between the front and reverse side images due to differences between both images caused by factors like document skews, different scales during image capture, and warped surfaces at books' spine areas. We are currently working on the development of a computer-assisted method to do the image overlay.

ACKNOWLEDGMENTS

This research was supported in part by the Agency for Science, Technology & Research and the Ministry of Education, Singapore, under research grant R252-000-071-303/112. The authors would like to thank the National Archives of Singapore for the use of their archival documents.

REFERENCES

[1] G. Nagy, "Twenty Years of Document Image Analysis in PAMI," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 1, pp. 38-62, Jan. 2000.

[2] R.G. Casey and E. Lecolinet, "A Survey of Methods and Strategies in Character Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 7, pp. 690-706, July 1996.

[3] H. Negishi, J. Kato, H. Hase, and T. Watanabe, "Character Extraction from Noisy Background for an Automatic Reference System," *Proc. Fifth Int'l Conf. Document Analysis and Recognition,* pp.143-146, Sept. 1999.

[4] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. System, Man, and Cybernetics,* vol. 9, no. 1, pp. 62-66, 1979.

[5] Y. Liu, S.N. Srihari, "Document Image Binarization Based on Texture Features," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 5, pp. 540-544, May 1997.

[6] S. Liang and M. Ahmadi, "A Morphological Approach to Text String Extraction from Regular Periodic Overlapping Text/Background Images," *Graphical Models and Image Processing,* vol. 56, no. 5, pp. 402-413, Sept. 1994.

[7] J.M. White and G.D. Rohrer, "Image Thresholding for Optical Character Recognition and Other Applications Requiring Character Image Extraction," *IBM J. Research Development,* vol. 27, no. 4, pp. 400-410, 1983.

[8] H.-S. Don, "A Noise Attribute Thresholding Method for Document Image Binarization," *Proc. Third Int'l Conf. Document Analysis and Recognition,* pp. 231-234, Aug. 1995.

[9] G. Sharma, "Cancellation of Show-through in Duplex Scanning," *Proc. Int'l Conf. Image Processing,* vol. 3, pp. 609-612, Sept. 2000.

[10] G. Sharma, "Show-through Cancellation in Scans of Duplex Printed Documents," *IEEE Trans. Image Processing,* vol. 10, no. 5, pp. 736-754, May 2001.

[11] D.L. Donoho, "Threshold Selection for Wavelet Shrinkage of Noisy Data," *Proc. 16th Ann. Int'l Conf. IEEE Eng. in Medicine and Biology Soc.,* vol. 1, pp. A24-A25, Nov. 1994.

[12] D.L. Donoho, "De-Noising by Soft-Thresholding," *IEEE Trans. Information Theory,* vol. 41, no. 3, pp. 613-627, May 1995.

[13] K. Berkner, M.J. Gormish, E.L. Schwarts, and M. Boliek, "A New Wavelet-Based Approach to Sharpening and Smoothing of Images in Besov Spaces with Applications to Deblurring," *Proc. Int'l Conf. Image Processing,* vol. 3, pp. 797-800, Sept. 2000.

[14] J. Lu, D.M. Healy, and J.B. Weaver, "Contrast Enhancement of Medical Images Using Multiscale Edge Representation," *Optical Eng.,* vol. 33, no. 7, pp. 2151-2161, July 1994.

[15] J. Lu, "Image Deblocking via Multiscale Edge Processing," *Proc. SPIE Wavelet Application in Signal and Image Processing IV,* M.A. Unser, A. Aldroubi, and A.F. Laine, eds., vol. 2825, part 2, pp. 742-751, Aug. 1996.

[16] S. Mallat and S. Zhong, "Characterization of Signals from Multiscale Edges," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 7, pp. 710-732, July 1992.

[17] W.L. Hwang, F. Chang, "Character Extraction from Documents Using Wavelet Maxima," *Proc. SPIE: Wavelet Applications in Signal and Image Processing IV,* vol. 2825, part 2, M.A. Unser, A. Aldroubi, and A.F. Laine, chairs/eds., pp.1003-1015, Aug. 1996.

[18] K. Etemad, D. Doerman, and R. Chellappa, "Multiscale Segmentation of Unstructured Document Pages Using Soft Decision Integration," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 1, pp. 92-96, Jan. 1997.

