HAL open science

Unsupervised Detection of Fibrosis in Microscopy Images Using Fractals and Fuzzy c-Means Clustering

S. K. Tasoulis, Ilias Maglogiannis, V. P. Plagianakos

To cite this version:

S. K. Tasoulis, Ilias Maglogiannis, V. P. Plagianakos. Unsupervised Detection of Fibrosis in Microscopy Images Using Fractals and Fuzzy c-Means Clustering. 8th International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2012, Halkidiki, Greece. pp.385-394, 10.1007/978-3-642-33409-2_40 . hal-01521436

HAL Id: hal-01521436

https://inria.hal.science/hal-01521436v1

Submitted on 11 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Unsupervised Detection of Fibrosis in Microscopy Images Using Fractals and Fuzzy c-means Clustering

S.K. Tasoulis, I. Maglogiannis, and V.P. Plagianakos

Department of Computer Science and Biomedical Informatics,
University of Central Greece, Papassiopoulou 2–4, Lamia, 35100, Greece
{stas,imaglo,vpp}@ucg.gr

Abstract. The advances in improved fluorescent probes and better cameras in collaboration with the advent of computers in imaging and image analysis, assist the task of diagnosis in many fields of biologic and medical research. In this paper, we introduce a computer-assisted image characterization tool based on a Fuzzy clustering method for the quantification of degree of Idiopathic Pulmonary Fibrosis (IPF) in medical images. The implementation of this algorithmic strategy is very promising concerning the issue of the automated assessment of microscopic images of lung fibrotic regions.

Key words: Image Analysis, Cluster Analysis, Fuzzy Clustering, Fractal Dimension.

1 Introduction

Idiopathic pulmonary fibrosis (IPF), also referred to as cryptogenic fibrosing alveolitis, is a chronic, progressive and usually lethal lung disorder of unknown aetiology. Median survival of newly diagnosed patients with IPF is about 3 years, similar to that of clinical non-small cell lung cancer. At present there are no proven therapies for IPF [18]. IPF usually affects patients aged 50–70 years, with a male preponderance, evidenced by a male-to-female ratio of 2:1. The incidence of IPF has been estimated at 10.7 cases per 100,000 men and 7.4 cases per 100,000 women and the incidence appears to be rising [6].

Advances in improved fluorescent probes [5] and better cameras [11, 21] have expanded the capabilities of the light microscope and its usefulness in biologic and medical research. The advent of computers in imaging and image analysis permits quantitative computer analysis. Thus, automatic classification and assessment of microscopic images is possible in an attempt to assist the task of diagnosis.

The field of microscopy image analysis has occupied several research teams and significant research work may be found in the literature of this field. In [14] a tool that classifies biological microscopic images of lung tissue sections with

idiopathic pulmonary fibrosis was presented. Similar tools have also been proposed for the assessment of liver fibrosis [4, 16, 22, 2], the study of micro vascular circulating leukocytes [9], the assessment of testicular interstitial fibrosis [19, 20], or that of lung fibrosis [12]. The use of pattern recognition or classification methods like Support Vector Machines or Neural Networks has enabled the design of decision-making algorithms, appropriate to microscopic data. Within this context, a method for evaluation of electron microscopic images of serial sections based on the Gabor wavelets and the construction of a mapping between the model and the target image has been proposed in [13].

In the present study, we employ one of the most widely used fuzzy clustering algorithms; namely the Fuzzy c-means [3]. The goal is to identify the presence and the degree of lung pathology caused by idiopathic pulmonary fibrosis (IPF), which is a chronic, progressive and usually lethal lung disorder of unknown etiology [1], whose variability concerning the severity of the lesions it incurs in the lung is great when assessed by microscopic histological images [12]. Fuzzy c-means clustering (and more over an improved version of fuzzy c-means) has also been applied in environmental cases [10]

The Fuzzy c-means clustering method was applied to digital images of sections, captured using a Nikon ECLIPSE E800 microscope and a Nikon digital camera DXM1200 at a magnification of 4x. Each image was partitioned into windows of specified size and features were extracted for each window. The main contribution of the this study is the calculation of the fractal dimension of the microscopic images and its incorporation to the clustering algorithm.

