1-1-1987

Maximum Entropy Reconstruction Of Moment Coded Images

Nicolaos S. Tzannes
University of Central Florida

Peter A. Jonnard

Maximum entropy reconstruction of moment-coded images

Nicolaos S. Tzannes
University of Central Florida
Department of Electrical Engineering
and Communication Sciences
Orlando, Florida 32816-0993

Peter A. Jonnard
Stanford University
Department of Electrical Engineering
Stanford, California 94305

Abstract. The maximum entropy principle (MEP) is applied to the problem of reconstructing an image from knowledge of a finite set of its moments. This new approach is compared to the existing method of moments approach and is shown to have a clear edge in performance in all of the applications attempted. Compression ratios more than twice as high as those previously achieved are possible with the new MEP method.

Subject terms: image reconstruction; maximum entropy; method of moments; moment coding.

Optical Engineering 26(10), 1077-1083 (October 1987).

CONTENTS

1. INTRODUCTION

The general problem of reconstructing data from a retained finite number of its moments is considered here from the point of view of the maximum entropy principle (MEP). The problem itself is quite old, but it has come to the forefront recently owing to its application to images, where moments can be calculated by using video-digital processors. Although the methods discussed here apply to both one- and two-dimensional data, we address mainly the problem of images.

The first important paper in the area is by Hu,[1] who used image moment invariants in two-dimensional pattern recognition problems. Later, Dudani et al.[2] used the approach in algorithms for automatic identification of aircraft. More recently, Teague[3] defined moment invariants with respect to Zernike polynomials and applied them to problems of image recognition, contrasting his results with the usual method of moments.

The basic notion behind all such problems is that a finite set of the image moments may contain enough information about the image to satisfy the problem at hand. It is, of course, known that all of the moments contain as much information about the image as the image itself since the coefficients of the power series of its characteristic function are uniquely related to them. The key question in each application is how small the set can be, i.e., how much the image data can be compressed by moment coding and still retain some key characteristic of the image pertinent to the application. In image transmission or storage, the moments can be used to compress the image in the same way as is done with various types of transform coding (Karhunen-Loeve, discrete cosine, slant, etc.). A few of them can be calculated and then transmitted or stored. Upon reception or retrieval, the moments can be used to reconstruct an approximation to the original image by an inverse procedure referred to as the method of moments. It is this inverse problem of reconstructing an approximation to an image by using a finite set of its moments that is of interest in the present work. More specifically, the work presented here proposes a new approach to reconstructing moment-compressed images that appears to have the edge over the method of moments in all of the cases we tried experimentally.

Section 2 briefly reviews the definition of the moments and the method of moments for the inverse problem. Section 3 discusses the proposed method of inversion by use of the

Paper 2339 received Sept. 22, 1986; revised manuscript received April 27, 1987; accepted for publication April 27, 1987; received by Managing Editor May 4, 1987. This paper is a revision of a paper presented at the SOUTHEASTCON '86 conference, March 1986, Richmond, Va.

MEP. Section 4 gives some examples to illustrate the superiority of the proposed MEP method.

2. METHOD OF MOMENTS

What follows in this section can be found in detail in Ref. 3. We review it here to set the stage for the proposed entropy method presented in Sec. 3.

Let us start by denoting a given image by the two-dimensional function $f(x,y)$, with x,y in bounded domains (the dimensions of the image). In defining the moments here, we assumed the image to be a continuous (or piecewise continuous) function of x, y and this is the reason for the integrals. In practical problems the image is spatially sampled and the integrals must be replaced by sums.

The zero-order moment of the image is given by

$$M(0,0) = \iint f(x,y)dxdy \ , \tag{1}$$

which represents the normalization factor of the image. The first-order moments are given by

$$M(1,0) = \iint xf(x,y)dxdy \ , \tag{2}$$

$$M(0,1) = \iint yf(x,y)dxdy \ , \tag{3}$$

which locate the centroid of $f(x, y)$. The third-order moments are given by

$$M(2,0) = \iint x^2f(x,y)dxdy \ , \tag{4}$$

$$M(1,1) = \iint xyf(x,y)dxdy \ , \tag{5}$$

$$M(0,2) = \iint y^2f(x,y)dxdy \ , \tag{6}$$

which characterize the size and orientation of the image. The general expression for the moments is given by

$$M(j,k) = \iint x^jy^kf(x,y)dxdy \ . \tag{7}$$

We note, however, that when we speak of having all of the moments up to the second moment, we actually mean that we have 3+2+1 numbers, and when we speak of having all of the moments up to the sixth one, we mean that we have 7+6+5+4+3+2+1 numbers, etc. Thus, if an $N \times N$ image is compressed by keeping six of its moments, the compression ratio would be $N \times N/28$ and not $N \times N/6$. In general, the compression ratio is given by $2N^2/(n+1)(n+2)$, where n is the number of moments retained.