[19] R. Cao, C.L. Tan, Q. Wang, and P. Shen, "Segmentation and Analysis of Double-Sided Handwritten Archival Documents," *Proc. Fourth IAPR Int'l Workshop Document Analysis Systems,* pp. 147-158, Dec. 2000.

[20] C.L. Tan, R. Cao, P. Shen, J. Chee, and J. Chang, "Removal of Interfering Strokes in Double-Sided Document Images," *Proc. Fifth IEEE Workshop Applications of Computer Vision,* pp. 16-21, Dec. 2000.

[21] W. Niblack, *An Introduction to Digital Image Processing,* pp. 115-116, Englewood Cliffs, N.J.: Prentice Hall, 1986.

[22] L. Feng, Y.Y. Tang, and L.H. Yang, "A Wavelet Approach to Extracting Contours of Document Images," *Proc. Fifth Int'l Conf. Document Analysis and Recognition,* pp. 71-74, Sept. 1999.

[23] M. Junker, R. Hoch, and A. Dengel, "On the Evaluation of Document Analysis Components by Recall, Precision, and Accuracy," *Proc. Fifth Int'l Conf. Document Analysis and Recognition,* pp. 713-716, Sept. 1999.

Estimating the Intrinsic Dimension of Data with a Fractal-Based Method

Francesco Camastra and Alessandro Vinciarelli

Abstract—In this paper, the problem of estimating the intrinsic dimension of a data set is investigated. A fractal-based approach using the Grassberger-Procaccia algorithm is proposed. Since the Grassberger-Procaccia algorithm performs badly on sets of high dimensionality, an empirical procedure that improves the original algorithm has been developed. The procedure has been tested on data sets of known dimensionality and on time series of Santa Fe competition.

Index Terms—Bayesian information criterion, correlation integral, Grassberger-Procaccia's algorithm, intrinsic dimension, nonlinear principal component analysis, box-counting dimension, fractal dimension, Kolmogorov capacity.

1 INTRODUCTION

PATTERN recognition problems involve data represented as vectors of dimension d. The data is then embedded in the space $\mathbb{R}^d$, but this does not necessarily mean that its intrinsic dimension (ID) is d. The ID of a data set is the minimum number of free variables needed to represent the data without information loss. In more general terms, following Fukunaga [9], a data set $\Omega \subset \mathbb{R}^d$ is said to have an ID equal to M if its elements lie entirely within an M-dimensional subspace of $\mathbb{R}^d$ (where $M < d$).

Estimation of the ID is important for many reasons. The use of more dimensions than strictly necessary leads to several problems. The first one is the space needed to store the data. As the amount of available information increases, the compression for storage purposes becomes even more important. The speed of algorithms using the data depends on the dimension of the vectors, so a reduction of the dimension can result in reduced computational time. Moreover, in the statistical learning theory approach [32], the capacity and the generalization capability of the classifiers depend on ID and the use of vectors with smaller dimension often leads to improved classification performance. Finally, when using an autoassociative neural network [18] to perform a nonlinear feature extraction (e.g., nonlinear principal component analysis), the ID can suggest a reasonable value for the number of hidden neurons.

This paper presents an approach to ID estimation based on fractal techniques. Fractal techniques have been successfully applied to estimate the attractor dimension of underlying dynamic systems generating time series [17]. The literature presents results in the study of chaotic systems (e.g., Hénon map, Rössler oscillator) [22], in the analysis of ecological time series (e.g. Canadian lynx population) [15], in biomedical signal analysis [31], in radar clutter identification [12], and in the prediction of financial time series [24]. Nevertheless, in pattern recognition, fractal methods are mainly used to measure the fractal dimension of an image [13]. As far as we know, the application of fractal approaches to the problem of ID estimation has never been proposed before. The proposed ID estimation method is tested on both artificial and real data showing good results.

The paper is organized as follows: Section 2 presents several techniques for estimating ID. In Section 3, fractal methods are reviewed. The procedure to estimate ID is described in Section 4. In Section 5, some experimental results are reported and, in Section 6, some conclusions are drawn.

- *F. Camastra is with the INFM-DISI, University of Genova, Via Dodecaneso 35, 16146 Genova, Italy. E-mail: camastra@disi.unige.it.*
- *A. Vinciarelli is with the Institut Dalle Molle d'Intelligence Artificielle Perceptive, Rue du Simplon 4, 1920 Martigny, Switzerland. E-mail: vincia@idiap.ch.*

Manuscript received 18 July 2001; revised 14 Jan. 2002; accepted 2 Apr. 2002. Recommended for acceptance by P. Meer.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 114549.

