Moment Matrices for Recognition of Spatial Pattern in Noisy Images

A. O. Hero, J. O'Neill and W.J. Williams

Abstract

We present a method for detection and classification of a spatial pattern in noise contaminated binary images which is based on performing subspace decomposition on a non-negative definite matrix of higher order moments of the image. We introduce a method which uses normalized power moments or ascending factorial moments as descriptors. While the set of p-th order factorial moments are in one-to-one correspondence with the set of p-th order power moments, the computation of factorial moments is much more numerically stable than the power moments. Indeed, using factorial moments we are able to implement pattern classifiers with over 30% more moment descriptors. We illustrate these techniques for word classification in binary document images.

I. INTRODUCTION

The problem of classification of patterns in noisy binary images has been a key component to many different areas including: automatic document processing, such as word spotting [1], character recognition [2], database retrieval; automatic target recognition (ATR); and astronomical cartography. We present a pattern classification method which is based on an underlying spatial Poisson point process model and uses higher order factorial moments of the intensity function as pattern discriminants. This extends our work on standard power moment methods presented at ICIP-96 [3]. Some justifications for moments are: 1) they provide a non-parametric pattern description; 2) combinations of moments have been identified with important invariances such as rotation, scale, and translation [4], [5]; 3) sample moments can usually be treated as jointly Gaussian random variables. With our introduction of non-negative definite moment matrices we provide another justification: they can be used to effectively separate signal pattern from noise background via noise subspace processing.

II. SPATIAL POINT PROCESS REPRESENTATIONS

Let $W = W(x, y)$ be a binary image indexed over $(x, y) \in \mathcal{A} = \{1, \ldots, n\} \times \{1, \ldots, m\}$. Define the coordinate locations (pixels) $\{(X_i, Y_i)\}_{i=1}^N$ as those pixels at which W is active, i.e., W not equal to zero. A general statistical model for $\{(X_i, Y_i)\}_{i=1}^N$ is a two-dimensional Bernoulli process specified by the probabilities $\rho(x, y) \in [0, 1]$ that pixel (x, y) is active. For a large number of pixels $(nm \gg N)$, the Bernoulli process is well approximated by a Poisson point process over the continuous rectangle $A = [1, n] \times [1, m]$. This Poisson point

process is completely described by the normalized intensity function $\rho(x, y)$ and the rate constant Λ

$$\lambda(x, y) = \Lambda \rho(x, y), \quad (x, y) \in A$$

where $\Lambda = E[N]$ is the average number of active pixels in the image. The statistical expectation of any function $g(X_i, Y_i)$ is

$$E[g(X_i, Y_i)] = \int_A g(x, y)\rho(x, y) \, dx dy.$$

III. SPATIAL MOMENTS

The spatial power moment (PM) $\mu_{X,Y}(k, l)$ of (integer) order k and l of positive integer random variables X, Y is defined as:

$$\mu_{X,Y}(k, l) = E[X^k Y^l].$$

Since the bivariate monomials form a basis for the space of all two-dimensional square integrable functions, the set of $2p$ power moments $\{\mu_{X,Y}(k, l)\}_{kl=1}^p$ completely characterizes the image as $p \to \infty$.

Fractional factorial moments come in two varieties: ascending and descending. In this work we use the spatial ascending fractional factorial moment (AFFM) $\alpha_{X,Y}(k, l, s)$ of (integer) order k and l and (real) fraction parameter s which is defined below

$$\alpha_{X,Y}^{kl,s} = E[[X]_s^k [Y]_s^l] \tag{1}$$

where we use the Pochammer symbol $[x]_s^k$ to denote the ascending fractional factorial

$$[x]_s^k = x(x + s) \cdots (x + s(k - 1)).$$

By convention we define $[x]_s^0 = 1$. When $s = 1$ and X is a positive integer $[x]_s^k = \frac{(x+k-1)!}{(x-1)!}$ is the standard ascending factorial moment [6].

