

Galaxy classification using fractal signature

S. Lekshmi¹, K. Revathy¹, and S. R. Prabhakaran Nayar²

¹ Department of Computer Science, University of Kerala, Kariavattom, Trivandrum 695581, India
e-mail: kr@md2.vsnl.net.in

² Department of Physics, University of Kerala, Kariavattom, Trivandrum 695581, India

Received 10 October 2002 / Accepted 11 March 2003

Abstract. Fractal geometry is becoming increasingly important in the study of image characteristics. For recognition of regions and objects in natural scenes, there is always a need for features that are invariant and they provide a good set of descriptive values for the region. There are many fractal features that can be generated from an image. In this paper, fractal signatures of nearby galaxies are studied with the aim of classifying them. The fractal signature over a range of scales proved to be an efficient feature set with good discriminating power. Classifiers were designed using nearest neighbour method and neural network technique. Using the nearest distance approach, classification rate was found to be 92%. By the neural network method it has been found to increase to 95%.

Key words. galaxies: fundamental parameters (classification) – techniques: image processing – methods: data analysis

1. Introduction

The multivariate nature of galaxies makes classification of galaxies difficult and not very accurate. The application of pattern recognition techniques could develop machine automated classifiers which simplify the above task. Several experts have derived galaxy classification and star-galaxy classification based on input features extracted from images (Burda & Feitzinger 1992; Han 1995; Naim et al. 1995; Zaritsky 1995; Philip et al. 2002). Mahonen & Frantti (2000) have developed a classifier for star-galaxy separation based on fuzzy set reasoning. Their results showed that neural networks give better results than direct fuzzy classification does. The classification of galaxies into simple and highly complex groups was done by Thanki et al. (2000) using capacity dimension and correlation dimension as features. It was found that ellipticals tend to have a lower value of the average correlation dimension than the spirals. Goderya & Lolling (2002) describe a computational scheme to develop an automatic galaxy classifier. The model employs either geometric shape features or direct pixel images for classification purposes. The classification of blended images was studied by Thayer et al. (1996) using artificial neural networks to classify images of APS catalog. The different algorithms for classification of galaxies like Principal Component Analysis, encoder neural networks, supervised neural networks using back-propagation or quasi-Newton algorithms were compared by Lahav et al. (1996). A comparison of neural

network algorithms and preprocessing methods for identifying stars and galaxies was done by Bazell & Peng (1998). Andreon & Davoust (1997) have classified galaxies in clusters based on the isophotal analysis of galaxy images. Storrie-Lombardi et al. (1992) have employed artificial neural networks for galaxy classification by assigning Bayesian probabilities to the corresponding morphological type of the galaxy. Neural network classifiers were developed by Odewahn (1995) for automated classification of galaxies using photometric parameters such as mean surface brightness, color, concentration index and surface brightness. Almost all of these classification procedures required a priori knowledge about how the input feature set is related to the galaxy type. Studies have been made on astrophysical bodies using fractal techniques. Feitzinger et al. (1987) observed that fractal feature is a constant parameter for certain galaxies. Elmegreen et al. (2001) have studied the fractal structure of star formation using the size distribution function of stellar aggregates. Perdan (1990) has classified astrophysical fractals into two groups, namely fractal configuration in space-time and fractal configuration in parameter spaces. Wu et al. (1997) have studied the application of multi-fractal formalism in order to classify the galaxy structure.

The different classification procedures developed in the past few decades employed either statistical methods or neural network techniques. In our earlier work, galaxies were classified making use of the fractal feature, namely fractal dimension (Revathy & Lekshmi 2003). The success rate was found to be 73%. By including another feature, namely the spectral flatness measure (Revathy et al. 2000) of the image, the success

Send offprint requests to: S. R. Prabhakaran Nayar,
e-mail: srp@md2.vsnl.net.in

rate was increased to 80%. In the present work, we have considered the fractal signature which has been found to give better results for the classification of galaxies.