Let us next consider the inverse problem of reconstructing an approximation of the image from a few of its moments by using the so-called method of moments. If the first two moments are known, the method of moments asserts that $f(x,y)$ can be reconstructed by the expression

$$g(x,y) = g(0,0) + g(1,0)x + g(0,1)y + g(2,0)x^2 + g(1,1)xy + g(0,2)y^2 \ , \tag{8}$$

where the coefficients $g(j,k)$ are picked so that the moments of the function $g(x, y)$ match those of $f(x, y)$. The general case of knowing the first N moments is an easy generalization of the above.

The method of moments is quite easy to apply, and it leads to a system of linear algebraic equations whose number is equal to the number of known moments as well as to the number of unknown coefficients.

Instead of the regular moments, one may use the central moments for either the compression or the decompression (reconstruction) of the image $f(x, y)$. In fact, one may view the problem in the realm of expansion theory and realize that although the set of functions $x^j y^k$ forms a complete set (Weierstrass' approximation theorem), it does not form an orthogonal one. To remedy this, Teague[3] reformulated the method of moments using Legendre polynomials. We briefly review this reformulation since it is to this approach that we compare the proposed method.

Since $f(x, y)$ is piecewise continuous, it can be expanded on Legendre polynomials as

$$f(x,y) = \sum\sum M'(m,n)P_m(x)P_n(y) \ , \tag{9}$$

where $P_m(x)$ are the known Legendre polynomials obeying

$$\int_{-1}^{+1} P_m(x)P_n(x)dx = \frac{2}{2m+1}\delta_{mn} \ , \tag{10}$$

i.e., they are orthogonal with the normalization factor as given by Eq. (10). The coefficients of the expansion $M'(m,n)$ are given by

$$M'(m,n) = \frac{(2m+1)(2n+1)}{4}\iint f(x,y)P_m(x)P_n(y)dxdy \ . \tag{11}$$

Teague calls them "orthogonal (Legendre) moments" and shows that they are related to the regular noncentral moments by simple expressions. They are often called "modified moments" since moments are scalars and the term "orthogonal" is usually reserved for vectors.

With this formulation, given a finite set of moments, one calculates the Legendre moments and then uses Eq. (9) to arrive at an approximation of the original $f(x,y)$, setting all the unknown Legendre moments equal to zero. Quite obviously, what Teague proposes is a Legendre polynomial version of the usual transform coding approaches, with the advantage that the coefficients are moments that are calculable by video processors. Furthermore, moments have symmetry properties that allow an efficient encoding of the image, which is important in data transmission and storage and in the development of automated pattern recognition schemes.

To demonstrate the usefulness of the approach, Teague applied it to coding and decoding letters of the alphabet. Some of his results are reviewed and compared to the proposed method in Sec. 4.

3. MAXIMUM ENTROPY PRINCIPLE RECONSTRUCTION

According to the maximum entropy principle proposed by Jaynes,[4] an unknown probability density function (PDF) $f(x)$ can be estimated by maximizing its entropy with respect to f(x), using any prior knowledge about the underlying random variable as constraints in the maximization. Phrased in terms of a two-dimensional PDF f(x,y), which is useful for the sequel, one estimates the unknown f(x,y) by minimizing the negentropy

$$H(x,y) = \iint f(x,y)\log[f(x,y)]dxdy \ , \tag{12}$$

subject to

$$K(i) = \iint f(x,y)h_i(x,y)dxdy \ , \tag{13}$$

$i = 0, 1, 2, \ldots, N$, which represents knowledge of the averages of the functions h(x,y) over the PDF f(x,y). When $i = 0$, $h(x,y) = 1$, so the constraint demanding that f(x,y) be integrated out to unity is always included in the formulation. The MEP was formulated for probability mass functions (PMFs), and it is somewhat controversial when applied to PDFs, but we ignore such problems here since the eventual application will be discrete. The above constraints are usually moments of the underlying random variable (X,Y), but they need not be; there is nothing in the original formulation of the principle that demands it.

