Fractal Analysis of Mammograms

Fredrik Georgsson*, Stefan Jansson, and Christina Olsén

Department of Computing Science,
Umeå University, SE-901 87 Umeå, Sweden
`fredrikg@cs.umu.se`

Abstract. In this paper it is shown that there is a difference in local fractal dimension between imaged glandular tissue, supporting tissue and muscle tissue based on an assessment from a mammogram. By estimating the density difference at four different local dimensions (2.06, 2.33, 2.48, 2.70) from 142 mammograms we can define a measure and by using this measure we are able to distinguish between the tissue types. A ROC-analysis gives us an area under the curve-value of 0.9998 for separating glandular tissue from muscular tissue and 0.9405 for separating glandular tissue from supporting tissue. To some extent we can say that the measured difference in fractal properties is due to different fractal properties of the unprojected tissue. For example, to a large extent muscle tissue seems to have different fractal properties than glandular or supportive tissue. However, a large variance in the local dimension densities makes it difficult to make proper use of the proposed measure for segmentation purposes.

1 Introduction

It has been put forward that knowledge regarding properties of tissue can be assessed by estimating the fractal properties of an image, or more specifically an x-ray image, of the tissue [1,2]. The results have been hard to reproduce by other researchers [3,4,5] and the approach has been rightfully questioned.

We have, however, shown [6] that by estimating the box dimension (or any equivalent measure of fractal dimension) for a projected set, bounds for the dimension of the unprojected set can be imposed. From a theoretical point of view this tells us that information regarding the three dimensional object, e.g. the tissue composition of the breast, might be found by performing fractal analysis of an image, e.g. on a mammogram. This is of course encouraging, but it still remains to be shown if we are able to exploit this theoretical advance in practice.

The aims of this paper are to evaluate whether it is possible to use fractal analysis of x-ray images (mammograms) to classify them into tissue types and to draw conclusions regarding the fractal properties of the depicted object (the breast).

* Part of this research was funded by the Swedish Research Council under grant number 621-2002-3795.

In [3] an attempt to differentiate tumor from healthy tissue using different fractal properties was carried out. The conclusion was that fractal properties might be useful for segmentation within an image, but it is not possible to compare fractal properties between images. In [1] the authors match the fractal dimension to Wolfe-grades [7] thus enabling a predictability on the probability of breast cancer. In [2] fractal geometry was used to describe the pathology of tumors but the approach was questioned in [8]. It has been argued that fractal properties alone are not sufficient for effective texture segmentation, thus it is quite common to use fractal features as part of feature vectors in texture classification [9,10]. However, we aim at investigating the correspondence between the property of the projected image and the properties of the unprojected organ and pure texture approaches are of little interest for us. Our work differ from all of the above since we make use of knowledge of the imaging system when estimating fractal properties from an image in order to gain knowledge of the structures that generated the image.

The paper is organized as follows; In section 1.1 some mathematical background to fractal geometry is given. In section 1.2 a brief description of mammography and mammographic imaging is given. In section 2 the method is accounted for and in section 3 the data set used is described. The results are presented in section 4, the results are discussed in section 5 and conclusions are drawn in section 6.

1.1 Fractal Geometry and Some New Projective Properties of Fractals

Fractal geometry has been put forward as being suitable for analyzing irregular sets and one of the most used measure of fractality is the fractal dimension. In the literature it is not always clear what is meant by fractal dimension but the most frequent interpretation seems to be that the fractal dimension is identical to the Box-dimension as defined in equation (1)¹ We let $N_\varepsilon(F)$ be the smallest number of sets of a diameter smaller than ε that is needed to cover the set F . The Box-dimension is then defined as

$$\dim_B(F) = \lim_{\varepsilon \rightarrow 0} \frac{\log N_\varepsilon(F)}{-\log \varepsilon}. \quad (1)$$

Since we aim at applying an estimate of the Box-dimension on a digital (discrete) image we have to modify equation (1) slightly. The reason for this is, of course, that we cannot let $\varepsilon \rightarrow 0$ since $\varepsilon \geq \Delta$ where Δ is the pixel size. The solution is to rewrite equation (1) according to

$$\log N_\varepsilon(F) + \log \varepsilon \cdot \dim_B(F) + h(\varepsilon) = 0 \quad (2)$$

where $h(\varepsilon)$ is assumed to be approximately constant.