2. Selection of galaxies

The galaxies considered for the present work are the nearby galaxies listed in the Nearby Galaxy Catalog (Frei et al. 1996). It consists of 113 galaxies and spans morphological types from -5 to 12 . The galaxies with morphological types -5 and -4 belong to ellipticals, -3 to -1 to lenticulars, 0 to 9 form spirals and the rest are irregulars. Fourteen galaxies belong to morphological type -5 and -4 , eight belong to morphological type -3 to -1 , ninety belong to morphological type from 0 to 9 and one belongs to morphological type greater than 9 . In this work, we have considered galaxies belonging to the first and third groups, leaving out lenticulars and irregulars.

3. Fractal feature and methodology

Fractals, introduced by Mandelbrot provide a promising approach used to model many homogeneous structures (Mandelbrot 1983). A theoretical fractal object is self-similar under all magnifications and the changes in properties with respect to change in scale are limited. There are many fractal features that can be generated from an image. Fractal dimension remains the primary characteristic. Mathematically it is invariant to changes in scale and can characterize the roughness of the surface. Fractals are used to model hierarchical structures in several areas of astrophysics, especially the distribution of matter at various scales in the universe. Several studies have been made regarding the fractal nature of galaxies and the distribution of galaxies in space (Jones 1999). The changes in picture properties with change in scale have been investigated by several authors (Peleg et al. 1984). One of the important properties of fractal objects is the surface area. The change in measured area taken at different scales with changing scale is used as the fractal signature of the image. Peleg et al. (1984) introduced a texture analysis method that measures the area of the gray level surface at varying resolutions. For a pure fractal gray level image the area is computed as

$$A(\epsilon) = F\epsilon^{2-D} \quad (1)$$

where ϵ is the resolution of the gray levels in the image, D is fractal dimension and F is a constant. The change in measured area with changing scale is used as the fractal signature of the texture. The surface area of the galaxy image is computed by the method suggested by Mandelbrot for curve measurement. The idea is to cover the gray level surface with a blanket having upper surface u_ϵ and lower surface l_ϵ . For $\epsilon = 0$, they are initialized to the gray level values of the image,

$$u_\epsilon = g(i, j) = l_\epsilon \quad (2)$$

where $g(i, j)$ represents gray level function.

From $\epsilon = 1$ onwards, u_ϵ is computed as the maximum of the upper surface for $u_{\epsilon-1}$ and l_ϵ is computed as the minimum of the lower surface for $l_{\epsilon-1}$. They are given by

$$u_\epsilon(i, j) = \max\left([u_{\epsilon-1}(i, j) + 1], \max_{|(m,n)-(i,j)\leq 1|}[u_{\epsilon-1}(m, n)]\right) \quad (3)$$

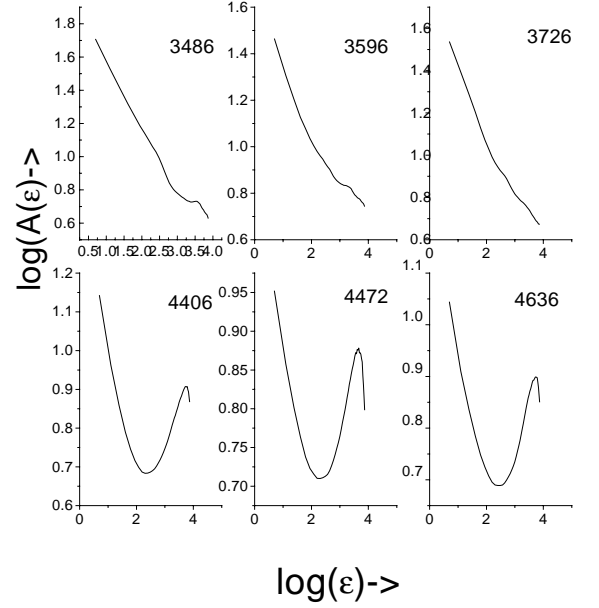


Fig. 1. Variation of $\log A(\epsilon)$ with $\log \epsilon$ on log–log scale for spirals (top) and ellipticals (bottom). The number in the figure indicates the NGC number of the galaxy.

