TEXTURE ANALYSIS OF MAMMOGRAMS USING THE TWO-DIMENSIONAL HURST OPERATOR

Ramachandran Chandrasekhar
Biomedical Engineering Department, Singapore General Hospital, Singapore 169608, Singapore
e-mail: gbercs@sgh.gov.sg

Yianni Attikiouzel and Christopher J S deSilva
Centre for Intelligent Information Processing Systems, Department of Electrical and Electronic Engineering
The University of Western Australia, Nedlands, WA 6907, Australia
e-mail: yianni@ee.uwa.edu.au and chris@ee.uwa.edu.au

Abstract A new texture operator, the two-dimensional Hurst operator, $\mathcal{H}$, is introduced in this paper. It is the two-dimensional extrapolation of an operator originally devised by H. E. Hurst to analyze time-series. If an image is visualized as a three-dimensional relief map, and the intensity taken to be the height above an x-y plane, this texture operator is useful in determining the relative jaggedness of different image regions. Initial results with mammograms reveal that this operator can discriminate between the predominantly smooth "fatty" and relatively rougher "fibroglandular" regions on mammograms, and has the potential for lesion detection.
Keywords: image processing, texture analysis, texture segmentation, Hurst operator, mammogram, fractal.

1. INTRODUCTION

Mammograms, or X-ray images of the breast, are currently considered to be the best method for the early detection of breast cancer in the asymptomatic female population. Interest in the computer analysis of mammograms is driven by the assumption that such analysis could assist in alleviating the workload faced by radiologists reading mammograms from population screening.

The properties of the different types of breast tissue and the superimposition of these tissues to produce a mammogram result in a fuzzy or cloudy appearance typical of a disordered texture. Texture analysis, especially using statistical or fractal measures [1], therefore seems a promising tool for analyzing mammograms. Several investigators [2, 3] have used established texture measures, principally for mammographic lesion detection or characterization. This paper presents a new texture measure, the *two-dimensional Hurst operator*, that has been found useful in the texture segmentation of mammograms.

2. THE TWO-DIMENSIONAL HURST OPERATOR

The two-dimensional Hurst operator, $\mathcal{H}$, defined below, is the two-dimensional extension of an operator originally devised by H. E. Hurst [4, 5] for analyzing the flow of water through the river Nile over long periods of time. Hurst's work subsequently received attention from Mandelbrot and co-workers [6, 7] who explained it on the basis of the fractal paradigm. The original one-dimensional formulation is available in the above references and will not be covered here.

A *digital image* is considered to be a two-dimensional real-valued data set $\zeta(x, y)$ where (x, y) is a point on a two-dimensional lattice. The number of data points in the x and y directions are n_x and n_y respectively. A *tile* is a closed, connected subset of ζ. Here, we shall deal only with *square tiles* T_t which are square regions of the image with side t, i.e., there are t data points in each of the x and y directions of the tile and t^2 in all. This is illustrated in Figure 1.

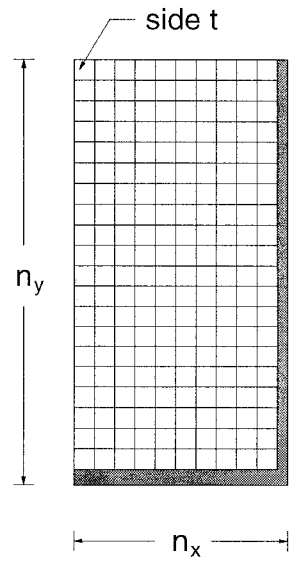

Fig. 1: Tiling of the image by a rectangular mosaic of square tiles. The image size is n_x by n_y as shown. The tiles are square of side t. The number of tiles is floor(n_x/t) and floor(n_y/t) in the x and y directions respectively. The shaded portion of the image is not covered by the tiles and therefore not subjected to the two-dimensional Hurst operator.