In [7] we prove that the set of $2p$ PM's $\{\mu_{X,Y}(k, l)\}_{kl=1}^p$ and the set of $2p$ AFFM's $\{\alpha_{X,Y}(k, l, s)\}_{kl=1}^p$ are theoretically equivalent: any of the sets can be expressed mathematically in terms of the other set. However, as will be seen, with finite precision arithmetic, the factorial moment computation is more numerically stable.

A. Power Moment Matrix

We will find it useful to center and scale the variates X, Y to improve numerical stability of the moment matrix computations. Recall that in the present imaging application $X \in \{1, \ldots, m\}$ and $Y \in \{1, \ldots, n\}$. To improve the stability of the PM we replace X and Y by $\tilde{X} = \frac{x - (m+1)/2}{\sigma_p}$ and $\tilde{Y} = \frac{x - (n+1)/2}{\sigma_p}$ where $\sigma_p = (\max(m, n) - 1)/2$. Define the $(2p + 1) \times (2p + 1)$ power moment matrix:

$$\mathbf{P}^{2p} = \left(\left(\mu_{\tilde{X}, \tilde{Y}}(i + j) \right) \right)_{ij=0, \ldots, p} \tag{2}$$

B. Factorial Moment Matrix

Since ascending factorials moments are not defined for negative variates, construction of an appropriately normalized factorial moment matrix requires more care than for the power moment matrix. First of all we must scale the positive integer variates X and Y differently. Define the variates $\tilde{X} = \frac{x-(m+1)/2}{\sigma_f}$ and $\tilde{Y} = \frac{x-(n+1)/2}{\sigma_f}$ where $\sigma_f = \left[[(\max(m,n)-1)/2]_1^{2p} \right]^{1/(2p)}$. Next, for any real number $T = T^+ - T^-$, we define T^+ as the positive part and T^- as the absolute value of the negative part of T. Define the $(2p+1) \times (2p+1)$ factorial moment matrix $\mathbf{A}^{2p}$:

$$\mathbf{A}^{2p} = \left(\left(\alpha_{\tilde{X}+,\tilde{Y}+}(i+j,1/\sigma_f) \right. \right. \tag{3}$$
$$\left. \left. +(-1)^{i+j}\alpha_{\tilde{X}-,\tilde{Y}-}(i+j,1/\sigma_f) \right) \right)_{ij=0,\ldots,p} .$$

IV. DENOISING VIA EIGENDECOMPOSITION

Here we show that any binary noise with moments that are known or can be accurately estimated can be removed from an image via eigendecomposition of the observed moment matrices. The following important properties of the power moment matrix (2) and the factorial moment matrix (3) are proven in [7]. Both are block Hankel matrices which are symmetric non-negative definite. Both matrices have all entries in the interval $[-1,1]$ and are sparse for any intensity which is symmetric about the center point $\left(\frac{m+1}{2}, \frac{n+1}{2} \right)$ of the image (Fig. 1). Finally, by construction the $(1,1)$ entry of these two matrices is equal to 1.

Let a signal-independent binary noise having intensity $\lambda_n(x,y) = \Lambda_n \rho_n(x,y)$ be added (modulo 2) to the image. Denote the noiseless image (signal pattern) intensity by $\lambda_s(x,y) = \Lambda_s \rho_s(x,y)$. Then the overall image intensity $\lambda(x,y) = \Lambda \rho(x,y)$ will be the sum

$$\begin{aligned} \lambda(x,y) &= \lambda_s(x,y) + \lambda_n(x,y) \\ &= \Lambda\left[\beta\rho_s(x,y) + (1-\beta)\rho_n(x,y)\right] \end{aligned}$$

where $\beta = \Lambda_s/(\Lambda_s + \Lambda_n) \in [0,1]$ is a monotone function of signal-to-noise ratio Λ_s/Λ_n. The above representation implies that for any function $g(X,Y)$ of the coordinate process (X,Y) we have the decomposition:

$$\begin{aligned} E[g(X,Y)] &= \beta E[g(X,Y)|\text{signal alone}] \\ &\quad +(1-\beta)E[g(X,Y)|\text{noise alone}] \end{aligned}$$