$$l_{\epsilon(i,j)} = \min\left([l_{\epsilon-1}(i, j) - 1], \min_{|(m,n)-(i,j)\leq 1|}[l_{\epsilon-1}(m, n)]\right). \quad (4)$$

The image points (m, n) with distance less than one from (i, j) were taken to be four immediate neighbors of (i, j) . For computing u_ϵ and l_ϵ at different points, the four immediate neighbors are considered. The difference between the upper and lower surfaces for a scale gives the volume of the blanket for that scale. The volume is given by

$$V = \sum_{i,j} (u_\epsilon(i, j) - l_\epsilon(i, j)). \quad (5)$$

The surface area, from which one can determine whether the surface is a fractal or not is computed as

$$A(\epsilon) = [V_\epsilon - V_{\epsilon-1}] / 2\epsilon. \quad (6)$$

The surface area gives a measure of the oscillations of the underlying surface for each scale (Espinal et al. 1998).

If the image is a fractal, the plot of $A(\epsilon)$ versus ϵ on a log–log scale is a straight line. Typical variations of $A(\epsilon)$ with ϵ for spiral and elliptical galaxies are given in Fig. 1. Note that the plots of spirals tend to be a straight line, which substantiates that the images of spiral galaxies have more pronounced fractal characteristics than those of ellipticals. The slope $S(\epsilon)$ of $A(\epsilon)$ is defined as the fractal signature. Here the fractal signature is computed using features derived for 47 scales. The fractal signatures for $\epsilon = 2, 3, \dots, 48$ were computed for elliptical and spiral galaxies. These are depicted in Figs. 2 and 3. It could be observed that fractal signatures of the two groups resemble each other. The fractal signatures of both types of galaxies vary with the morphological type and this is more pronounced in spirals than in ellipticals.

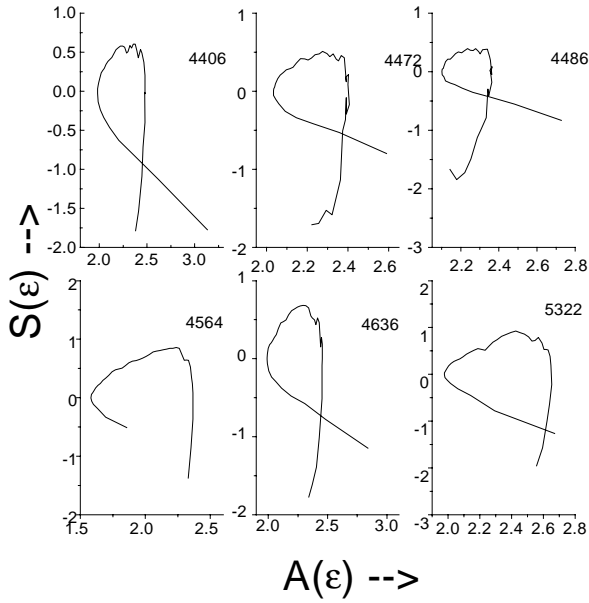


Fig. 2. Variation of fractal signature $S(\epsilon)$ with area $A(\epsilon)$ for elliptical galaxies. The number in the figure indicates the NGC number of the galaxy.