The rest of the paper is structured as follows. In Section 2 we briefly review the Fuzzy c-means clustering algorithm and we outline the fractal dimension and the box-counting method for its computation. Next in Section 3 we introduce the proposed methodology and finally in Section 4 we present the experimental analysis and results. The paper ends with concluding results and pointers for future work.

2 Background Information

2.1 Fuzzy Clustering

Data clustering is the process of partitioning a set of data vectors into disjoint groups (clusters), so that objects of the same clusters are more similar to each other than objects in different clusters. Different measures of similarity may be used in the clustering process, where the similarity measure controls how the clusters are formed. Some examples of measures that can be used in clustering include distance, connectivity, and intensity. In traditional (hard) clustering, data are divided into distinct clusters, where each data vector belongs to exactly one cluster. On the other hand, in fuzzy (soft) clustering, data vectors can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning

these membership levels and then using them to assign data elements to one or more clusters.

One of the most widely used fuzzy clustering algorithms is the Fuzzy C-Means (FCM) algorithm. FCM is a clustering method that assigns each each data point to a cluster to some degree that is specified by a membership grade. This technique was originally introduced by Jim Bezdek in [3] as an improvement on earlier clustering methods and attempts to partition a finite collection of data vectors into a collection of fuzzy clusters with respect to some given criterion. A theoretical discussion of FCM can be found in [7].

Given a finite set of data, the algorithm returns a list of cluster centers and a partition matrix indicating the degree to which each element belongs to a given cluster. Like the k-means algorithm, the FCM aims to minimize an objective function, like the following:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m \leq \infty, \tag{1}$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the i-th of d-dimensional measured data, c_j is the d-dimension center of the cluster, and $\|\cdot\|$ is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} (\frac{\|x_i - c_j\|}{\|x_i - c_k\|})^{\frac{2}{m-1}}}, \tag{2}$$

where

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}.$$

This iteration stops when

$$max_{ij}\{|u_{ij}^{k+1} - |u_{ij}^k|\} \leq \epsilon, \tag{3}$$

where k is the iteration number and ϵ is a constant between 0 and 1 that controls the termination of algorithm. This procedure converges to a local minimum or a saddle point of J_m.

2.2 Fractals and Fractal Dimension

It is known that a fractal designates a rough or fragmented geometric shape that can be subdivided into parts, each of which is a reduced-size copy of the whole [8]. Fractals are generally self-similar and independent of scale. In general nature conforms to fractals much more than it does to classical shapes and hence fractals can serve as models for many natural phenomena.

Contrary to classical geometry, fractals are not regular and may have a non-integer dimension. Fractal concepts have provided a new approach for quantifying the geometry of complex or noisy shapes and objects. Fractal geometry has been proven capable of quantifying irregular patterns, such as tortuous lines, crumpled surfaces and intricate shapes, and estimating the ruggedness of systems [15].

In this study, we incorporate the fractal dimension in an attempt to guide the FCM algorithm to construct meaningful clusters. The box-counting approach is one of the frequently used techniques to estimate the FD of an image. Among a number of techniques [15], it was found to be the most appropriate method of fractal dimension estimation.

The box-counting method is easy, automatically computable, and applicable for patterns with or without self-similarity [8, 17]. In the box counting procedure, each input image is covered by a sequence of grids of descending sizes and for each of the grids, two values are recorded. These are the number of square boxes intersected by the image, $N(s)$, and the side length of the squares, s.

Then the fractal dimension is calculated by the regression slope D ($1 \leq D \leq 2$) of the straight line formed by plotting $\log(N(s))$ against $\log(1/s)$ [15]. An image having a fractal dimension of 1, or 2, is considered as completely differentiable, or very rough and irregular, respectively. The linear regression equation used to estimate the fractal dimension is

$$\log(N(s)) = \log(K) + D \log(1/s), \tag{4}$$

where K is a constant and $N(s)$ is proportional to $(1/s)D$.