The general solution to the above maximization (minimization) problem is easily obtained using Lagrangian multipliers for the constraints. It is

$$f(x,y) = \exp[-\lambda_0 - \lambda_1 h_1(x,y) - \lambda_2 h_2(x,y) - \ldots - \lambda_N h_N(x,y)] \ , \tag{14}$$

with the Lagrangian multipliers calculable by inserting the solution into the constraints and solving them simultaneously. The ensuing equations are not linear, and Newton-Raphson methods are used for these calculations. When the formulation involves a one-dimensional PDF f(x), all of the above double integrals become single ones, and the prior averages concern the one-dimensional h(x). And when the unknown probability function is a PMF, all of the integrals must be replaced by summation signs.

The MEP method has enjoyed success in PMF estimation as well as in the problem of spectral estimation; the key paper here is that of Burg.[5] In all of these applications, the success is attributed to the fact that it is intuitively sound (maximizing the entropy makes sense) and that it utilizes all of the values of known constraints while remaining noncommittal about unknown ones. The MEP also has been used in image processing, albeit for a different problem, that of enhancing an image in the presence of noise.[6,7] It is actually surprising that no one has yet used it in moment decoding of images, where it would seem to have its most natural setting, as is obvious from what follows.

To apply the MEP method to the reconstruction of moment-coded images, all that is needed is to assume that the image f(x,y) is a PDF of an underlying random variable (X,Y) whose nature is of no consequence to the problem. There is no mathematical difficulty with this assumption since images are everywhere positive and can be easily normalized to integrate to unity. In fact, there are even physical arguments for such an assumption,[8] but we do not concern ourselves with them here. Actually, any function that is everywhere positive can be thought of as a PDF, and thus if the setting is right, it is subject to the MEP approach, as one of the applications in Sec. 4 illustrates.

Let us see how one applies it to the problem of reconstructing an image whose first two regular moments have been retained, and let us contrast it to the solution to the method of moments reflected in Eq. (8). The MEP solution has the form

$$g(x,y) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 y - \lambda_3 x - \lambda_4 xy - \lambda_5 y) \ , \tag{15}$$

where g(x,y) is the reconstructed image and the λ are Lagrangian multipliers calculated by inserting the above into Eqs. (1) through (3) and solving them simultaneously for the constraints.

The MEP also can be applied to the orthogonal (Legendre) moment type of coding directly, without the need to calculate the regular moments. The coefficients $M'(m,n)$ given by Eq. (11) are viewed simply as averages of the functions $P_m(x)P_n(y)$ over the PDF f(x,y), and they are the constraints of the extremization. The solution in this case is given by

$$f(x,y) = \exp\left[-\sum\sum\lambda_{mn}P_m(x)P_n(y)\right] \ , \tag{16}$$

with the Lagrangian multipliers calculable from the constraints.

In closing, we point out that the MEP method does not lead to only a reconstruction whose moments are equal to the retained moments of the original image, as the method of moments does. Such reconstructions are many, and the method of moments solution is not unique unless all moments are retained. The MEP approach picks the solution that maximizes the entropy, and this solution is unique. Furthermore, the MEP solution utilizes all of the information known a priori, without assuming anything about what is not known, i.e., about the unretained moments (regular or Legendre), whereas the method of moments assumes them to be equal to zero.

4. APPLICATIONS AND EXAMPLES

The proposed MEP solution to the decoding of moment-coded data was applied to various cases of practical data, and the solution was compared to that obtained by using the "orthogonal" method of moments outlined in Sec. 2. In all cases the data were discrete. The reconstructions in the two-dimensional examples were quantized to two values (one and zero, shown in the figures as a dot and a blank, respectively) using the average value as a threshold to facilitate viewing. The mean squared errors, however, were computed using the actual values. The maximization problem was numerically solved by using simple Newton-Raphson algorithms, and the initial guesses for the Lagrangian multipliers were all zeroes. Since the original images were known, one could actually calculate initial guesses for the Lagrangian multipliers by solving Eq. (14) for the λ values, but no such effort was made. The solution converges without much difficulty since, as is known,[9] entropy is a convex function with only a single maximum.