Let $\mathbf{M}$, $\mathbf{M}_s$, and $\mathbf{M}_n$ denote the signal plus noise, signal alone, and noise alone moment matrices ($\mathbf{M}$ denotes either power moment or factorial moment matrix). We have the important result:

$$\mathbf{M} = \beta\mathbf{M}_s + (1-\beta)\mathbf{M}_n$$

Now let $\mathbf{K}$ be the known Cholesky factor of $\mathbf{M}_n$, i.e. $\mathbf{M}_n = \mathbf{K}\mathbf{K}^T$, and define $\tilde{\mathbf{M}} = \mathbf{K}^{-1}\mathbf{M}\mathbf{K}^{-T}$. Then we have:

$$\tilde{\mathbf{M}} = \beta\tilde{\mathbf{M}}_s + (1-\beta)\mathbf{I} \tag{4}$$

where $\mathbf{I}$ is the $(2p+1) \times (2p+1)$ identity matrix, and $\tilde{\mathbf{M}}_s = \mathbf{K}^{-1}\mathbf{M}\mathbf{K}^{-T}$ is the **whitened moment matrix** of the signal pattern.

Except for the presence of the signal-to-noise ratio dependent scalar β, (4) is the standard additive decomposition of a "measurement covariance" matrix $\tilde{\mathbf{M}}$ into unknown "signal covariance" $\tilde{\mathbf{M}}_s$ and "white noise covariance" $\mathbf{I}$. While the latter is full rank with $2p+1$ constant eigenvalues $\{1,\ldots,1\}$, the eigenvalues of the signal matrix $\tilde{\mathbf{M}}_s$ are non-negative and typically fall off rapidly to zero. This permits exact recovery of $\mathbf{M}_s$ via application of the eigendecomposition: $\tilde{\mathbf{M}} = \sum_{i=1}^{2p+1} \gamma_i \underline{\xi}_i \underline{\xi}_i^T$ where γ_i and $\underline{\xi}_i$ are eigenvalues (rank ordered) and eigenvectors.

To wit, since the eigenvectors of $\tilde{\mathbf{M}}$ and $\tilde{\mathbf{M}}_s$ are identical, from (4) we have

$$\tilde{\mathbf{M}} = \sum_{i=1}^{q} [\beta\gamma_i^s + (1-\beta)]\underline{\xi}_i\underline{\xi}_i^T + (1-\beta)\sum_{i=q+1}^{2p+1} \underline{\xi}_i\underline{\xi}_i^T$$

where $\{\gamma_i^s\}_{i=1}^q$ are the non-zero eigenvalues of $\tilde{\mathbf{M}}_s$. Thus only the q largest eigenvalues $\gamma_i = \beta\gamma_i^s + (1-\beta)$ of $\tilde{\mathbf{M}}$ are related to the signal pattern and the rest are pure noise eigenvalues $\{(1-\beta)\ldots,(1-\beta)\}$. In particular, at least if we know β a priori, then $\tilde{\mathbf{M}}_s$ can be exactly recovered from the eigendecomposition of $\tilde{\mathbf{M}}$ via

$$\tilde{\mathbf{M}}_s = \frac{1}{\beta}\sum_{i=1}^{q}[\gamma_i - (1-\beta)]\underline{\xi}_i\underline{\xi}_i^T .$$

However, (surprise!) it turns out that prior knowledge of β is not required. Indeed, since the $(1,1)$ element of $\mathbf{M}_s$ is equal to 1, and $\underline{e}_1^T\mathbf{K}\underline{\xi}_i = \underline{e}_1^T\underline{\xi}_i = \xi_{i1}$, we can determine β from the linear equation $\beta = \sum_{i=1}^q [\gamma_i - (1-\beta)]|\xi_{i1}|^2$:

$$\beta = \frac{\sum_{i=1}^q [\gamma_i - |\xi_{i1}|^2]}{1 - \sum_{i=1}^q |\xi_{i1}|^2} .$$