¹ From a mathematical point of view the Box-dimension is a simplification of the dimensionality concept, since the dimension should be based on a measure to be interesting. Examples of such dimensions are the Hausdorff and Packing dimensions.

Equation (2) shows that there is a linear relationship between $\log \varepsilon$ and $\log N_\varepsilon(F)$ where $\dim_B(F)$ is the constant of proportionality. Thus, the box dimension $\dim_B(F)$ is found as the slope of a straight line fitted to the points $\{(\log \varepsilon_i, \log N_{\varepsilon_i}(F))\}_{i=1}^N$, where $\varepsilon_1 < \varepsilon_2 < \dots < \varepsilon_N \leq w_1$. The fitting of the line to the points is done in a least square sense. The upper boundary w_1 is defined in section 2.

As an extension of work in [6] we have proved that

$$\dim_B(F) \leq \dim_B(\mathcal{F}) \leq \frac{12 \cdot \dim_B(F)}{12 - \dim_B(F)} \quad (3)$$

for almost all directions of projection. In equation (3), $F \subset \mathbb{R}^2 \times \mathbb{R}$ is the projection² of a set $\mathcal{F} \subset \mathbb{R}^3 \times \mathbb{R}$. In fact, we can show that it is not possible to find tighter bounds than the ones presented in equation (3). By using equation (3) we find bounds for the dimension of the unprojected set (for instance the breast) by estimating the Box-dimension of the projected set (for instance the mammo-gram). By using the upper and lower bounds from equation (3) it is possible to establish the required difference in box-dimensions for us to be able to say anything regarding the difference in the original set. This is illustrated in Fig. 1. Furthermore, by using equation (3) and an estimated dimension $\dim_B(F)$ we can find an interval of dimensions for which we cannot exclude the possibility that the fractal properties of the projected volumes do overlap. This interval is given by

$$I_F = \left[\max \left(0, \frac{12 \cdot \dim_B(F)}{12 + \dim_B(F)} \right), \min \left(3, \frac{12 \cdot \dim_B(F)}{12 - \dim_B(F)} \right) \right]. \quad (4)$$

By using I_F we can make assertions like: If $\dim_B(F_2) \notin I_{F_1}$ then $\dim_B(\mathcal{F}_2) \neq \dim_B(\mathcal{F}_1)$. The reverse is not true, i.e. if $\dim_B(F_2) \in I_{F_1}$ we cannot conclude that $\dim_B(\mathcal{F}_1) = \dim_B(\mathcal{F}_2)$.

There are many versions of the Box-dimension that differ in what type of covering sets are used and how they are allowed to be arranged etc. In the continuous case they all have identical mathematical properties [11] while this is not necessarily the case in the discrete case, a fact that many users of fractal geometry do not seem to be aware of.

We denote a method for estimating a fractal dimension as an estimator and in [12,13] we have evaluated the performance of a substantial number of estimators of intensity images. Based on these results we have picked the method most suitable for determining the local dimension. This was determined to be a method known as the probability method [14] with some modifications. The modifications consists mainly of performing the calculations in local neighbourhoods of a maximal size $w_1 \times w_1$. The estimated local dimensions are then averaged over a neighborhood of size $w_2 \times w_2$. A third parameter, *scale*, is the number of discrete intensity levels used in the calculations. That is, the original m levels of intensity are linearly mapped to *scale* levels.

² We have shown this for parallel, central and realistic x-ray projection.

Fig. 1. (a) A schematic illustration of the projective setting used in mammography. A volume $\mathcal{F}_i$ is projected down to an area F_i . (b) The relationship between an estimated fractal dimension $\dim_B(F)$ from a projection of a set and the upper boundary (dash-dotted line) and the lower boundary (dotted line) of the original, unprojected set $\dim_B(\mathcal{F})$ as given by equation (3).

1.2 Mammography

As mentioned earlier, we have used mammography as the domain for our study. A mammogram is an x-ray image of breast tissue and since the difference in contrast between tissue types (glandular tissue, fat tissue and supporting tissue) is very low the technical requirements imposed on a mammographic system are very hard.

The reasons for using mammograms for a study of this kind are two-fold. Firstly, a thorough fractal analysis requires rather high resolution and contrast and we have shown in [15] that of the more popular medical modalities, mammography is the most suitable one for fractal analysis. Since part of our aim is to evaluate the usefulness for fractal geometry as a method for segmentation it is important to know that the data we apply it to are as good as possible. Secondly, the proper method to use for the segmentation of mammograms into tissue types is an open research question and, furthermore, the relative proportion of dense tissue can be mapped to a probability of risk for developing tumours [7].