4.1. One-dimensional letter E

As indicated in Sec. 3, the MEP method is also applicable to one-dimensional data, particularly if it is everywhere nonnegative (if not, a bias must be added whose effect must be removed at the outset). This first example pertains to a one-dimensional version of the letter E. The original letter E,

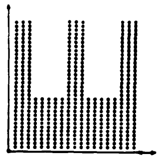

Fig. 1. One-dimensional letter E.

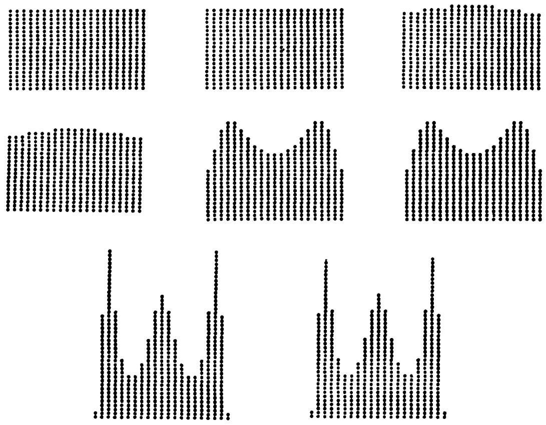

Fig. 2. One-dimensional MEP reconstruction of the letter E. First row (left to right): reconstructions using zeroth through second moments. Second row: third through fifth. Third row: sixth and seventh.

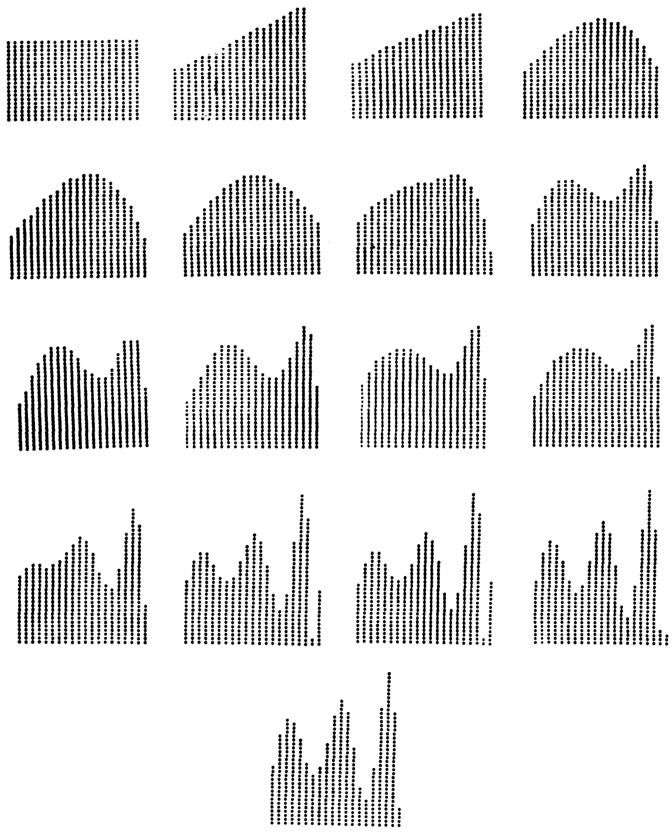

Fig. 3. One-dimensional MoM reconstruction of the letter E. First row (left to right): reconstructions using zeroth through third moments. Second row: fourth through seventh. Third row: eighth through eleventh. Fourth row: twelfth through fifteenth. Fifth row: sixteenth.

facing upward, is shown in Fig. 1. Moments of this letter were calculated and then used to reconstruct it by the two methods described in this paper. Figure 2 shows the evolution of the reconstruction of the letter E as each moment is utilized in the MEP approach, while Fig. 3 shows the same for the method of moments (MoM) approach. The MEP approach evolved to a reasonable approximation of the original letter with the use of six moments (a compression ratio of 21/6). The method of moments was nowhere near a reasonable facsimile of the letter at that point and required 16 moments before something resembling the letter E began to appear. Table I summarizes the mean squared errors (MSEs) of the evolutions of the two methods and illustrates the significant difference between the two approaches.