V. A WORDSPOTTING EXAMPLE

We generated postscript versions of two words "van" and "vax" in various font sizes, pitches, and font types. Varying levels of spatially homogeneous salt and pepper noise were added modulo-2 to the bitmaps of each word. Note that modulo-2 addition produces noise which is not strictly additive or linear. Raw moments of various mixed orders were computed empirically and sample power and sample factorial moments matrices were constructed using Matlab 4.0. Note that the number of pixels, or window size for each word depends on the number and width of letters in the word. To standardize the computation the bitmap coordinates for each word were scaled to a square of length 1 on a side. The Cholesky factor $\mathbf{C}$ of the spatially homogeneous noise moment matrix was applied to prewhiten the empirical word moment matrix. An eigendecomposition was performed on the prewhitened empirical moment matrix, with the signal subspace dimension determined by a threshold rule, and the original noiseless moment matrix was recovered by eliminating the noise subspace and renormalization, as discussed in the previous section of this paper.

A representative example of the noiseless and noise degraded bitmaps and raw power moment matrices is shown in Figs. 2 and 3 for "van" and "vax." Here SNR is equal to 0dB ($\beta = 0.5$ the number of random bit flips equal to

379

the number of active pixels for each word) using $L = 18$ moments (moment matrices of dimension 37×37). Note from Fig. 3 that while the noiseless raw moment matrices provide a small perceptible discrimination between words, addition of noise completely masks the word differences. Contrast this with the high discrimination power of the prewhitened moment domain evident from the clearly perceptible differences between "van" and "vax" encoded in the noiseless and recovered prewhitened matrices shown in Fig. 4.

We defined a simple discriminant based on computing the mean square distances between noisy moment matrix to the corresponding noiseless moment matrices for "van" or "vax." The discrimination was implemented using a minimum distance decision rule. Three classes of discriminants were compared: (D1) mean square distances between the raw empirical moment matrix M and the noiseless moment matrices M_s for the two words; (D2) mean square distances between the recovered prewhitened signal moment matrices $\tilde{M}_s$ and noiseless prewhitened moment matrices $C^{-T} M_s C^{-1}$ for the two words; (D3) the mean square distances between the recovered signal moment matrix $\hat{M}_s = C^T \tilde{M}_s C$ and the noiseless moment matrices M_s for the two words. In each case a mask was used to screen out elements of the moment matrices which were not substantially different from "van" to "vax."

The probability of decision error for each of discriminators rules is shown in Figs. 5 and 6 as a function of SNR. For unwhitened moment matrix discriminators D1 and D3 we compare probability of error using only lower order moments extracted from the noisy and recovered 37×37 moment matrices, respectively. The number used range from 3 moments $(\mu_{X,Y}(1,0), \mu_{X,Y}(1,1), \mu_{X,Y}(0,1))$, denoted by $L = 1$ in the figures, to 99 moments $(\mu_{X,Y}(i,j),\ i,j = 0,\ldots,9,\ i,j$ not simultaneously equal to zero), denoted by $L = 9$ in the figures. The performance of D1 is uniformly worse than that of D3 for all L values. Note that the use of more moments in the raw moment discriminant D1 actually degrades discrimination performance. This is consistent with the well known variance increase in estimation of higher order statistics as the order increases [8]. Interestingly, the opposite trend is observed in the whitened moment discriminant D3 where variance reduction has been acheived in the higher order moments via the subspace eigendecomposition of the raw 37×37 moment matrix. Finally, as expected, note that D2 attains very low probability of error by using minimum distance discrimination directly in the whitened moment matrix domain. We suspect that the reason that D3 is incapable of matching the excellent performance of D2 is due to poor condition number of the Cholesky factor C.