3 Proposed methodology

For this study, Age- and sex- matched, 6-8 week-old mice were used for the induction of pulmonary fibrosis by a single intravenous injection with a dose of 100mg/Kg of body weight (100 mg/kg body weight; 1/3 LD50; Nippon Kayaku Co. Ltd., Tokyo). Bleomycin administration initially induces lung inflammation that is followed by a progressive destruction of the normal lung architecture. To monitor disease initiation and progression, mice were sacrificed at 7, 15 and 23 days after bleomycin injection. Mice injected with saline alone and sacrificed 23 days post injection, served as the control group. For pathology assessment, at each time point, bronchoalveolar lavage (BAL) has been performed (3x 1ml Saline) for the estimation of total and differential cell populations. Finally after perfusion of lungs via the heart ventricle with 10ml Phosphate Buffer Saline, lungs were then removed, weighed dissected and collected, for histology. Sagittal sections from the right lung were used for Hematoxylin and Eosin staining and histopathologic analysis.

When we retrieve the microscopic images the procudure continues as follows. Initially, the input microscopic images are being binirized using the Otsu's method. This method is one of many binarization algorithms, i.e. it automatically performs reduction of a graylevel image to a binary one. The algorithm

assumes that the image contains two classes of pixels (e.g. foreground and background) and calculates the optimum threshold separating those two classes so that their combined spread (intra-class variance) is minimal. At the next step we segment the binary image to $n \times n$ windows and then we calculate Fractal Dimension (FD) of each window. Fractal dimension is a useful feature proposed to characterize roughness and self-similarity in a picture, which has been used for texture segmentation, shape classification, and graphic analysis in many fields. It can be defined as a ratio providing a statistical index of complexity comparing how detail in a pattern changes with the scale at which it is measured. Finally based on the Fractal Dimension and other image properties we perform clustering using Fuzzy c-means (FCM) clustering algorithm. The fuzzification parameter m that we use in this work is set to the default value 1.25. The complete scheme of the proposed method is shown at Figure 1. Next at Figure 2 an example of an image is illustrated in various steps of the procedure. From left to right we can see the input microscopic image, the greyscale image, the binary image, and the image that shows the clustering result.

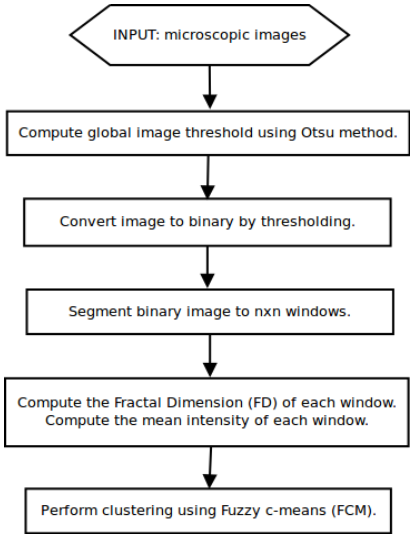

Fig. 1. Diagram of the proposed methodology.

4 Experiment setup and results

The results that are presented in this section are based on images that correspond to 7, 15 and 23 days after bleomycin administration in mice (see Figure 3). The lung images captured using a Nikon ECLIPSE E800 microscope and a Nikon DXM1200 digital camera, were initially separated to windows of size 100×100

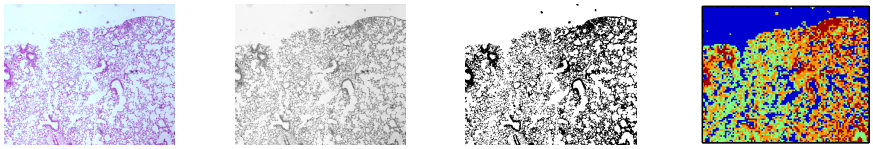

Fig. 2. An example of an image in various steps of the proposed methodology

pixels and the box-counting method was applied at each window to calculate the corresponding fractal dimension. Next, the images were separated into windows of smaller sizes (5, 10, and 20, respectively). Each of this windows corresponds to a data vector, with attributes the fractal dimension of the initial 100×100 window that contains it and the mean value of the intensity of the grayscale window image. Finally, to cluster each of the datasets constituted by such data points, we employed the Fuzzy c-Means clustering algorithm, for 3 and 4 clusters, respectively.

