Multifractal feature descriptor for histopathology

Chamidu Atupelage\textsuperscript{a,}, Hiroshi Nagahashi\textsuperscript{b}, Masahiro Yamaguchi\textsuperscript{c}, Michie Sakamoto\textsuperscript{d} and Akinori Hashiguchi\textsuperscript{d}

\textsuperscript{a}Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Tokyo, Japan
\textsuperscript{b}Imaging Science and Engineering Laboratory, Tokyo Institute of Technology, Tokyo, Japan
\textsuperscript{c}Global Scientific Information and Computing Center, Tokyo Institute of Technology, Tokyo, Japan
\textsuperscript{d}Department of Pathology, School of Medicine, Keio University, Keio, Japan

Abstract. Background: Histologic image analysis plays an important role in cancer diagnosis. It describes the structure of the body tissues and abnormal structure gives the suspicion of the cancer or some other diseases. Observing the structural changes of these chaotic textures from the human eye is challenging process. However, the challenge can be defeat by forming mathematical descriptor to represent the histologic texture and classify the structural changes via a sophisticated computational method.

Objective: In this paper, we propose a texture descriptor to observe the histologic texture into highly discriminative feature space.

Method: Fractal dimension describes the self-similar structures in different and more accurate manner than topological dimension. Further, the fractal phenomenon has been extended to natural structures (images) as multifractal dimension. We exploited the multifractal analysis to represent the histologic texture, which derive more discriminative feature space for classification.

Results: We utilized a set of histologic images (belongs to liver and prostate specimens) to assess the discriminative power of the multifractal features. The experiment was organized to classify the given histologic texture as cancer and non-cancer. The results show the discrimination capability of multifractal features by achieving approximately 95% of correct classification rate.

Conclusion: Multifractal features are more effective to describe the histologic texture. The proposed feature descriptor showed high classification rate for both liver and prostate data sample datasets.

Keywords: Histologic images, fractal, multifractal, texture classification

1. Introduction

Histology is the study of the structures of the body tissues and widely used technique of identifying cancers of different body organs. Diagnosing histologic image is a challenging task and the decisions are subject to human error. Figure 1 gives evidence: how difficult to distinguish cancer and non-cancer texture of histologic images. However, Computer Aided Diagnosis (CAD) systems can be utilized to characterize histologic images by modeling the complex texture patterns into an extensive statistical model. This paper proposes a new method to discriminate the histologic images, which transforms the histologic texture into more discriminative feature space by utilizing self-similarity properties of the texture. Fractal dimension is the measure, which appropriately describes the self-similarity of the self-similar structures. However, natural images are not fractals itself, instead we can find statistically distributed self-similar patterns in different scales, and these structures are called random fractals. Multifractal dimension is the measure introduced for random structures or natural objects. Therefore, we utilized the multifractal computation to transform the histologic texture from pixel domain to multifractal domain. In multifractal domain, we can obtain detail information of the textures, which is not observable in the pixel domain. This multifractal information is employed to construct the feature space and
be used in texture classification problem. We utilized the proposed feature descriptor to classify cancer and non-cancer regions of a set of annotated histologic data samples obtained from prostate and liver specimens.

2. Methodology

A fractal is a geometrical shape, where the reduced copy of the original structure shows an approximately similar behavior as its origin [1]. This property is called self-similarity. The statistical quantity that we used to compute the visibility of the self-similarity is called fractal dimension.

Consider a bounded set $S$ in Euclidean $n$-space and the structure is covered by a countable collection of subsets of diameter $d_i$. Define the measure of the set $S$ as,

$$H_D^D(S) = \inf_{D>0} \sum d_i^D$$

(1)

where $D > 0$ is a real quantity. The limit of $H_D^D$ (when $\epsilon$ tends to zero) is the Hausdorff measure of the set $S$. The value of $D$ where $H_D^D$ varies from $\infty$ to 0 is subsequently called the Hausdorff dimension or fractal dimension [2].