TABLE I. Mean squared errors for one-dimensional letters E and F.

Order of moments used	Letter E		Letter F	
	MoM	MEP	MoM	MEP
0	0.32	0.32	0.35	0.35
1	0.38	0.32	0.51	0.32
2	0.36	0.32	0.35	0.28
3	0.35	0.32	0.30	0.28
4	0.36	0.29	0.32	0.27
5	0.36	0.29	0.31	0.23
6	0.33	0.063	0.31	0.065
7	0.33	0.063	0.31	0.066
8	0.35	0.063	0.31	0.070
9	0.34	0.067	0.29	0.070
10	0.34	0.067	0.29	0.070
11	0.35	0.067	0.29	
12	0.26		0.23	
13	0.26			
14	0.27			
15	0.21			
16	0.22			

4.2. One-dimensional letter F

Example 2 is the same as example 1 except that it is for the letter F, shown in Fig. 4. The results are shown in Figs. 5 and 6 and in Table I. Although the beginnings of the letter appeared using six moments in the MEP approach, the results are obviously not as good as in the first example owing to the letter's lack of symmetry. The method of moments was unable to produce an F-looking figure even after 16 moments were used.

4.3. Two-dimensional letter E

The letter E is now viewed as a two-dimensional 21×21 pixel image of ones and zeroes, and the moments are spatial. The original E and the evolving reconstructions using the MEP method are shown in Figs. 7 and 8, respectively. The results using the MoM approach are shown in Fig. 9. Recall that when we talk about using seven moments we actually mean $8+7+6+\ldots+2+1 = 36$ numbers. The MEP approach produced a very acceptable reconstruction with the use of seven moments (a compression ratio of 441/36), whereas for the same number of moments the MoM approach resulted in the blob shown in Fig. 9. Even after 12 moments the MoM approach had not resulted in a reasonable looking E, as shown in the figure. It took 15 moments, i.e., 136 numbers

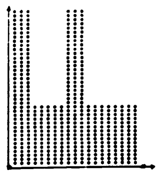

Fig. 4. One-dimensional letter F.

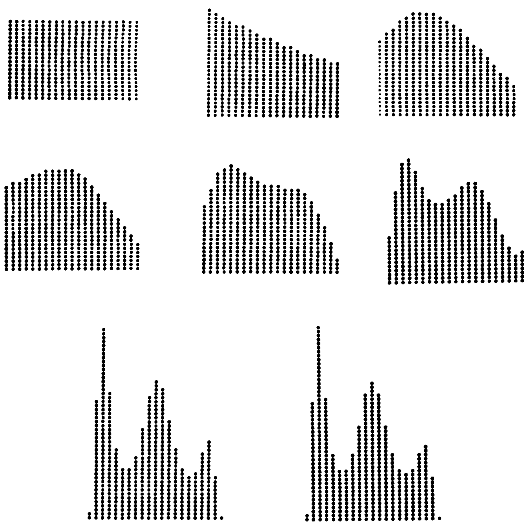

Fig. 5. One-dimensional MEP reconstruction of the letter F. First row (left to right): reconstructions using zeroth through second moments. Second row: third through fifth. Third row: sixth and seventh.