2 Method

Given an image $F : D \rightarrow \{0, \dots, m-1\}$ we use an estimator to calculate an image of the same size but the intensity at each spatial point corresponds to the local fractal dimension. We denote this image $F_{loc} : D \rightarrow [0, \dots, 3]$. In our data, see section 3, we have binary markings $M^k \subset D$ corresponding to an anatomical feature M given by expert k . Let $F_{loc}[M]$ denote the local dimensions at the

Fig. 2. Examples of a mammogram and its local dimensions. The calibration is only carried out once and the calibration surfaces can be dismissed after completing the calibration.

spatial coordinates given in M and let $H_{loc}[M]$ denote the histogram over these dimensions. The hypothesis is that the estimated distributions $H_{loc}[M^k]$ and $H_{loc}[N^k]$ differ if $M \neq N$. It is also of interest to see if there is a difference between the domain experts, that is if $H_{loc}[M^k]$ and $H_{loc}[M^l]$ if $k \neq l$.

Optimal values for the three parameters (w_1 , w_2 and *scale*) that control the performance of the estimator were found by a simple optimization procedure. We inserted patches into the mammogram with known fractal properties, Fig. 2(a). The four patches to the left of the mammogram were generated by approximating a fractal Brownian motion [16] and have from the top dimensions of 2.1, 2.3, 2.5 and 2.7 respectively. We also added patches with a plane (dimension 2), points (dimension 0) and lines (dimension 1) to the right of the mammogram. The points and lines were not used in the optimization of parameters.

By using the calibration areas we could determine what sets of parameters that yield a linear mapping between estimated dimension and true dimension. The initial parameter space contains three discrete parameters and it is possible to impose upper and lower boundaries for each parameter and thus we have a finite number of parameter combinations to try out. A coarse discretization of the parameter space allowed us to determine where in the space the optimal parameter combinations were situated. Within this area an exhaustive search was carried out yielding the chosen parameter values. The optimization of parameters was only performed on the calibration patches in Fig. 2. The actual mammogram did not take part in this process.

Next we calculated an F_{loc} version for every mammogram in our database. For each F_{loc} we extracted $H_{loc}[M^k]$, $\forall M, k$ and used these for further analyses.

Fig. 3. A distribution over local fractal dimensions ($H_{loc}[Gland]$, $H_{loc}[Pect]$, $H_{loc}[Supp]$). The peaks $dim_{loc} = 2.060 \pm 0.03$, $dim_{loc} = 2.335 \pm 0.03$, $dim_{loc} = 2.485 \pm 0.03$ and $dim_{loc} = 2.700 \pm 0.03$ have been marked.

3 Data

The mammograms used in this study are randomly selected from the MIAS-database (8 bit, 1024×1024 images). Three domain experts were asked to mark the pectoralis muscle, the mamilla, the breast border and the glandular disc on the images. In total we have used 142 different mammograms and we have $3 \cdot 142$ different markings of regions. In this paper "Gland" denotes the glandular tissue, "Pect" denotes the pectoral muscle and "Supp" denotes the supporting tissue (including fat) that surrounds the glandular tissue. Each image in the MIAS-database is adjusted to be of square size, and since a mammogram is rectangular we always have empty black areas to the left and right of the actual mammogram. In these areas we inserted the calibration patches described in section 2. Thus we did not alter any mammographic information when adding the patches. The mammograms were all adjusted so that the pectoral muscle is to the left in the image, but no other preprocessing took place.

4 Results

For $D = \{1, \dots, 1024\} \times \{1, \dots, 1024\}$ images with intensity levels $L = \{0, \dots, 255\}$ the optimal set of parameters was found to be $w_1 = 47$, $w_2 = 5$ and $scale = 51$. Using these parameters an F_{loc} was calculated for each image in our database, and $H_{loc}[M^k]$ were extracted from each F_{loc} .