0-7803-4137-6/97/$10.00©1977 IEEE

The image data is deliberately restricted to an *analysis window* or *lag* which is identified with the radius of a two-dimensional *mask* that is associated with the operator. An *octagonal mask* of lag ρ equal to 4 is illustrated in Figure 2 and the respective Euclidean distances r_k from the central pixel are shown in Table 1. The octagonal mask was chosen because it approximates a circle, which is the ball of radius ρ in two-dimensions.

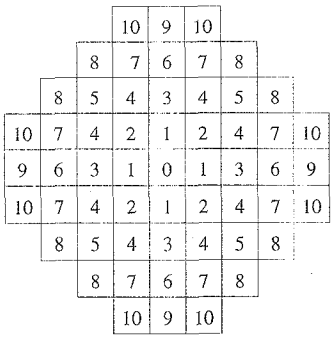

Fig. 2: Octagonal mask of "radius" 4 and diameter 9. Pixels at the same Euclidean distance from the central pixel, numbered 0 above, are labelled with the same index. See Table 1 for the relevant distances.

Pixel Label k	No. of Pixels	Euclidean Distance r_k
0	1	0
1	4	1
2	4	$\sqrt{2}$
3	4	2
4	8	$\sqrt{5}$
5	4	$2\sqrt{2}$
6	4	3
7	8	$\sqrt{10}$
8	8	$\sqrt{13}$
9	4	4
10	8	$\sqrt{17}$

Table: 1: The Euclidean distances corresponding to the indices shown on the octagonal mask in Figure 2.

Table 2 gives an explanation of the symbols used in defining the Hurst operator. The mask of radius ρ is centred at the pixel with co-ordinates $\mathbf{p} = (a, b)$. Let the set of pixels that lie within the mask be called M_ρ. The position of the general pixel within the mask is $\mathbf{q} = (i, j) \in M_\rho$ and its value is $\zeta(\mathbf{q})$. Then,

$$N_\rho = \sum_{\mathbf{q} \in M_\rho} \sum (1) \tag{1}$$

$$\mu(\rho, \mathbf{p}) = \frac{1}{N_\rho} \sum_{\mathbf{q} \in M_\rho} \sum \zeta(\mathbf{q}) \tag{2}$$

$$\sigma(\rho, \mathbf{p}) = \left[\frac{1}{N_\rho} \sum_{\mathbf{q} \in M_\rho} \sum [\zeta(\mathbf{q}) - \mu(\rho, \mathbf{p})]^2 \right]^{\frac{1}{2}} \tag{3}$$

$$C(r, \rho, \mathbf{p}) = \sum_{\substack{\mathbf{q} \in M_\rho \\ 1 \leqslant r \leqslant \rho \\ \|\mathbf{q}-\mathbf{p}\| \leqslant r}} \sum [\zeta(\mathbf{q}) - \mu(\rho, \mathbf{p})] \tag{4}$$

$$R(\rho, \mathbf{p}) = \max_{1 \leqslant r \leqslant \rho} [C(r, \rho, \mathbf{p})] - \min_{1 \leqslant r \leqslant \rho} [C(r, \rho, \mathbf{p})] \tag{5}$$

Masks of *different* lags are superimposed on the central pixel, $\mathbf{p}$, and the means are computed for these different lags. These means are called $\mu(\rho, \mathbf{p})$ to show their dependence on ρ as well as $\mathbf{p}$. The number of pixels associated with any single normed distance r from the central pixel could be 4, 8 or some other number as shown in Table 1. These equidistant pixels are *all* treated as *one data point* corresponding to the distance r_k when they are accumulated. The mean is subtracted from *each pixel* before this accumulation is performed.

The equations (1) to (5) refer to a *single* central pixel within a square *tile* T_t of side t. The values for $\mu(\rho, \mathbf{p})$ to $R(\rho, \mathbf{p})$ are then computed by taking *each* pixel within the tile in turn to be the central pixel. The *mean* of all these values is then given by

$$\begin{aligned} \left\langle \frac{R(\rho, \mathbf{p})}{\sigma(\rho, \mathbf{p})} \right\rangle &= \frac{1}{N_t} \sum_{\mathbf{p} \in T_t} \sum \frac{R(\rho, \mathbf{p})}{\sigma(\rho, \mathbf{p})} \\ &\triangleq \left\langle \frac{R(\rho)}{\sigma(\rho)} \right\rangle \end{aligned} \tag{6}$$