0162-8828/02/$17.00 © 2002 IEEE

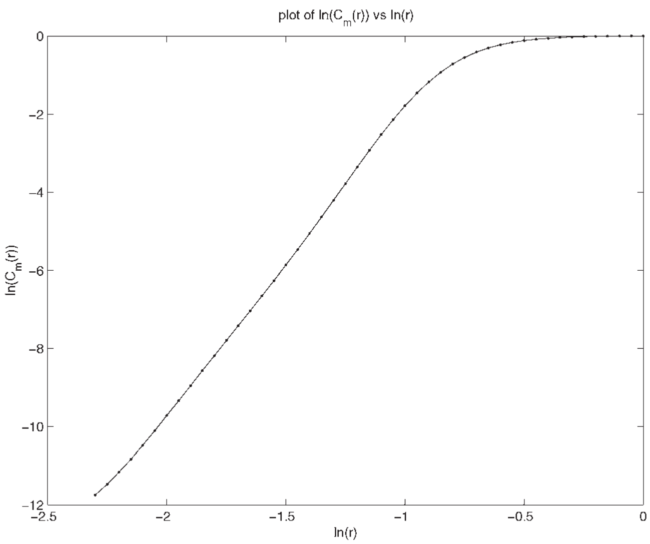

Fig. 1. Plot of $\ln(C_m(r))$ versus $\ln(r)$.

2 ID Estimation Techniques

Following the classification proposed in [16], there are two approaches for estimating ID. The first (*global*) estimates the dimensionality of the data set as a whole. The second (*local*) estimates ID using the information contained in sample neighborhoods, avoiding the projection of the data onto a lower-dimensional space.

The literature presents only a few results, mostly based on *projection techniques* and topological methods (*topological dimension of data* [3]). Projection techniques, which are generally used as global methods, search for the best subspace to project data by minimizing the reconstruction error. These methods can be divided into two families: linear and nonlinear. Linear methods such as *Principal Component Analysis* (PCA) are inadequate estimators since they tend to overestimate the ID [1]. For example, a set formed by points lying on a curve for PCA has dimension 2 rather than 1. On the other hand, though nonlinear PCA [18] performs better than linear PCA in some contexts [7], it presents problems when estimating ID [19]. Among projection techniques it is worth mentioning the *Whitney reduction network* recently proposed by Broomhead and Kirby [2], [18]. This method is based on Whitney's concept of *good projection*, namely, a projection obtained by means of an injective mapping. As pointed out in [18], finding good projections can be difficult and can sometimes involve empirical considerations. Finally, the fractal approach presented in this paper can be classified as a global method.

Topological methods that are local try to estimate the *topological dimension* of the data manifold. Some authors [8] use Topology Representing Networks (TRN) [21], that optimally represent the data topology to estimate ID.

The number n of cross-correlations is used as an indicator of the local dimension of Ω, assuming that n is close to the number k of spheres which touch a given sphere, in the *Sphere Packing Problem* [4]. This approach presents some drawbacks: the above assumption has not been proven yet, the number k is known exactly only for dimension values between 1 and 8, and k tends to grow exponentially with the space dimension. This last peculiarity strongly limits the use of the conjecture in practical applications where data can have high dimensionality. Other authors [10] perform a *Voronoi tesselation* of data space and, in each Voronoi set, a local PCA is performed. Finally, Bruske and Sommer [3] after the Voronoi tesselation of the data space, compute, by means of TRN, an optimal topology preserving map G. Then, for each node $i \in G$, a PCA is performed on the set Q_i consisting of the differences between the node i and all of its m_i neighbors in G. The above mentioned [3], [10] methods present some limits: Since none of the eigenvalues of the covariance matrix will be null due to noise, it is necessary to use heuristic thresholds in order to decide whether an eigenvalue is significant or not.