The accumulated distributions $H_{loc}[M]$ are shown in Fig. 3 (based on 142 images) and here we clearly see that there is a difference in the relative shape

Fig. 4. Distributions over ρ -values for glandular tissue, muscle tissue and supportive tissue as calculated from 142 images. Optimal segmentation is obtained by thresholding at $\rho_1 = -0.174$ and $\rho_2 = 0.0218$ yielding a lowest combined probability of misclassification $P(\text{err}|\text{Gland}) = 0.248$, $P(\text{err}|\text{Pect}) = 0.0213$ and $P(\text{err}|\text{Supp}) = 0.0423$.

of the distributions, specially between the pectoral muscle vs. glandular and supportive tissue. The pectoral muscle has clear peaks at $\text{dim}_{loc} = 2.060$ and $\text{dim}_{loc} = 2.485$ whilst glandular and supportive tissue have peaks at $\text{dim}_{loc} = 2.335$ and $\text{dim}_{loc} = 2.700$. The peaks were estimated to have a width of $2\delta = 0.06$. We define a measure

$$\varphi_\ell(x) = \frac{\sum_{|\ell-x|<\delta} H_{loc}[\cdot](x)}{\sum_x H_{loc}[\cdot](x)}$$

and calculate

$$\boldsymbol{\rho} = [\varphi_{2.060}(x) \quad \varphi_{2.335}(x) \quad \varphi_{2.485}(x) \quad \varphi_{2.700}(x)]^T. \quad (5)$$

The vector $\boldsymbol{\rho}$ describes the relative density in the histogram at each of the identified peaks. Since we have observed that glandular tissue has peaks at the second ($\text{dim}_{loc} = 2.485$) and fourth ($\text{dim}_{loc} = 2.700$) values we project the vector $\boldsymbol{\rho}$ onto the base vector $\mathbf{b} = [-1 \quad 1 \quad -1 \quad 1]^T$ with the inner product

$$\rho = \mathbf{b}^T \boldsymbol{\rho}. \quad (6)$$

It is easily shown that the value of ρ will be in the -1 to 1 interval, where a value of 1 corresponds to a histogram with all observed data at $\text{dim}_{loc} = 2.485$ and/or $\text{dim}_{loc} = 2.700$. A value of -1 corresponds to a histogram with all observed data at $\text{dim}_{loc} = 2.060$ and/or $\text{dim}_{loc} = 2.485$.

By calculating ρ for all 142 images we are able to estimate a distribution $p(\rho)$ for different tissue types. In Fig. 4 we can see $p(\rho)$ given the three tissue

Fig. 5. Receiver operating characteristic (ROC) curve for using the measure to differentiate tissue types when using ground truth from a specific domain expert (F)

types. The optimal, in the sense least probability for misclassification, thresholds are $\rho_1 = -0.174$ and $\rho_2 = 0.0218$. The thresholds were found by a simple linear search of ρ (exhaustive). The discriminating power obtainable by using equations (5) and (6) is illustrated in Fig. 5 where we can see receiver operating characteristic for the tissue types. The performance differs slightly depending on what expert we used for the ground truth markings, see Table 1. The best A_z values (area under the ROC-curve) obtained for separating tissue types were 0.9998 for glandular vs. pectoralis, 0.9915 for supporting tissue vs. pectoralis and 0.9405 for glandular tissue vs. supporting tissue.

5 Discussion

Based on the ROC-curves from the $p(\rho)$ data (Fig. 5) we conclude that by using the local fractal dimension it is possible to separate tissue types. The ROC analyses are based on accumulated data from all 142 images but to be useful for segmentation, we have to use an estimate $\widehat{H_{loc}}[\cdot]$ from a neighbourhood in F_{loc} . It turns out that the variance is quite substantial which makes it rather difficult use the local density distribution. Thus equations (5) and (6) are not necessarily useful for segmentation purposed.

We have observed that there is a significant peak at about dimension 2, see Fig. 3. To some extent this can be explained by over and under exposure of the images. Similarly it seems that fractal analysis is very good at detecting film scratches and other artifacts. Both these observations are interesting, but since mammography as a modality slowly is going digital, where problems with

Table 1. Intra variance amongst domain experts regarding the area under the ROC-curve

Expert	Area under ROC-curve		
	Gland. vs. Pect	Gland vs. Supp	Supp vs. Pect
A	0.9995	0.8463	0.9994
C	0.9983	0.8874	0.9882
F	0.9998	0.9405	0.9915

Fig. 6. Enlarged part of the normalised histograms for $H_{loc}[Gland^k]$, $H_{loc}[Pect^k]$ and $H_{loc}[Supp^k]$, $k \in \{A, C, F\}$. The observable intra variance is very low for the pectoral muscle and the highest intra variance is found in the supportive tissue.

exposure and film scratches are significantly smaller, it is not deemed interesting to investigate this further.