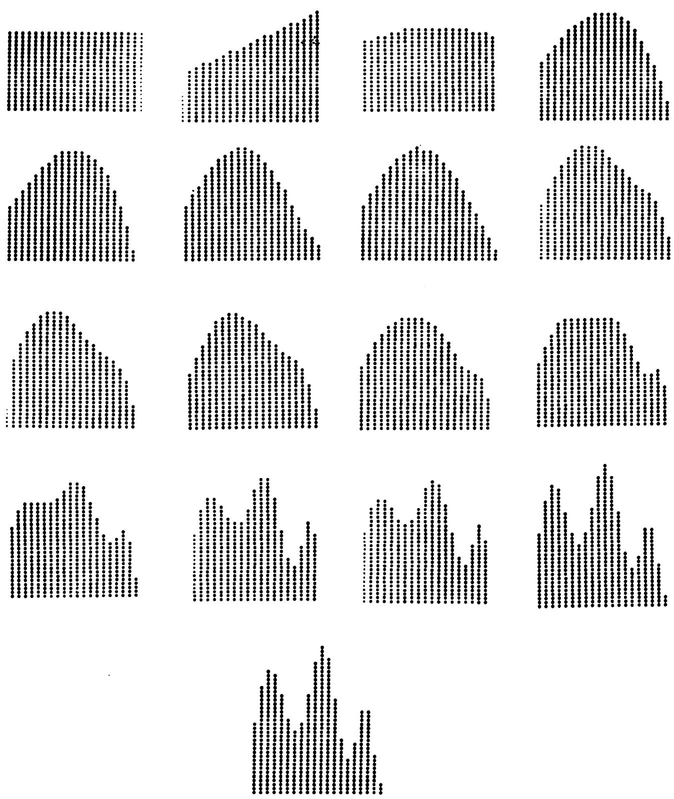

Fig. 6. One-dimensional MoM reconstruction of the letter F. First row (left to right): reconstructions using zeroth through third moments. Second row: fourth through seventh. Third row: eighth through eleventh. Fourth row: twelfth through fifteenth. Fifth row: sixteenth.

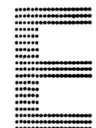

Fig. 7. Two-dimensional letter E (21×21 pixels).

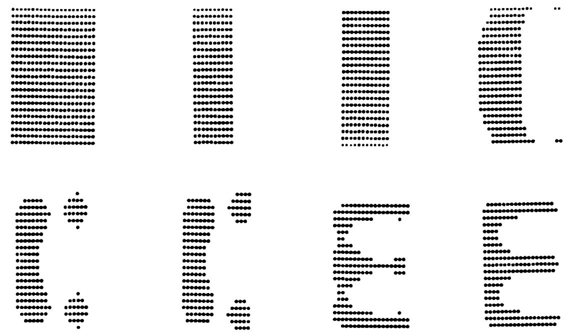

Fig. 8. Two-dimensional MEP reconstruction of the letter E. First row (left to right): reconstructions using zeroth through third moments. Second row: fourth through seventh.

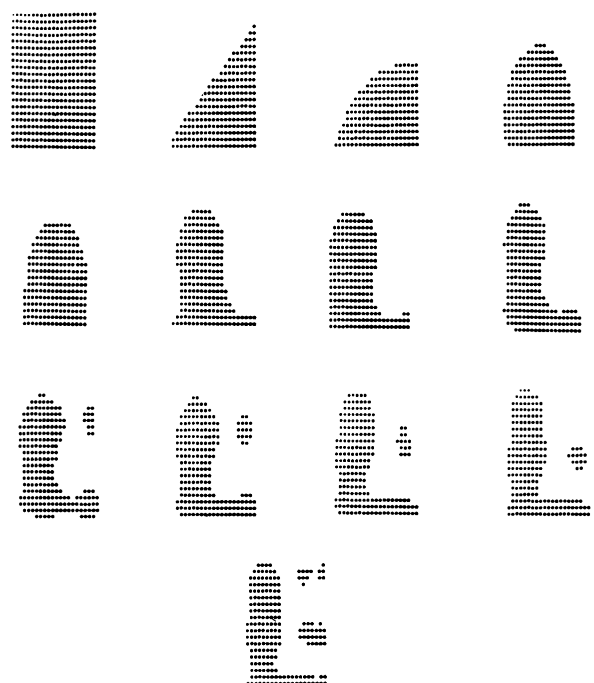

Fig. 9. Two-dimensional MoM reconstruction of the letter E. First row (left to right): reconstructions using zeroth through third moments. Second row: fourth through seventh. Third row: eighth through eleventh. Fourth row: twelfth.

(not shown), to produce the same quality E. The results check favorably with those reported by Teague.[3] The MSEs for both methods are given in Table II.

4.4. Two-dimensional letter F

Example 4 is the same as example 3 except that it is for the two-dimensional letter F (21×21 pixels), shown in Fig. 10. The results are shown in Figs. 11 and 12 and Table II, and the superiority of the proposed method is evident. Again, the

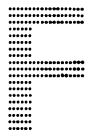

Fig. 10. Two-dimensional letter F (21×21 pixels).