Several computation algorithms have been proposed to measure the fractal dimension of the given structure.
in literature. Among them, Box Counting method is popular and widely used method for fractal computation [3]. The Box Counting method: cover a structure $\Sigma$ with a grid of $s$-dimensional boxes with side length $\epsilon$ and count the number of non-empty boxes $N(\epsilon)$. When $\epsilon$ tends to zero, the limiting value of $N(\epsilon)$ is proportional to $\epsilon^{-D_h}$. Therefore, $D_h$ is the box counting dimension of a given fractal, and it is defined as,
\[
D_h = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log \epsilon} \tag{2}
\]

On the other hand, the natural structures (images) are not self-similar. However, we can find statistically distributed self-similarity over the natural structures in different scales. The fractal perspective that we use to characterize random fractals is known as multifractal [4].

The usual way of characterizing the multifractal property of a structure is: first, find a non-integer exponent $\alpha$ known as Hölder exponent, which describes the local singularity of the object [5]. Let the structure $\Sigma$ be divided into non-overlapping boxes $S_i$ of size $\epsilon$ such that $\Sigma = \bigcup S_i$. Each box $S_i$ is characterized by some local singularity measure $\mu(S_i)$. The coarse Hölder exponent of each subset $S_i$ is defined by,
\[
a_i = \lim_{\epsilon \to 0} \frac{\log (\mu(S_i))}{\log (\epsilon)} \tag{3}
\]

Consequently, the second step is finding the frequency distribution $f(\alpha)$ of the exponent measure $\alpha$, which is known as multifractal spectrum. This function describes the global regularity of the observed structure and it can be defined as,
\[
f(\alpha) = \lim_{\epsilon \to 0} \frac{\log N_i(\alpha_\epsilon))}{\log (\epsilon)} \tag{4}
\]
where $N_i(\alpha_\epsilon)$ is the number of boxes $S_i$ containing particular values of $\alpha_i$.

In this work, we extract texture characteristics in the multifractal domain as $a$ and $F(a)$, which correspond to each pixels of a given image. Our early work has shown that different local singularity characteristic measures ($\mu(S_i)$) can be utilized to compute multifractal features of a given image [6]. We have empirically defined six local singularity characteristic measures to construct the multifractal feature descriptor. The formulations of these measures are tabulated in Table 1.

![Image](image1.png)

**Table 1**

| Measure Type | Formula |
|-------------|---------|
| Maximum $\mu(m, n)$ | $\max_{(i, l)} g(x, l)$ |
| Minimum $\mu(m, n)$ | $\min_{(i, l)} g(x, l)$ |
| Sum $\mu(m, n)$ | $\sum_{(i, l)} g(x, l)$ |
| Modified box counting $\mu(m, n)$ | $\left( \max_{(i, l)} g(x, l) - \min_{(i, l)} g(x, l) \right) / l$ |
| Gradient magnitude $\mu(m, n)$ | $\sqrt{\nabla x \cdot \nabla y (x, n) \cdot \nabla x \cdot \nabla y (x, n)}$ |
| Gradient angle $\mu(m, n)$ | $\arctan \left\{ \frac{\nabla x \cdot \nabla y (x, n)}{\nabla x \cdot \nabla y (x, n)} \right\}$ |

Where $\mu$ is an exponent measure, $\epsilon$ indicates a box size, whose center is located at observed pixel $(m, n)$, $\Omega$ represents all pixels $(k, l)$ belongs to the box, $\Omega'$ is a set of non-zero pixels within a particular box, $g(x, l)$ is a grayscale intensity at point $(k, l)$, $\nabla$ and $\nabla$ are gradient operators towards $x$ and $y$ directions, $l$ is the standard deviation of Gauss kernel.

### 3. Experiment and results

We utilized a set of hematoxylin and eosin (H&E) stained biopsy specimens of several liver and prostate cancer patients to obtain the sample dataset. These data are scanned (20x magnification) and visualized using a tool called Nano-Zoomer [7] (the tool visualizes each data sample in 8-different magnifications: 1.25x, 2.5x, 5.0x, 10x, 20x, 40x, 63x, and 100x) and experienced pathologists annotated the cancer regions of each of the data samples. Based on this annotated data sample, for each organ (liver and prostate) we constructed 24 categories of experimental datasets (three sizes of image patches: 64×64, 128×128, 256×256 pixels with eight magnifications). Each category consists around 300–400 sample patches. Using the proposed feature description method, we constructed the multifractal features space for each sample patch. That feature space is utilized with typical texton based classification method with Support Vector Machine (SVM). We computed correct classification rate for each data category utilizing n-fold cross validation to minimize the over-fitting of the classifier.