At Figures 4 and 5 we can see the clustering results with respect to the window size for all images where windows of the same cluster are colored equally for the 4 and 3 clusters case respectively. In the 4 cluster case the cluster colored in red corresponds to the severe pathology, the orange cluster corresponds to the mild pathology and green colored cluster corresponds to the normal lung. Finally the blue colored cluster denotes the background. In the 3 cluster case the mild pathology class is missing and the colors red, green and blue corresponds to the severe pathology, normal lung and background class respectively. The results are quite encouraging with respect to their medical content. The clustering algorithm seems to work pretty fine in distinguishing pathologic from normal. Thus the majority of the blocks that are clustered according to their fractal dimension into pathological clusters (i.e. severe and mild) belong into the regions indicated by the expert pathologists as fibrotic.

Tables 1 and 2 report the percentage, with respect to the window size for all images (retrieved after 7, 15 and 23 days), of each cluster found. The computed percentages coincide with manual annotation and scoring performed by our collaborating experts pathologists.

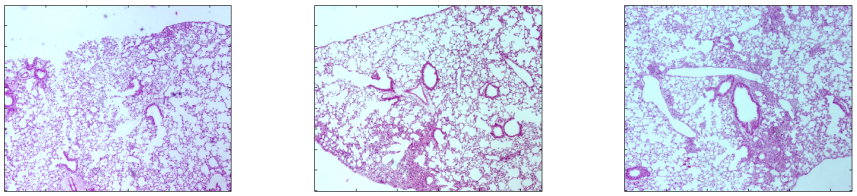

Fig. 3. Original IPF Microscopy Images retrieved after 7, 15 and 23 days respectively

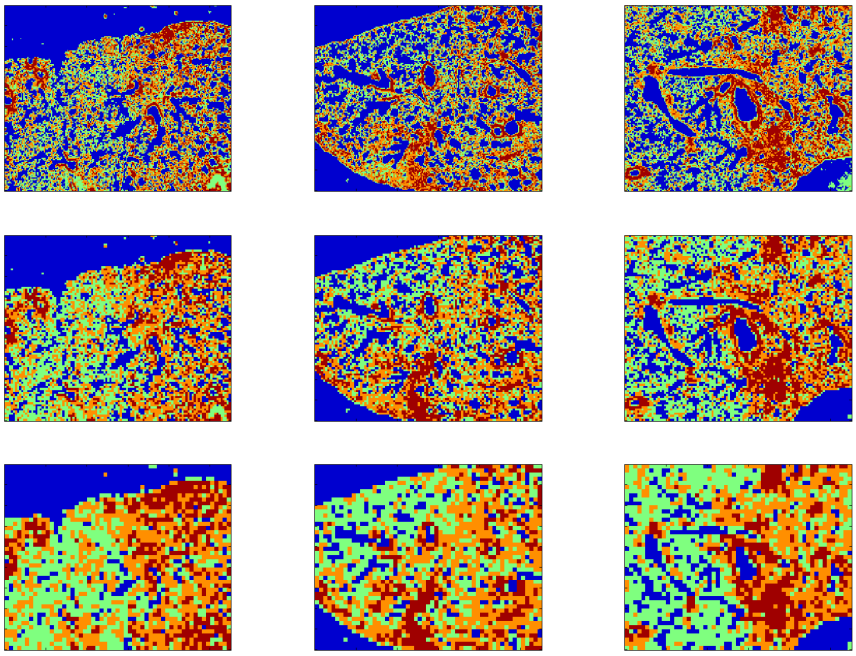

Fig. 4. Clustering result for the 4 clusters case with respect to the window sizes for the images retrieved after 7, 15 and 23 days respectively

Table 1. Percentages of the 4 clusters found by the clustering algorithm.

Window Size	Severe pathology	Mild pathology	Normal lung	Backround
	Image retrieved after 7 days			
5	10.89	21.59	23.19	44.31
10	12.59	26.86	26.72	33.81
20	16.55	31.55	28.48	23.40
	Image retrieved after 15 days			
5	11.49	21.99	22.00	44.50
10	14.17	27.01	25.16	33.64
20	14.52	33.33	30.56	21.58
	Image retrieved after 23 days			
5	15.01	24.74	25.50	34.74
10	16.83	27.19	29.33	26.64
20	16.25	29.95	35.82	17.96