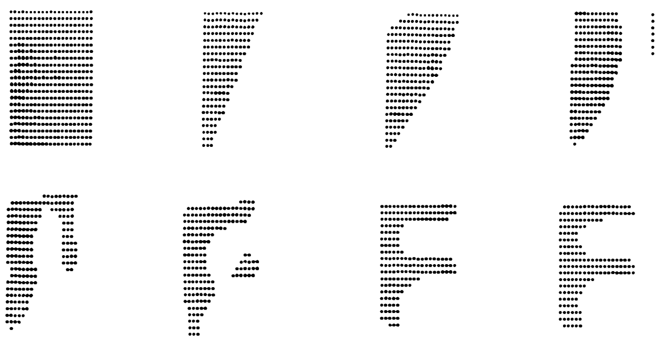

Fig. 11. Two-dimensional MEP reconstruction of the letter F. First row (left to right): reconstructions using zeroth through third moments. Second row: fourth through seventh.

TABLE II. Mean squared errors for two-dimensional letters E and F.

Order of moments used	Letter E		Letter F	
	MoM	MEP	MoM	MEP
0	0.91	0.91	1.3	1.3
1	1.10	0.84	1.6	1.1
2	0.94	0.82	1.4	0.97
3	0.94	0.76	1.1	0.90
4	0.93	0.62	1.2	0.71
5	0.90	0.62	1.0	0.62
6	0.88	0.43	1.0	0.40
7	0.75	0.33	0.91	0.37
8	0.71		0.90	
9	0.65		0.84	
10	0.57		0.77	
11	0.58		0.71	
12	0.61			

results obtained using the method of moments agree reasonably well with those reported by Teague.[3]

4.5. Two-dimensional letter O

This last example pertains to the letter O, shown in Fig. 13, which cannot be treated in a one-dimensional manner. The results are shown in Figs. 14 and 15. The MEP approach produced a perfect O using only the first four spatial moments, i.e., with a data compression ratio of 441/15, a truly remarkable result. The method of moments was used up through the twelfth moment, and still the result was quite unacceptable, even though it resembled an O. The MSEs of the two methods are given in Table III. It is interesting that the MEP method produced a perfect O with four moments even though its MSE was larger than that of the MoM with 12 moments, but one must recall that the MSE is based on actual values and not quantized ones.

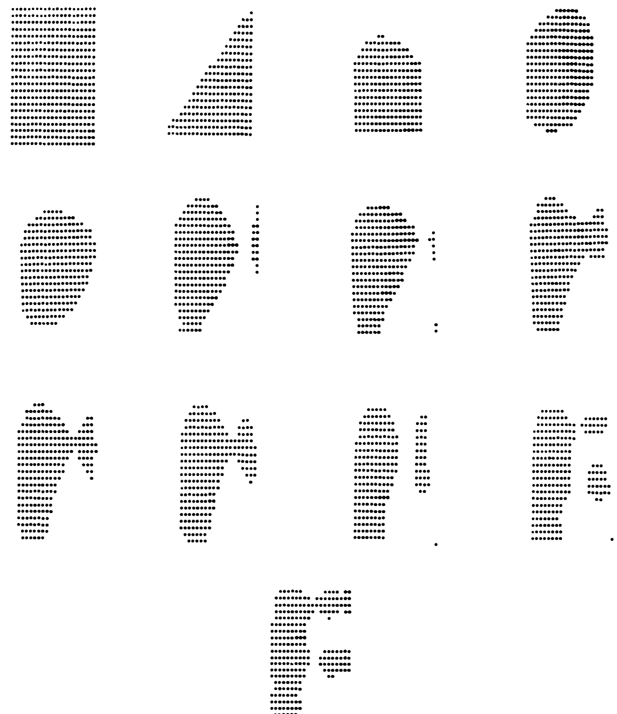

Fig. 12. Two-dimensional MoM reconstruction of the letter F. First row (left to right): reconstructions using zeroth through third moments. Second row: fourth through seventh. Third row: eighth through eleventh. Fourth row: twelfth.

5. CONCLUSION

On the basis of the examples reported here, it appears that the MEP approach is substantially better at reconstructing moment-coded data than the other available method, the method of moments. Its superiority was both visual and mathematical (MSE). The next step would be to try the method on photographic image data, and if it proves successful there as well for the same compression ratios reported here, then moment-coding could be established as a practicable alternative to all of the existing types of transform coding.