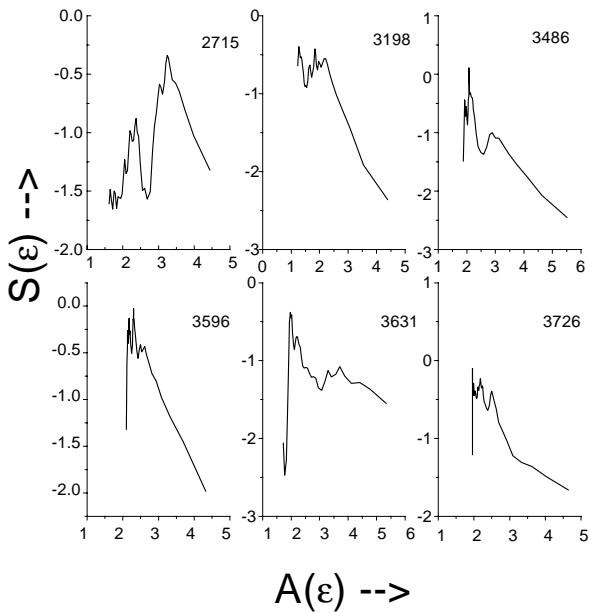


Fig. 3. Variation of fractal signature $S(\epsilon)$ with area $A(\epsilon)$ for spiral galaxies. The number in the figure indicates the NGC number of the galaxy.

4. Classification methods

4.1. Nearest distance approach

Galaxy images were classified into ellipticals and spirals by comparing the distances between their fractal signatures (M1). For two images i_1 and i_2 , whose signatures are S_1 and S_2 , the distance is defined as,

$$D(i_1, i_2) = \sum (S_1(\epsilon) - S_2(\epsilon))^2 \log \left(\left(\epsilon + \frac{1}{2} \right) / \left(\epsilon - \frac{1}{2} \right) \right). \quad (7)$$

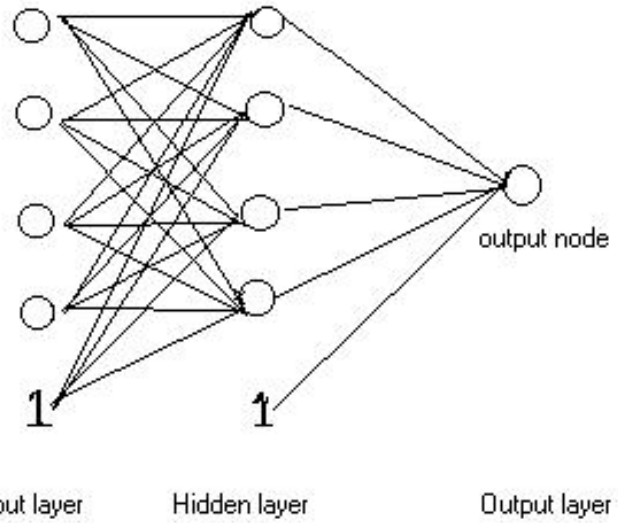


Fig. 4. Neural network architecture.

The first three images of elliptical galaxies were used to define the elliptical cluster and the first ten images of spiral galaxies were used to define the spiral cluster. The clusters were defined by taking the average of the fractal signature for the various scales defined. Out of 14 ellipticals, only one was misclassified and out of 90 spirals, only seven were misclassified.

4.2. Neural network approach

To obtain a better classification rate, a classifier based on neural network was designed using these features as input parameters (M2). The back-propagation algorithm is used for the purpose. The architecture of the network is a multi-layered one where the nodes in a layer are fully connected to the nodes in the next layer (Gose et al. 2000). The neural network architecture is shown in Fig. 4. The input layer contains nodes representing the forty seven signature values and the hidden layer contains four nodes. The output layer consists of the output node. A threshold is applied to the value of the output node to determine the class. Also, a bias node whose value equals one is added to input layer and hidden layer. The back propagation algorithm consists of two steps:

(i) A feed forward step in which the output of the nodes comprising the hidden layers and the output layer are computed. The output values are calculated as a linear combination of weights and node values of the previous layer. This result is then operated on by the sigmoid function given by $R(s) (= \frac{1}{1+e^{-s}})$. The output value of the j th computed node is

$$x_j^{(k+1)} = R \left(\sum w_{ij}^{(k+1)} x_j^{(k)} \right). \quad (8)$$

Here $x_j^{(k)}$ is the value of the j th node in k th layer and $w_{ij}^{(k+1)}$ is the weight of the link connecting i th node in k th layer to j th node in $(k + 1)$ th layer.