3 Fractal Methods

A unique definition of the dimension has not been given yet. A popular definition, among many proposed, is the so-called *box-counting* dimension [20], belonging to a family of dimension definitions which are based on *fractals* [5]. The box-counting dimension is a simplified version of Haussdorff dimension [5], [22]. The box-counting dimension D_B (or *Kolmogorov capacity*) of a set Ω is defined as follows: If $\nu(r)$ is the number of the boxes of size r needed to cover Ω, then D_B is

$$D_B = \lim_{r \to 0} \frac{\ln(\nu(r))}{\ln(\frac{1}{r})}. \quad (1)$$

Unfortunately, the box-counting dimension can be computed only for low-dimensional sets because the algorithmic complexity grows exponentially with the set dimension. Therefore, in our opinion, a good substitute for the box-counting dimension can be the *correlation dimension* [11]. Due to its computational simplicity, the correlation dimension is successfully used to estimate the dimension of attractors of dynamical systems. The correlation dimension is defined as follows: let $\Omega = \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x_N}$ be a set of points in $\mathbb{R}^n$ of cardinality N. If the *correlation integral* $C_m(r)$ is defined as:

$$C_m(r) = \lim_{N \to \infty} \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} I(\|\mathbf{x_j} - \mathbf{x_i}\| \leq r), \quad (2)$$

where I is an *indicator function*,[1] then the *correlation dimension* D of Ω is:

$$D = \lim_{r \to 0} \frac{\ln(C_m(r))}{\ln(r)}. \quad (3)$$

It has been proven [11] that the correlation dimension is a lower bound of the box-counting dimension, but, because of noise, the difference between the two is negligible in real applications. Some methods [28], [27] have been studied to obtain an optimal estimate for the correlation dimension, but all of these techniques work only when the correlation integral assumes a given form;[2] otherwise, the estimators can perform poorly [29]. Moreover, these methods generally require some heuristics to set the radius r [30]. Therefore, in our work, we used the original procedure (*GP algorithm*) proposed by Grassberger and Procaccia that consists of plotting $\ln(C_m(r))$ versus $\ln(r)$ and measuring the slope of the linear part of the curve (Fig. 1).

It has been proven [6], [26] that, in order to get an accurate estimate of the dimension D, the set cardinality N has to satisfy the following inequality:

$$D < 2 \log_{10} N. \quad (4)$$

Inequality (4) shows that the number N of data points needed to accurately estimate the dimension of a D-dimensional set is at least $10^{\frac{D}{2}}$. Even for low-dimensional sets this leads to huge values of N. The effect of N on the measure of the dimension can be seen in Table 1. This table reports the value of the measures of D obtained using the GP algorithm over sets of points randomly distributed in 10-dimensional hypercubes (supposed to have $D = 10$). The dimension is measured for different values of N and the error with respect to the supposed true dimension is reduced by increasing the number of data points from 10^3 to $10^{\frac{D}{2}} = 10^5$. In order to cope with this problem and to improve the reliability of the measure for low values of N, we develop an empirical procedure as described in the following section.

4 Intrinsic Dimension Estimation Procedure

Consider the set Ω of cardinality N. Our empirical procedure (EP) consists of the following steps:

1. $I(\lambda)$ is 1 iff condition λ holds, 0 otherwise.
2. For example, Takens' method is optimal iff $C_m(r) = ar^D[1 + br^2 + o(r^2)]$ where a, b are constants.

TABLE 1
Dependence of the Estimated Correlation Dimension on the Number of Data Points Used (the Actual Dimension of Data is 10)

points number	estimated dimension
1000	7.83
2000	7.94
5000	8.30
10000	8.56
30000	9.11
100000	9.73

1. A set Ω', whose ID d is known, with the same cardinality N as Ω is created. For instance, Ω' could be constituted by N points randomly generated in a d-dimensional hypercube.
2. The correlation dimension D of Ω' is measured with the GP algorithm.
3. The previous steps are repeated for T different values of d. The set $C = \{(d_i, D_i) : i = 1, 2, \dots, T\}$ is obtained.
4. A best-fitting to the points in C is performed. A plot (*reference curve*) Γ of D versus d is generated (see Fig. 2). The reference curve allows to estimate the value of D when d is known.
5. The correlation dimension D of Ω is computed and, using Γ, the intrinsic dimension of Ω can be estimated.

The above method is based on the following assumptions:

1. Γ depends on N.
2. Since the GP algorithm gives close estimates for sets of the same dimensionality and cardinality, the dependence of Γ on the Ω' sets used for its setup is negligible.

In comparison with topological methods, our approach offers the following advantages: it allows one to estimate the ID of high-dimensional data, unlike TRN-based method. Moreover, the proposed approach is based on the estimation of a fractal dimension and, therefore, allows one to obtain noninteger values. This latter advantage is quite important, since, due to the presence of noise, real data can sometimes lie within a *fractal-like* submanifold, whose dimension is usually noninteger.