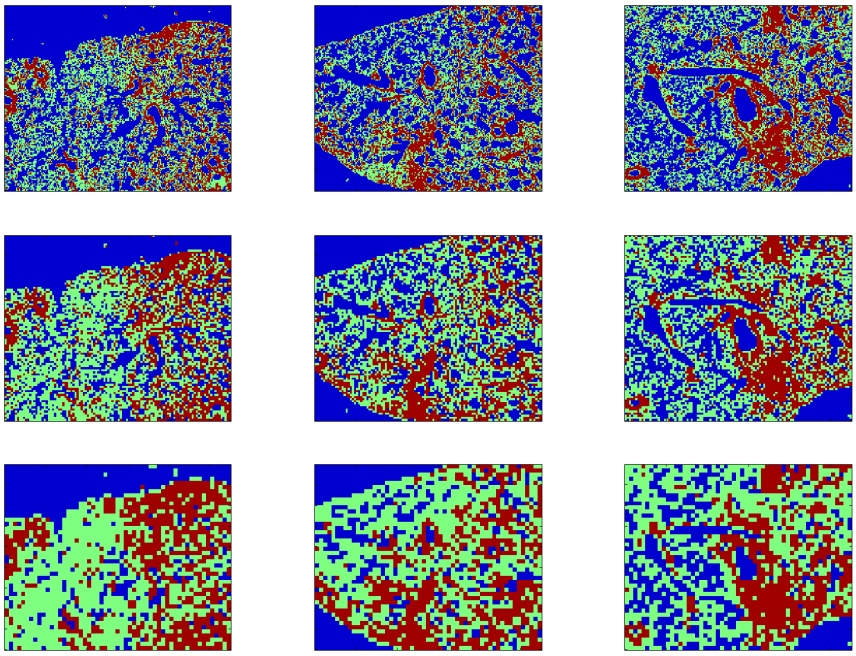

Fig. 5. Clustering result for the 3 clusters case with respect to the window sizes for the images retrieved after 7, 15 and 23 days respectively

Table 2. Percentages of the 3 clusters found by the clustering algorithm.

Window Size	Pathology	Normal lung	Backround
	Image retrieved after 7 days		
5	18.68	31.30	50.00
10	23.07	38.37	38.55
20	28.76	45.19	26.03
	Image retrieved after 15 days		
5	20.52	30.56	48.91
10	23.79	37.40	38.79
20	29.64	44.98	25.37
	Image retrieved after 23 days		
5	23.61	33.58	42.80
10	24.68	39.87	35.44
20	26.72	44.43	28.83

5 Conclusions

In the present paper, a computer-assisted image characterization tool based on a Fuzzy clustering method was introduced for the quantification of degree of

IPF in medical images. The implementation of this algorithmic strategy is very promising concerning the issue of the automated assessment of microscopic images of lung fibrotic regions. The results obtained so far show that the proposed strategy addresses the vital biological issues concerning their imaging part as this is contained in the specific type of microscopic images.

In a future study, we intent to use the degree of membership to postprocess the clustering results in an attempt to eliminate small isolated regions and increase the accuracy of the proposed approach.

Acknowledgments: The authors would like to thank the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: "Heracleitus II. Investing in knowledge society through the European Social Fund." for financially supporting this work.