(ii) A back-propagation step where the weights are updated backwards from the output layer to one or more hidden layers. The back propagation step uses the steepest descent method to

Table 1. NGC number of the galaxy and classification results. DSC: distance to Spiral cluster, DEC: distance to Elliptical cluster, DVC: derived class, DS: desired class, M1: nearest neighbour technique M2: neural network technique OP: output value by neural network method.

NGC No.	Nearest neighbour (M1)			Neural network (M2)		DS
	DEC	DSC	DVC	OP value	DVC	
2768	0.09	0.32	E	0.01	E	E
3377	0.02	0.56	E	0.007	E	E
3379	0.04	0.61	E	0.005	E	E
4125	0.06	0.46	E	0.47	E	E
4365	1.33	0.52	S	0.99	S	E
4374	0.20	0.29	E	0.34	E	E
4406	0.13	0.25	E	0.42	E	E
4472	0.33	1.07	E	0.29	E	E
4486	0.37	0.96	E	0.06	E	E
4621	0.22	1.19	E	0.004	E	E
4636	0.06	0.60	E	0.044	E	E
5322	0.08	0.84	E	0.005	E	E
5813	0.09	0.43	E	0.01	E	E
4564	0.77	2.13	E	0.004	E	E
3166	0.30	0.27	S	0.99	S	S
5701	1.20	0.22	S	0.99	S	S
3623	1.38	0.91	S	0.99	S	S
4594	4.86	2.95	S	0.99	S	S
5377	0.39	0.37	S	0.99	S	S
2775	0.30	0.11	S	0.99	S	S
2985	0.98	0.24	S	0.99	S	S
3031	0.46	0.93	E	0.99	S	S
3368	0.57	0.15	S	0.99	S	S
4192	0.67	0.36	S	0.99	S	S
4450	0.38	0.04	S	0.99	S	S
4569	0.43	0.32	S	0.99	S	S
4725	0.46	0.28	S	0.99	S	S
4826	0.41	0.45	E	0.99	S	S
2683	2.43	1.17	S	0.99	S	S
3351	0.56	0.32	S	0.99	S	S
3675	1.28	0.44	S	0.99	S	S
4013	1.98	0.70	S	0.99	S	S
4216	0.13	0.40	E	0.01	E	S
4394	1.93	0.89	S	0.99	S	S
4501	0.70	0.17	S	0.99	S	S
4548	0.16	0.55	E	0.96	S	S
4579	0.05	0.40	E	0.91	S	S
4593	0.34	0.06	S	0.98	S	S
5746	1.94	1.44	S	0.99	S	S
5792	1.47	0.64	S	0.99	S	S
5850	1.24	0.54	S	0.99	S	S
5985	1.38	0.57	S	0.99	S	S
2903	0.69	0.24	S	0.99	S	S
3147	1.33	0.64	S	0.99	S	S
3344	2.20	0.87	S	0.99	S	S
3953	1.06	0.26	S	0.99	S	S
4030	3.53	1.72	S	0.99	S	S
4088	6.44	3.89	S	0.99	S	S
4258	1.27	0.54	S	0.99	S	S
4303	3.76	1.66	S	0.99	S	S
4321	0.87	0.27	S	0.99	S	S
4689	3.01	1.43	S	0.99	S	S

Table 2. NGC number of the galaxy and classification results. DSC: distance to Spiral cluster, DEC: distance to Elliptical cluster, DVC: derived class, DS: desired class, M1: nearest neighbour technique M2: neural network technique OP: output value by neural network method.