By making use of equation (3) we can say that the structures yielding local dimensions of $dim_{loc} = 2.060$ and $dim_{loc} = 2.485$ must have different dimensions in three dimensions. We cannot say the same for $dim_{loc} = 2.335$ and $dim_{loc} = 2.700$. Thus we conclude that muscle tissue has slightly different fractal properties than glandular tissue and supportive tissue. The difference could to some extent be explained by the fact that the muscle quite often is under exposed in the images.

From Fig. 4 it appears that supporting tissue and glandular tissue have more in common than muscle tissue and any other two tissue types. This might be explained by the fact that glandular tissue and supportive tissue to a larger extent are projected on top of each other. That is, within each projective cone, see Fig. 1 (a) there are both tissue types.

In Fig. 6 we can observe the intra variance amongst the domain experts. Although the observed difference is relatively low it will give rise to a significant difference in the area under the ROC curve values, see Table 1.

6 Conclusion

We conclude that there is a difference in local fractal dimension estimated by the box dimension from a mammogram. To part, we can say that this difference is due to difference in the dimension of the three dimensional unprojected tissue. However, the high variance of the local density function of the local dimension makes it difficult to make use of this difference in tissue segmentation.

References

1. Caldwell, C.B., Stapleton, S.J., Holdsworth, D.W., Jong, R.A., Weiser, W.J., Cooke, G., Yaffe, M.J.: Characterization of mammographic parenchymal pattern by fractal dimensions. *Physics in Medicine and Biology* 35, 235–247 (1990)
2. Baish, J.W., Jain, R.: Fractals and cancer. *Cancer Research* 60, 3683–3688 (2000)
3. Veenland, J., Grashuis, J.L., van der Meer, F., Beckers, A.L.D., Gelsema, E.S.: Estimation of fractal dimension in radiographs. *Medical Physics* 23, 585–594 (1996)
4. Huang, Q., Lorch, J., Dubes, R.: Can the fractal dimension of images be measured. *Pattern Recognition* 27, 339–349 (1994)
5. Jelinek, H., Fernandez, E.: Neurons and fractals: how reliable and useful are calculations of fractal dimensions? *Journal of Neuroscience Methods* 81, 9–18 (1998)
6. Nilsson, A., Georgsson, F.: Projective properties of fractal sets. *Chaos, Solitons and Fractals*, Accepted for publication (2006)
7. Wolfe, J.N.: Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer* 37, 2486–2492 (1976)
8. Chung, H.W., Chung, H.J.: Letter to the editor; correspondence re: J. W. Baish and R. K. Jain, *Fractals and Cancer*. *Cancer Res.* 61, 8347–8348 (2001)
9. Lee, W.L., Chen, Y.C., Chen, Y.C., Hsieh, K.S.: Unsupervised segmentation of ultrasonic liver images by multiresolution fractal feature vectors. *Information Sciences* 175, 177–199 (2005)
10. Chen, C., Daponte, J., Fox, M.: Fractal feature analysis and classification in medical imaging. *IEEE Transactions on Medical Imaging* 8, pp. 133–142 (1989) nr: 30
11. Falconer, K.: *Fractal Geometry*. Wiley, Chichester (1990)
12. Jansson, S., Georgsson, F.: Evaluation of methods for estimating the fractal dimension of intensity images. In: Georgsson, F., Börjén, N., (eds.): *Proceedings of SSBA. Swedish Society for Automated Image Analysis*, Number UMINF-06.11 in ISSN 0348-0542, pp. 69–72 (2006)
13. Jansson, S.: Evaluation of methods for estimating fractal properties of intensity images. Umnad: 699/06, Dept. of Comp. Sc. Umeå University (2006)
14. Keller, J., Chen, S., Crownover, R.: Texture description and segmentation through fractal geometry. *Computer Vision, Graphics, and image processing* 45, 150–166 (1989)
15. Georgsson, F.: Fractal dimensions and the geometry of x-ray imaging. In: Bengtsson, E., (ed.): *Proceedings of SSBA 2004*, pp. 150–153 (2004)
16. Peitgen, H.O., Saupe, D. (eds.): *The Science of Fractal Images*. Springer, Heidelberg (1988)