References

1. Antoniou, K.M., Pataka, A., Bouros, D., Siafakas, N.M.: Pathogenetic pathways and novel pharmacotherapeutic targets in idiopathic pulmonary fibrosis. Pulmonary Pharmacology and Therapeutics 20(5), 453–461 (2007)
2. Bedossa P., Dargére D., P.V.: Sampling variability of liver fibrosis in chronic hepatitis c. Hepatology 38(6), 1449–57 (2003)
3. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers, Norwell, MA, USA (1981)
4. Caballero, T., Pérez-Milena, A., Masseroli, M., O'Valle, F., Salmerón, F., Del Moral, R., Sánchez-Salgado, G.: Liver fibrosis assessment with semiquantitative indexes and image analysis quantification in sustained-responder and non-responder interferon-treated patients with chronic hepatitis c. Journal of Hepatology 34(5), 740–747 (2001)
5. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., Prasher, D.C.: Green fluorescent protein as a marker for gene expression. Science (New York, N.Y.) 263(5148), 802–805 (Feb 1994)
6. Coultas, D.B., Zumwalt, R.E., Black, W.C., Sobonya, R.E.: The epidemiology of interstitial lung diseases. Am. J. Respir. Crit. Care Med. 150(4), 967–972 (1994)
7. Cox, E.: Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration. Elsevier Inc., USA (2005)
8. Foroutan-pour, K., Dutilleul, P., Smith, D.: Advances in the implementation of the box-counting method of fractal dimension estimation. Applied Mathematics and Computation 105(23), 195 – 210 (1999), http://www.sciencedirect.com/science/article/pii/S0096300398100966
9. Hussain, M., Merchant, S., Mombasawala, L., Puniyani, R.: A decrease in effective diameter of rat mesenteric venules due to leukocyte margination after a bolus injection of pentoxifylline digital image analysis of an intravital microscopic observation. Microvascular Research 67(3), 237–244 (2004), http://www.sciencedirect.com/science/article/pii/S0026286204000081

10. Iliadis, L., Vangeloudh, M., Spartalis, S.: An intelligent system employing an enhanced fuzzy c-means clustering model: Application in the case of forest fires. Computers and Electronics in Agriculture 70(2), 276 – 284 (2010), http://www.sciencedirect.com/science/article/pii/S0168169909001367, ¡ce:title¿Special issue on Information and Communication Technologies in Bio and Earth Sciences¡/ce:title¿
11. Inoué, S., Spring, K.: Video Microscopy. New York: Plenum (1997)
12. Izbicki, G., Segel, M., Christensen, T., Conner, M., Breuer, R.: Time course of bleomycin-induced lung fibrosis. Int J. Exp. Path 83(3), 111–9 (2002)
13. König, P., Kayser, C., Bonin, V., Würtz, R.: Efficient evaluation of serial sections by iterative gabor matching. Journal of Neuroscience Methods 111(2), 141–150 (2001), http://www.sciencedirect.com/science/article/pii/S0165027001004393
14. Maglogiannis, I., Sarimveis, H., Kiranoudis, C.T., Chatziioannou, A.A., Oikonomou, N., Aidinis, V.: Radial basis function neural networks classification for the recognition of idiopathic pulmonary fibrosis in microscopic images. IEEE Transactions on Information Technology in Biomedicine 12(1), 42–54 (2008)
15. Mandelbrot, B.: The Fractal Geometry of Nature. W.H. Freeman and Co., New York, USA (1983)
16. Masseroli, M., Caballero, T., O'Valle, F., Del-Moral, R., Pérez-Milena, A., Del-Moral, R.: Automatic quantification of liver fibrosis: design and validation of a new image analysis method: comparison with semi-quantitative indexes of fibrosis. Journal of Hepatology 32(3), 453–464 (2000)
17. Peitgen, H.O., Jürgens, H., Saupe, D.: Chaos and Fractals: New Frontiers of Science. Springer (Feb 1993), http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20\&path=ASIN/0387979034
18. Selman, M., Pardo, A.: Idiopathic pulmonary fibrosis: an epithelial/fibroblastic cross-talk disorder. Respiratory Research 3(1), 3 (2002), http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=64814&tool=pmcentrez&rendertype=abstract
19. Shiraishi, K., Takihara, H., Naito, K.: Quantitative analysis of testicular interstitial fibrosis after vasectomy in humans. Aktuelle Urologie 34(4), 262–264 (2003)
20. Shiraishi, K., Takihara, H., Naito, K.: Influence of interstitial fibrosis on spermatogenesis after vasectomy and vasovasostomy. Contraception 65(3), 245–249 (2002)
21. Shotton, D.: Image resolution and digital image processing in electronic light microscopy. Cell Biology, a Laboratory Handbook 3, 85–98 (1998)
22. Yagura, M., Murai, S., Kojima, H., Tokita, H., Kamitsukasa, H., Harada, H.: Changes of liver fibrosis in chronic hepatitis c patients with no response to interferon-alpha therapy: including quantitative assessment by a morphometric method. J Gastroenterol 35(2), 105–111 (2000), http://www.biomedsearch.com/nih/Changes-liver-fibrosis-in-chronic/10680665.html