The MEP method requires more computations for the reconstruction, and its complexity increases rapidly with the inclusion of additional moments. The computer time needed for a reconstruction is dependent on the initial guesses for the values of the Lagrangian multipliers in the Newton-Raphson method and thus is difficult to compare with the method of moments. Its effectiveness, however, more than makes up for these disadvantages, particularly in applications of transmission of images over low rate channels, or in storage of images with a requirement of high compression ratios. Furthermore, research into efficient algorithms can alleviate this problem in two dimensions as it has already done in one-dimensional problems.

Before closing this work, we report that a recent paper by Kavehrad and Joseph[10] compared the method of moments to the MEP approach in the problem of estimating one-dimensional PDFs whose moments are known up to a finite number. Their conclusions on the performance of the two methods are much the same as reported here.

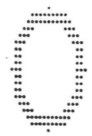

Fig. 13. Two-dimensional letter O (21×21 pixels).

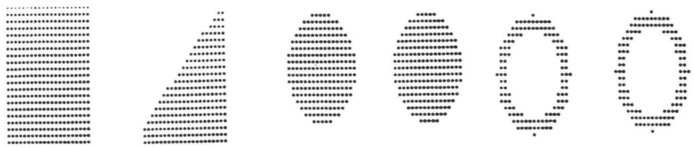

Fig. 14. Two-dimensional MEP reconstruction of the letter O. Left to right: reconstructions using zeroth through fifth moments.

Fig. 15. Two-dimensional MoM reconstruction of the letter O. First row (left to right): reconstructions using the zeroth through sixth moments. Second row: seventh through twelfth.

6. REFERENCES

1. M. K. Hu, "Visual pattern recognition by moment invariants," IRE Trans. Inf. Theory IT-8, 179–187 (1962).
2. S. A. Dudani, K. J. Breeding, and R. B. McGee, "Aircraft identification by moment invariants," IEEE Trans. Comput. C-26, 39–45 (1977).
3. M. R. Teague, "Image analysis via the general theory of moments," J. Opt. Soc. Am. 70(8), 920–930 (1980).
4. E. T. Jaynes, "Prior probabilities," IEEE Trans. Syst. Sci. Cybern. SSC-4, 227–241 (1968).
5. J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Stanford University, microfilms No. 75-25, 499 (1975).
6. B. R. Frieden, "Restoring with maximum entropy and maximum likelihood," J. Opt. Soc. Am. 62(4), 511–518 (1972).
7. A. J. Daniell and S. F. Gull, "Maximum entropy algorithm applied to image enhancement," IEE Proc. 127, Pt. F, No. 5, 170–172 (1980).

TABLE III. Mean squared errors for two-dimensional letter O.

Order of moments used	MoM	MEP
0	0.31	0.31
1	0.31	0.31
2	0.32	0.31
3	0.31	0.31
4	0.31	0.60
5	0.31	0.60
6	0.30	
7	0.27	
8	0.27	
9	0.25	
10	0.22	
11	0.23	
12	0.22	

8. S. F. Gull and J. Skilling, "The entropy of an image," in *Maximum-Entropy and Bayesian Methods in Inverse Problems*, C. R. Smith and W. T. Grandy, Jr., eds., pp. 287–301, D. Reidel Publ. Co., Hingham, Mass. (1985).
9. R. G. Gallager, *Information Theory and Reliable Communication*, p. 85, Wiley and Sons, New York (1968).
10. M. Kavehrad and M. Joseph, "Maximum entropy and the method of moments in evaluation of probability of error in digital communication systems," in *Proc. 20th Annual Conference on Information Sciences and Systems* (Princeton University, March 1986).

Nicolaos S. Tzannes is a professor in and chairman of the Department of Electrical Engineering and Communication Sciences at the University of Central Florida. He holds BSEE, MSEE, and Ph.D. degrees from the University of Minnesota, Syracuse University, and the Johns Hopkins University, respectively. Dr. Tzannes has taught at Johns Hopkins, Tufts University, the Universities of Virginia and Oklahoma, and the University of Patras, Greece. He is the author or editor of several books, the most recent being *Communication and Radar Systems* (Prentice-Hall, 1985). He also is the author or coauthor of more than 70 scientific papers on communications systems, digital signal processing, mathematical linguistics, etc., published in various journals or conference proceedings.

Peter A. Jonnard received his BSEE degree from the University of Virginia. He is presently a graduate student in the Department of Electrical Engineering at Stanford University.