NGC No.	Nearest neighbour (M1)			Neural network (M2)		DS
	DEC	DSC	DVC	OP value	DVC	
5055	0.72	0.07	S	0.99	S	S
5248	4.33	2.10	S	0.99	S	S
5364	2.07	0.90	S	0.99	S	S
5371	3.09	1.37	S	0.99	S	S
6384	2.00	0.66	S	0.99	S	S
2715	3.07	1.91	S	0.99	S	S
2976	2.82	1.53	S	0.99	S	S
3079	6.06	4.52	S	0.99	S	S
3198	2.92	1.14	S	0.003	E	S
3486	3.76	1.72	S	0.99	S	S
3596	1.28	0.27	S	0.99	S	S
3631	3.45	1.80	S	0.99	S	S
3672	3.09	1.80	S	0.99	S	S
3726	1.86	0.64	S	0.99	S	S
3810	4.88	2.86	S	0.99	S	S
3877	1.93	1.07	S	0.99	S	S
3893	2.68	1.52	S	0.99	S	S
3938	3.79	1.91	S	0.99	S	S
4123	1.38	0.57	S	0.99	S	S
4136	2.56	0.91	S	0.99	S	S
4157	1.30	1.07	S	0.99	S	S
4254	7.84	4.58	S	0.99	S	S
4414	2.26	0.87	S	0.99	S	S
4535	1.88	0.56	S	0.99	S	S
4651	2.17	0.83	S	0.99	S	S
5033	0.55	0.22	S	0.99	S	S
5334	19.47	14.95	S	0.99	S	S
2403	2.52	0.97	S	0.99	S	S
2541	9.89	6.60	S	0.003	E	S
3184	2.43	1.18	S	0.99	S	S
3319	9.60	6.42	S	0.99	S	S
3556	4.46	2.28	S	0.99	S	S
4144	2.92	2.00	S	0.99	S	S
4189	1.98	0.71	S	0.99	S	S
4487	1.79	0.57	S	0.99	S	S
4559	1.98	0.70	S	0.99	S	S
4654	2.69	1.00	S	0.99	S	S
4731	4.12	1.92	S	0.99	S	S
5669	2.35	0.80	S	0.99	S	S
6015	4.28	2.17	S	0.99	S	S
6118	1.42	0.76	S	0.99	S	S
6503	5.17	3.01	S	0.99	S	S
4498	4.98	2.90	S	0.99	S	S
4571	5.08	2.94	S	0.99	S	S
5585	1.97	0.66	S	0.99	S	S
4178	3.66	1.72	S	0.99	S	S
4242	24.50	18.56	S	0.99	S	S
4861	4.46	2.28	S	0.99	S	S
5204	7.49	4.54	S	0.99	S	S
4449	6.70	4.03	S	0.99	S	S
4527	0.19	0.26	E	0.99	S	S
5005	0.59	0.73	E	0.02	E	S

Table 3. Class-wise performance using nearest neighbour classifier.

Galaxy type	Classification		
	Ellipticals	Spirals	Total
Ellipticals	13	1	14
Spirals	7	83	90
Total	20	84	104

Table 4. Class-wise performance using neural network classifier.

Galaxy type	Classification		
	Ellipticals	Spirals	Total
Ellipticals	13	1	14
Spirals	4	86	90
Total	17	87	104

Table 5. Representation of success rates using the two classifiers.

Galaxy type	Success rate	
	Nearest neighbor	Neural network
Ellipticals	92.8%	92.8%
Spirals	92.2%	95.5%
Total	92.3%	95.1%

update the weights so that the error function $E = \frac{1}{2} \sum (x_j^{(k)} - d_j)$ is minimized where d_j is the desired output class.