It was found that $L = 23$ could be used in the factorial moment matrix without running into run time errors due to numerical roundoff. This is to be compared to the upper limit of $L = 20$ encountered for the power moment matrix computations. Note that this 15% increase in L translates into an over 30% increase in the number of distinct mixed raw moments that can be used for discrimination ($23^2 - 1 = 528$ as compared to $20^2 - 1 = 399$).

REFERENCES

[1] F. Chen, L. Wilcox, and D. Bloomberg, "Word spotting in scanned images using hidden Markov modeling," in *Proc. IEEE Int. Conf.*
Acoust., Speech, and Sig. Proc., volume V, pp. 1-4, Minneapolis, MN, 1993.

[2] S. Kahan, T. Pavlidis, and H. Baird, "On the recognition of printed characters of any font and size," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 9, no. 2, pp. 274-288, Mar, 1987.

[3] J. O'Neill, A. O. Hero, and W. J. Williams, "Word spotting via spatial point processes," in *IEEE Int. Conf. on Image Processing*, volume 2, pp. 215-219, Laussane, Switzerland, Sept. 1996.

[4] M. Hu, "Pattern recognition by moment invariants," *Proc. of the Institute of Radio Engineers (IRE)*, p. 1428, 1961.

[5] Y. Li, "Reforming the theory of invariant moments for pattern recognition," *Pattern Recognition*, no. 7, pp. 723-730, 1992.

[6] N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate discrete distributions (2nd Ed)*, Wiley, New York, 1992.

[7] A. O. H. J. O'Neill and W. J. Williams, "Moment matrix methods for pattern classification," Technical Report 307, Comm. and Sig. Proc. Lab. (CSPL), Dept. EECS, University of Michigan, Ann Arbor, Mar., 1997.

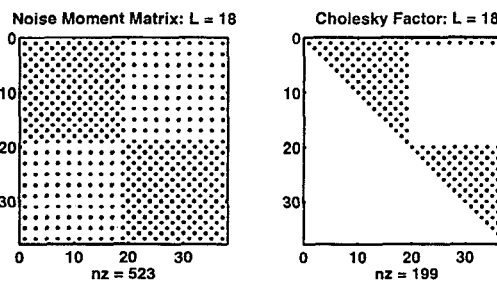[8] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Springer-Verlag, New York, 1981.

Fig. 1. The noise alone power and factorial moment matrices and their Cholesky factors have sparse structure which can be exploited to reduce roundoff error and explore structure of moment invariants.
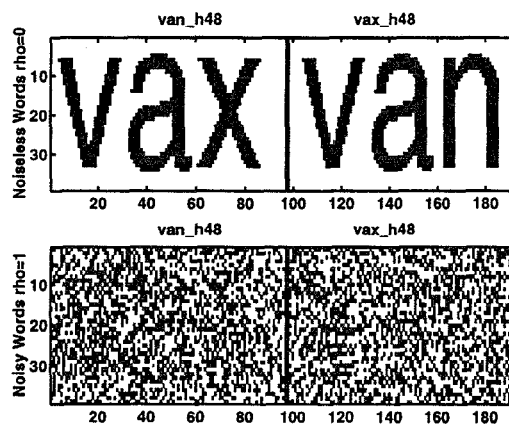

Fig. 2. Top: words *van* and *vax* in Helvetica 48 font. Bottom: same words corrupted by 0dB bit flip noise.
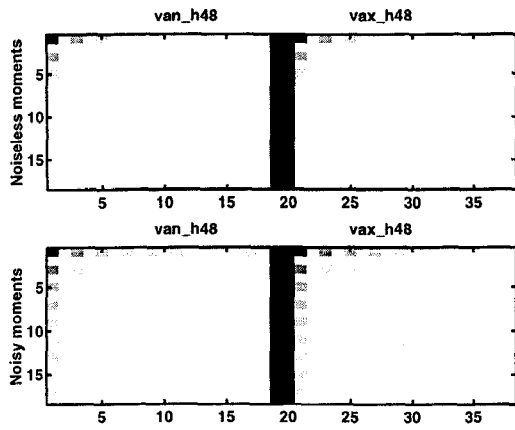
380

Fig. 3. Top: Power moment matrices for words *van* and *vax* in Helvetica 48 font for $L = 18$. Bottom: same moment matrices computed from words corrupted by 0dB bit flip noise. Note absence of strongly distinctive features between words even for noiseless case.