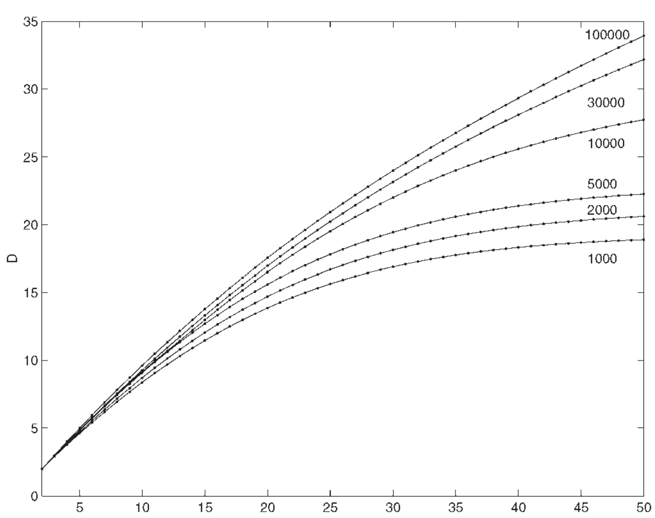

Fig. 2. Reference curves for different values of the cardinality N.

TABLE 2
ID Estimation by the GP Algorithm and EP of 8- and 23-Dimensional Data Sets

points	GP (d=8)	EP (d=8)	GP (d=23)	EP (d=23)
1000	6.83	7.86	14.99	22.95
2000	6.94	7.75	15.76	22.48
5000	7.42	7.98	17.09	23.21
10000	7.51	8.20	18.04	22.43
30000	7.65	8.13	19.10	23.20
100000	7.83	8.03	19.78	23.24

The first column reports the number of points used for the estimation. The second and third columns show the value of d estimated with GP and EP, respectively, for the 8-dimensional set. The last two columns show the estimates of d obtained with GP and EP for the 23-dimensional set.

5 EXPERIMENTAL RESULTS

The EP was tested by first creating reference curves for different values of the cardinality N, then by using each of them to estimate the dimension of data sets of the same cardinality and known dimension.

During our experimentation, the sets Ω' used for the reference curve setup were formed by randomly generated points in a d-hypercube. A plot was generated for the following cardinality values: 1,000, 2,000, 5,000, 10,000, 30,000, and 100,000. Correspondence to each value, a pair (d, D) was calculated for

$$d \in \{2, 3, 5, 10, 15, 18, 20, 25, 28, 30, 33, 35, 38, 40, 43, 45, 48, 50\}. \quad (5)$$

TABLE 3
Estimation of the Attractor Dimension of the Series *D* and *A* of Santa Fe Competition

points	GP (D)	EP (D)	GP (A)	EP (A)
1000	7.54	8.84	2.00	2.01
2000	7.87	8.90	2.01	2.02
5000	8.13	8.83	2.03	2.03
10000	8.25	9.09	2.03	2.03
30000	8.48	9.07		

The first column reports the number of points used for the estimation. The dimension values estimated for D series with GP and EP, respectively, are reported in second and third column. Last two columns show the d estimated with GP and EP using the data of the A series. A series has only 10,000 points.

The plot function is estimated by a multilayer-perceptron (MLP) [1]. Its structure was set up by means of the *Bayesian information criterion* [25]. The resulting reference curves can be seen in Fig. 2.

In order to test the method, several sets (with cardinalities corresponding to the values indicated above) were created composed of random points generated in hypercubes in spaces with dimension 8 and 23. These sets were assumed to have ID 8 and 23, respectively, and were not used to generate the reference curves Γ.

Following the procedure described in Section 4, the GP algorithm was first applied for each set, then the plot Γ corresponding to the same cardinality as the set being measured was used to compute the ID. The results are reported in Table 2.

The table shows the dimension estimation obtained with the GP algorithm and with the empirical procedure proposed here. Indeed, a remarkable improvement is obtained when the cardinality is low. Afterwards, in order to validate the EP procedure, the data set A [14] and D [23] of the Santa Fe time series competition were considered. Data Set A is a real data time series generated by a Lorenz-like chaotic system, implemented by $\mathrm{NH_3}$-FIR lasers. The data set D is a synthetic time series generated by a particle motion, simulated on a computer, with nine freedom degrees. The goal of the experimentation was to estimate, with the GP procedure, the attractor dimension of time series A and D. In order to estimate the attractor dimension, we used the method of delays [17], [22].

