The CM's with repetition  $N$  have orders  $(N + n, n)$  with nonnegative integer  $n$ . According to (2.7), the difference in CM's of this form between class  $g_1$  and  $g_2$  is given by

$$\Delta C_{(N+n)n} = 2\pi \int_0^1 r^{N+2n+1} (a(r) - b(r)) dr. \quad (5.2)$$

Substituting the polynomial  $P(r)$  for  $(a(r) - b(r))$  in (5.2) and carrying out the integration for each power of  $r$ , we get a finite sum of the form

$$\Delta C_{(N+n)n} = \sum_{k=1}^K \frac{A_k}{N + 2n + M_k} \quad (5.3)$$

where  $A_k$ 's are constants and  $M_k$ 's,  $K$  are constant integers. On the other hand, according to (2.15), the standard deviation of  $C_{(N+n)n}$  due to additive white noise is given by

$$\sigma_{C_{(N+n)n}} = \frac{B}{(N + 2n + 1)^{1/2}} \quad (5.4)$$

where  $B$  is a constant depending on noise intensity. The last two equations show that while the difference in value between the CM of two classes goes like  $1/(p + q)$  (at most), the standard deviation due to white noise goes like  $1/\sqrt{p + q}$ . The ratio of the two, which can be taken as the signal-to-noise ratio, is proportional to  $1/\sqrt{p + q}$ .

This means that as the degree of the CM goes up, it becomes more vulnerable to white noise. The result is expected in view of the nature of the CM kernel  $r^{p+q} e^{i(p-q)\theta}$ . As  $p + q$  increases, both the signal and the noise components of  $C_{pq}$  decrease. However, since this kernel becomes negligible except near  $r = 1$ , the *cancellation* effect in the white noise due to its incoherence goes down and hence the noise part increases *relative* to the signal part which is coherent and does not have cancellations anyway.

VI. CONCLUSION

We have been concerned here with evaluating moment invariants as a feature space in a pattern recognition problem *avant projet*, i.e., predicting how they will behave given the classes and noise conditions of the problem. For this purpose, we accumulated enough background to help understand the nature of moment invariants and derived basic performance results in terms of the classes and noise.

We have also been concerned with the general question: *how good are moment invariants?* The reason for considering moment invariants as pattern recognition features in the first place is their invariance properties. Does this invariance justify their use for discrimination between patterns? When noise is present, as it *always* is, moment invariants will change in value and therefore will be no longer *invariant*. Is this change severe enough to overrule their initial invariance advantage?

We found that moment invariants are not *in general* good features. They suffer from information loss, suppression, and redundancy which limit their discrimination power. However, there are specific instances when these drawbacks do not essentially degrade the performance. For example, when the images in question do not have significant information in the higher order coefficients of their circular harmonic expansion, nothing is really lost, and when the central parts of the images have little useful detail, nothing is really suppressed. We also found that although Zernike moment invariants suffer from the same information loss problems of ordinary moment invariants, they do not have the information suppression or redundancy as such.



As far as noise is concerned, the situation is reversed. If the expected noise is mainly of high angular variation or mostly centralized, the situation is favorable. Using the second-order statistics of the noise process, it is possible to estimate a second-order approximation of the probability of misclassification. In the case of white noise, we concluded that higher order moment invariants were more vulnerable. In a problem where higher order circular harmonics are essential features of the patterns, this vulnerability puts a rigid limitation on how well moment invariants can provide the necessary discrimination information. This factor will determine if moment invariants are usable in the problem, or else some other feature space must be sought.

Based on the results of this paper, we suggest that the coefficient distribution of the circular harmonic expansion of interesting pattern recognition objects such as alphanumeric characters be investigated to determine which moment invariants, if any, are worth consideration as recognition features. Second-order statistics of commonly encountered noise processes can also be used to estimate the noise vulnerability of moment invariants in these cases. Derivation of Zernike-like variations of moment invariants to avoid the drawbacks described here is also an interesting ground for further work.

#### ACKNOWLEDGMENT

We should like to acknowledge S. Venkatesh for his assistance.

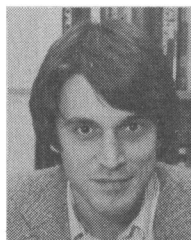
#### REFERENCES

- [1] M. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 179-187, Feb. 1962.
- [2] M. Teague, "Image analysis via the general theory of moments," *J. Opt. Soc. Amer.*, vol. 70, pp. 920-930, Aug. 1980.
- [3] Y. Abu-Mostafa and D. Psaltis, "Image normalization by complex moments," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, June 1983, pp. 114-120.
- [4] Y. Abu-Mostafa, "Information-theoretic characterization of linear feature spaces," to be published.
- [5] S. Dudani *et al.*, "Aircraft identification by moment invariants," *IEEE Trans. Comput.*, vol. C-26, pp. 39-45, Jan. 1977.
- [6] S. Reddi, "Radial and angular moment invariants for image identification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, pp. 240-242, Mar. 1981.
- [7] Y. Hsu *et al.*, "Rotation invariant digital pattern recognition using circular harmonic expansion," *Applied Opt.*, vol. 21, pp. 4012-4015, Nov. 1982.
- [8] S. Maitra, "Moment invariants," *Proc. IEEE*, vol. 67, pp. 697-699, Apr. 1979.
- [9] A. Cormack, "Representation of a function by its line integrals, with some radiological applications," *J. Appl. Phys.*, vol. 34, pp. 2722-2727, Sept. 1963.
- [10] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [11] J. Rice, *Matrix Computations and Mathematical Software*. New York: McGraw-Hill, 1981.
- [12] M. Born and E. Wolf, *Principles of Optics*. New York: Pergamon, 1975.
- [13] R. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1977.



**Yaser S. Abu-Mostafa** was born in Giza, Egypt, on March 3, 1957. He received the B.Sc. degree in electrical engineering from Cairo University in 1979, the M.S.E.E. degree from Georgia Institute of Technology in 1981, and the Ph.D. degree in electrical engineering and computer science from California Institute of Technology in 1983.

He was awarded the 1982-83 Milton and Francis Clauser Prize for the most original doctoral thesis at Caltech. He was appointed Assistant Professor of Electrical Engineering and Computer Science at Caltech in 1983. His current research interests include information theory, computational complexity, pattern recognition, and signal processing.



**Demetri Psaltis** received the B.Sc. degree in electrical engineering and economics in 1974, and the M.Sc. and Ph.D. degrees in electrical engineering in 1975 and 1977, respectively, all from Carnegie-Mellon University, Pittsburgh, PA.

After completing the Ph.D. dissertation he remained at Carnegie-Mellon as a Research Associate and later as a visiting Assistant Professor for three years. In 1980 he joined the faculty of the Department of Electrical Engineering at the California Institute of Technology, Pasadena, CA, where he is now Assistant Professor and a consultant to industry. His research interests are in the areas of optical information processing, image processing, pattern recognition, and optical devices. He has published over 80 technical papers in these areas. He is currently working on acousto-optic image processing architectures, adaptive optical processing methods, acousto-optic and photorefractive devices, and numerical optical computers.