It took 15 000 iterations for the algorithm to converge, where the learning rate is 0.2. Out of 14 ellipticals, only one was misclassified and out of 90 spirals, only four were misclassified. The performance of the classifiers designed using nearest neighbour method (M1) and neural network technique (M2) is given in Tables 1 and 2. The tables contain NGC numbers of the galaxies and the output results of the two methods (M1 and M2). The second and third columns represent the distances evaluated by M1, to the elliptical (DEC) and spiral clusters (DSC) respectively. Column 5 represents the values of the output node of the neural network (OP M2). DVC M1 and DVC M2 are the results of the computed class using methods M1 and M2 and DS is the expected class.

Table 3 presents the result of the classification for the elliptical and spiral galaxies using the nearest neighbour classifier. Table 4 presents the results for the elliptical and spiral galaxies using the neural network classifier. Table 5 compares the success rate for the two classes of galaxies using nearest neighbour and neural network techniques. It can be observed from

Tables 3–5 that, in general, the neural network (M2) approach succeeded in classifying certain galaxies for which the nearest neighbour technique failed.

5. Conclusion

We have shown that fractal signature is a promising feature for galaxy classification. Using the nearest distance approach, classification rate was found to be 92%, and by neural network, it was 95%. The dimensionality reduction technique will be included in future work, to reduce the complexity of the algorithm.

Acknowledgements. The authors thank the Indian Space Research Organization for the partial support of this work through the RESPOND project.

References

- Andreon, S., & Davoust, E. 1997, *A&A*, 319, 747
 Bazell, D., & Peng, Y. 1998, *ApJS*, 116, 47
 Burda, P., & Feitzinger, J. V. 1992, *A&A*, 261, 697
 Elmegreen, B. G., & Elmegreen, D. M. 2001, *AJ*, 121, 1507
 Espinal, F., Hunberger, T. L., Jawerth, B., & Kubota, T. 1998, *Opt. Eng.*, 37, 166
 Feitzinger, J. V., & Galinski, T. 1987, *ApJ*, 179, 249
 Frei, Z., Guhathakurta, P., Gunn, J. E., & Tyson, T. 1996, *AJ*, 111, 174
 Goderya, S. N., & Lolling, S. 2002, *Ap&SS*, 279, 377
 Gose, E., Johnsonbaugh, R., & Jost, S. 2000, *Pattern Recognition and Image Analysis: PHI*, 2000
 Han, M. 1995, *ApJ*, 442, 504
 Jones, B. J. T. 1999, *MNRAS*, 307, 376
 Lahav, O., Naim, A., Sodre, L., Jr., et al. 1996, *MNRAS*, 283, 207
 Mahonen, P., & Frantti, T. 2000, *ApJ*, 541, 261
 Mandelbrot, B. B. 1983, *The Fractal Geometry of Nature* (Oxford: W. H. Freeman)
 Naim, A., Lahav, O., Sodre, L., Jr., & Storrie-Lombardi, M. C. 1995, *MNRAS*, 275, 567
 Odewahn, S. C. 1995, *PASP*, 107, 770
 Peleg, S., Naor, J., Hartley, R., & Avnir, D. 1984, *IEEE Trans. PAMI*, 6, 518
 Perdang, J. 1990, *Vistas Astron.*, 33, 249
 Philip, N. S., Wadadekar, Y., Kembhavi, A., & Joseph, K. B. 2002, *A&A*, 385, 1119
 Revathy, K., & Lekshmi 2003, *Fractals*, in press
 Revathy, K., Raju, G., & Nayar, S. R. P. 2000, *Fractals*, 8, 247
 Storrie-Lombardi, M. C., Lahav, O., Sodre, L., Jr., & Storrie-Lombardi, L. J. 1992, *MNRAS*, 259, 8
 Thanki, S., Rhee, G., & Lepp, S. 2000, *Am. Astron. Soc. Meet.*, 196, 2905
 Thayer, D. N., Webster, J. G., Larsen, J. A., et al. 1996, *BAAS*, 28, 907
 Wu, K. K. S., Alexander, P., & Sleath, J. P. 1997, *MNRAS*, 286, 777
 Zaritsky, D. 1995, *AJ*, 110, 1602