Given a time series $x(t)$, with $(t = 1, 2, \ldots, N)$, a set of points $\{X(t) : X(t) = [x(t), x(t-1), \ldots, x(t-d+1)]\}$ was obtained. If d is adequately large, between the manifold M, generated by the points $X(t)$ and the attractor U of the dynamic system that generated the time series, there is a diffeomorphism. Therefore, it is adequate to measure the dimension of M to infer the dimension of U.

We applied the method of delays to the data set A, considering its first 1,000, 2,000, 5,000, 10,000 points. The results, obtained with the GP and EP algorithms, are reported in Table 3. Since the value of the fractal dimension of the attractor of Lorenz's system is approximately 2.06, the result can be considered satisfactory. We then applied the method of delays to the data set D, considering the first 1,000, 2,000, 5,000, 10,000, 30,000 points. The results, with GP and EP, are shown in Table 3. Since the system that generated the data D has 9 degrees of freedom, the result can be considered particularly satisfactory.

6 Conclusions

This work has presented a new intrinsic dimension estimation technique based on a fractal method, namely, the Grassberger-Procaccia algorithm. The GP algorithm is effective when the dimension is low, but presents severe limits for high dimensionalities. Therefore, an empirical procedure, that improves GP algorithm, has been developed.

A comparison performed over sets, randomly generated in hypercubes of known dimensionality shows that the proposed procedure gives a significant improvement over the GP algorithm for high-dimensional sets. Further experiments, performed on data series of the Santa Fe competition, confirm this opinion. It is worth mentioning that the EP algorithm must not necessarily be based on the GP algorithm. It can be based also on other fractal dimension estimation methods (e.g., box-counting).

The EP algorithm is based on the assumption that the dependence of reference curves on the data used for their generation is negligible. In our experimentation, reference curves were obtained using data sets of uniformly distributed patterns whose dimension is known. It can be interesting to use other kinds of data to generate reference curves. This will allow one to evaluate the validity of the assumption that the dependence of the curves on the data used to generate them is negligible.

Acknowledgments

The authors would like to thank S. Bengio, M. Kanevsky, A. Morris, I. McCowan, D. Ugolini, and A. Verri. The authors are indebted to the anonymous reviewers for their valuable comments.

References

[1] C. Bishop, *Neural Networks for Pattern Recognition.* Cambridge, UK: Cambridge Univ. Press, 1995.

[2] D.S. Broomhead and M. Kirby, "A New Approach to Dimensionality Reduction: Theory and Algorithms," *SIAM J. Applied Math.,* vol. 60, no. 6, pp. 2114-2142, 2000.

[3] J. Bruske and G. Sommer, "Intrinsic Dimensionality Estimation with Optimally Topology Preserving Maps," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 5, pp. 572-575, May 1998.

[4] J.H. Conway and N.J.A. Sloane, "Sphere Packings, Lattices and Groups," *Grundlehren der Mathematischen Wissenschaften 290,* New York: Springer-Verlag, 1988.

[5] J.P. Eckmann and D. Ruelle, "Ergodic Theory of Chaos and Strange Attractors," *Rev., Modern Physics,* vol. 57, pp. 617-659, 1985.

[6] J.P. Eckmann and D. Ruelle, "Fundamental Limitations for Estimating Dimensions and Lyapounov Exponents in Dynamical Systems," *Physica,* vol. D56, pp. 185-187, 1992.

[7] D. Fotheringhame and R.J. Baddeley, "Nonlinear Principal Component Analysis of Neuronal Spike Train Data," *Biological Cybernetics,* vol. 77, pp. 282-288, 1997.

[8] F. Frisone, F. Firenze, P. Morasso, and L. Ricciardiello, "Application of Topological-Representing Networks to the Estimation of the Intrinsic Dimensionality of Data," *Proc. Int'l Conf. Artificial Neural Networks,* 1995.

[9] K. Fukunaga, "Intrinsic Dimensionality Extraction," *Classification, Pattern Recognition and Reduction of Dimensionality, Volume 2 of Handbook of Statistics,* P.R. Krishnaiah and L.N. Kanal, eds., pp. 347-360, Amsterdam: North Holland, 1982.