Table 2 shows the correct classification rates for all sample data categories. For both liver and prostate data the larger patch size (256×256 pixels) shows the highest classification rates. Also we found that the images captured at larger magnifications (over 20.00x)
Table 2
Classification rates (%) of liver and prostate data sample

| Mag. scale | Patch size: 64 × 64 pixels | Patch size: 128 × 128 pixels | Patch size: 256 × 256 pixels |
|------------|----------------------------|-------------------------------|----------------------------|
|            | Liver | Prostate | Liver | Prostate | Liver | Prostate | Liver | Prostate |
| 1.25       | 89.0  | 89.0     | 85.8  | 85.8     | 95.3  | 95.3     | 94.8  | 94.8     |
| 2.50       | 90.3  | 90.3     | 85.5  | 85.5     | 94.3  | 94.3     | 95.5  | 95.5     |
| 5.00       | 90.5  | 90.5     | 89.3  | 89.3     | 96.5  | 96.5     | 96.3  | 96.3     |
| 10.00      | 82.8  | 82.8     | 91.8  | 91.8     | 93.0  | 93.0     | 96.8  | 96.8     |
| 20.00      | 62.5  | 62.5     | 88.5  | 88.5     | 77.5  | 77.5     | 91.5  | 91.5     |
| 40.00      | 66.3  | 66.3     | 83.0  | 83.0     | 73.3  | 73.3     | 83.0  | 83.0     |
| 63.00      | 55.5  | 55.5     | 73.8  | 73.8     | 66.5  | 66.5     | 73.8  | 73.8     |
| 100.00     | 53.0  | 53.0     | 67.3  | 67.3     | 59.8  | 59.8     | 67.3  | 67.3     |

comprise some visual distortions, because Nano-Zoomer generates the magnified images by interpolating the original data, which captured in 20.00x. As a clue, we found very low classification rates for larger magnification in Table 2. Furthermore pathologists are also infrequently used larger magnification for their diagnosis process, thus we put less consideration for classification results corresponds to larger magnifications. At the bottom line, we can summarize that the proposed multifractal feature descriptor is appropriate for lower magnifications (undistorted images), showing around 95% of classification rate for both liver and prostate data samples.

4. Conclusion

In this paper, we proposed a new multifractal feature descriptor to characterize histologic images. The merit of the proposed work is assessed by a classification experiment utilizing with histologic dataset obtained from liver and prostate cancer patients. Experimental results indicated that multifractal features characterize the histologic image approximately 95% of classification rate. It gives an evident for the significance utilizing of multifractal analysis into prostate image classification.

Acknowledgments

This research is supported by New Energy and Industrial Technology Development Organization (NEDO) under the Research and Development Project for pathological image recognition.

References

[1] B.B. Chaudhuri and N. Sarkar, “Texture segmentation using fractal dimension,” IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (1995), 72–77.
[2] R. Andrews, M. Blackledge and J. Turner, Fractal Geometry in Digital Imaging Elsevier Science, 1998.
[3] Andrzej Z Górski, “Pseudo-fractals and the box counting algorithm,” 34 (2001), 39.
[4] P. Mignot and J. Levy Vehel, “Multifractal Segmentation of images,” in Fractals 2 (1994), 379–382.
[5] P. Mignot and J. Levy Vehel, “Multifractal Segmentation of images,” (1994), 379–382.
[6] Chamidu Atupelage, Hiroshi Nagahashi, Masahiro Yamaguchi, Michie Sukamoto and Akihiti Haseguchi “Multifractal Feature Based Cancer Detection for Pathological Images,” in The 5th International Conference on Bioinformatics and Biomedical Engineering, Wuhu, China 2011.
[7] NanoZoomer 2.0-HT. [Online]. http://jp.hamamatsu.com/ products/life-science/pd26751013/nanozoom/index.jsp.html.