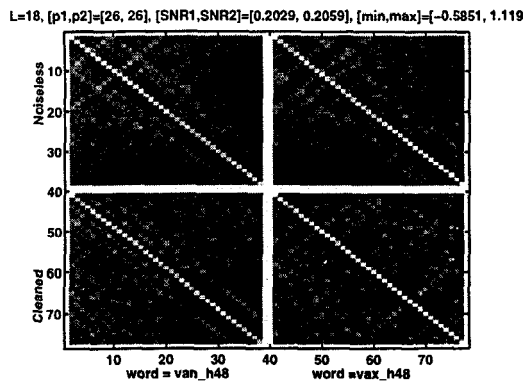L=18, [p1,p2]=[26, 26], [SNR1,SNR2]=[0.2029, 0.2059], [min,max]=[-0.5851, 1.119]

Fig. 4. Top: Noiseless whitened power moment matrices for words *van* and *vax*. Bottom: denoised empirical moment matrices from noise corrupted words shown in Fig 2. Note that distinctive features of noiseless whitened power moments are recovered after denoising.
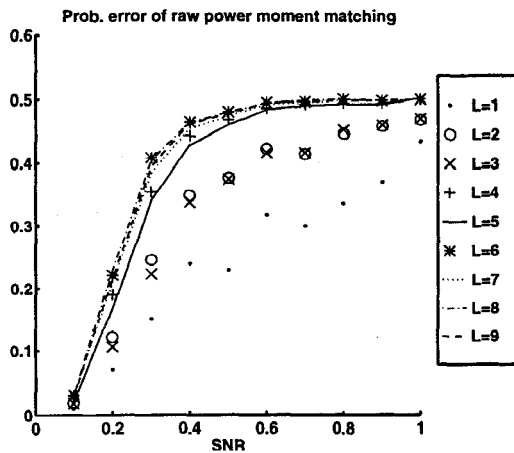
Prob. error of raw power moment matching

Fig. 5. Probability of error curves for D1 = minimum distance decision rule based on raw moments. Here we are matching different numbers of empirical moments to corresponding noiseless moments obtained from in Fig. 3. SNR is the relative number of bit flip errors as compared to active pixels in noiseless bitmap.
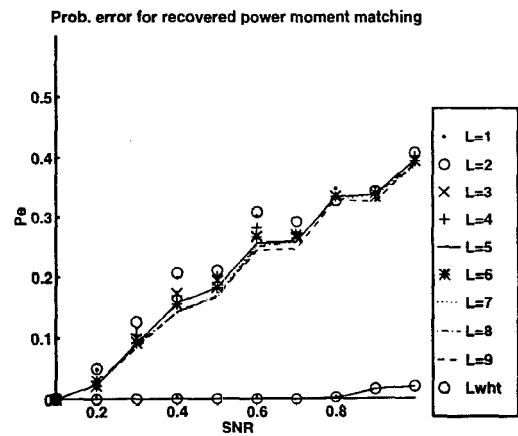
Prob. error for recovered power moment matching

Fig. 6. Bottom curve: probability of error for D2 = minimum distance decision rule based on moment matching in denoised prewhitened moment matrix domain. Here we are matching the empirical whitened moment matrices to noiseless whitened moment matrices shown in Fig. 4. Upper curves: same for D3 = minimum distance decision rule based on moment matching in recovered moment matrix domain.

381