[10] K. Fukunaga and D.R. Olsen, "An Algorithm for Finding Intrinsic Dimensionality of Data," *IEEE Trans. Computer,* vol. 20, no. 2, pp. 165-171, Feb. 1976.

[11] P. Grassberger and I. Procaccia, "Measuring the Strangeness of Strange Attractors," *Physica,* vol. D9, pp. 189-208, 1983.

[12] S. Haykin and X. Bo Li, "Detection of Signals in Chaos," *Proc. IEEE,* vol. 83, no. 1, pp. 95-122, 1995.

[13] Q. Huang, J.R. Lorch, and R.C. Dubes, "Can the Fractal Dimension of Images Be Measured?" *Pattern Recognition,* vol. 27, no. 3, pp. 339-349, 1994.

[14] U. Hübner, C.O. Weiss, N.B. Abraham, and D. Tang, "Lorenz-Like Chaos in $\mathrm{NH_3}$-FIR Lasers," *Time Series Prediction. Forecasting the Future and Understanding the Past,* A. Weigend and N.A. Gershenfeld, eds., pp. 73-104, Reading, Mass.: Addison Wesley, 1994.

[15] V. Isham, "Statistical Aspects of Chaos: A Review," *Networks and Chaos-Statistical and Probabilistic Aspects,* O.E. Barndorff-Nielsen, J.L. Jensen, and W.S. Kendall, eds., pp. 124-200, London: Chapman & Hall, 1993.

[16] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data.* Englewood Cliffs, N.J.: Prentice Hall, 1988.

[17] D. Kaplan and L. Glass, *Understanding Nonlinear Dynamics.* New York: Springer-Verlag, 1995.

[18] M. Kirby, *Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns.* New York: John Wiley and Sons, 2001.

[19] E.C. Malthouse, "Limitations of Nonlinear PCA as Performed with Generic Neural Networks," *IEEE Trans. Neural Networks,* vol. 9, no. 1, pp. 165-73, 1998.

[20] B. Mandelbrot, *Fractals: Form, Chance, and Dimension.* San Francisco: Freeman, 1977.

[21] T. Martinetz and K. Schulten, "Topology Representing Networks," *Neural Networks,* vol. 3, pp. 507-522, 1994.

[22] E. Ott, *Chaos in Dynamical Systems.* Cambridge, UK: Cambridge Univ. Press, 1993.

[23] F.J. Pineda and J.C. Sommerer, "Estimating Generalized Dimensions and Choosing Time Delays: A Fast Algorithm," *Time Series Prediction. Forecasting the Future and Understanding the Past,* A. Weigend and N.A. Gershenfeld, eds., pp. 367-385, Reading, Mass.: Addison Wesley, 1994.

[24] J.A. Scheinkman and B. Le Baron, "Nonlinear Dynamics and Stock Returns," *J. Businness,* vol. 62, pp. 311-337, 1989.

[25] G. Schwartz, "Estimating the Dimension of a Model," *Annals of Statistics,* vol. 6, pp. 497-511, 1978.

[26] L.A. Smith, "Intrinsic Limits on Dimension Calculations," *Physics Letters,* vol. A133, pp. 283-288, 1988.

[27] R.L. Smith, "Optimal Estimation of Fractal Dimension," *Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity,* M. Casdagli and S. Eubank, eds., vol. XII, pp. 115-135, 1992.

[28] F. Takens, "On the Numerical Determination of the Dimension of an Attractor," *Proc. Groningen 1984 Dynamical Systems and Bifurcations,* B. Braaksma, H. Broer, and F. Takens, eds., no. 1125, pp. 99-106, 1985.

[29] J. Theiler, "Lacunarity in a Best Estimator of Fractal Dimension," *Physics Letters,* vol. A133, pp. 195-200, 1988.

[30] J. Theiler, "Statistical Precision of Dimension Estimators," *Physical Rev.,* vol. A41, pp. 3038-3051, 1990.

[31] W.S. Tirsch, M. Keidel, S. Perz, H. Scherb, and G. Sommer, "Inverse Covariation of Spectral Density and Correlation Dimension in Cyclic EEG Dimension of the Human Brain," *Biological Cybernetics,* vol. 82, pp. 1-14, 2000.

[32] V. Vapnik, *Statistical Learning Theory.* New York: John Wiley and Sons, 1998.