where

$$N_t = \sum_{\mathbf{p} \in T_t} \sum (1) \tag{7}$$

A straight line is fitted to the data pairs comprising $(\log \rho, \log\langle R(\rho)/\sigma(\rho)\rangle)$, for *masks of different lags*, varying from $\rho = \rho_1$ to $\rho = \rho_n$, $\rho_n > \rho_1$. Therefore, there are n *distinct* lags between ρ_1 and ρ_n at which the computations are performed. Stated another way, there are n ordered pairs of data or n is the cardinality of the set of lags from ρ_1 to ρ_n used in the computations.

In the above and subsequent equations, log stands for $\log_{10}$. The straight line fit is performed to minimize the square of the error between the line and *all* data points. A total of *three* parameters result from this procedure:

1. the gradient, m;
2. the vertical intercept c;
3. the square of the correlation coefficient, η^2;

PARAMETER	SYMBOL	MEANING
Record Size	T_t	square tile of side t containing t^2 pixels
Random Variable or Observation	$\zeta(i,j) = \zeta(\mathbf{q})$	pixel value at position $\mathbf{q} = (i,j)$ in the digital Cartesian grid
Length of Analysis Window or Lag	ρ	radius of mask
Datum	$\mathbf{p}$	position of central mask pixel $\mathbf{p} = (a,b)$ in digital Cartesian grid
Mean	$\mu(\rho, \mathbf{p})$	mean of ζ over mask of radius ρ centred on $\mathbf{p}$
Standard Deviation	$\sigma(\rho, \mathbf{p})$	standard deviation of ζ over mask of radius ρ centred on $\mathbf{p}$
Cumulative Deviation from the Mean	$C(r,\rho,\mathbf{p}) = \sum_j \sum_i [\zeta(i,j) - \mu(\rho,\mathbf{p})]$ for $1 \leqslant r \leqslant \rho$ and $\|\mathbf{q}-\mathbf{p}\| \leqslant r$	accumulated sum of the difference between each observation and the mean
Range	$R(\rho,\mathbf{p}) = \max[C(r,\rho,\mathbf{p})] - \min[C(r,\rho,\mathbf{p})]$ for $1 \leqslant r \leqslant \rho$	difference between maximum and minimum of $C(r,\rho,\mathbf{p})$

Table: 2: Explanation of symbols used in defining the Hurst operator.

The equations below state explicitly how the m, c and η^2 values arise from the application of the two-dimensional Hurst operator to the tile T_t:

$$\langle u \rangle = \frac{1}{n} \sum_{\rho=\rho_1}^{\rho_n} \log \rho \tag{8}$$

$$\langle v \rangle = \frac{1}{n} \sum_{\rho=\rho_1}^{\rho_n} \log \left\langle \frac{R(\rho)}{\sigma(\rho)} \right\rangle \tag{9}$$

$$s_{uu} = \frac{1}{n} \sum_{\rho=\rho_1}^{\rho_n} (\log \rho)^2 - \langle u \rangle^2 \tag{10}$$

$$s_{vv} = \frac{1}{n} \sum_{\rho=\rho_1}^{\rho_n} \left[\log \left\langle \frac{R(\rho)}{\sigma(\rho)} \right\rangle \right]^2 - \langle v \rangle^2 \tag{11}$$

$$s_{uv} = \frac{1}{n} \sum_{\rho=\rho_1}^{\rho_n} \left[(\log \rho) \log \left\langle \frac{R(\rho)}{\sigma(\rho)} \right\rangle \right] - \langle u \rangle \langle v \rangle \tag{12}$$

$$m \triangleq \frac{s_{uv}}{s_{uu}} \tag{13}$$

$$c \triangleq \langle v \rangle - \frac{s_{uv}}{s_{uu}} \langle u \rangle = \langle v \rangle - m \langle u \rangle \tag{14}$$

$$\eta^2 \triangleq \frac{s_{uv}^2}{s_{uu} s_{vv}} = m \frac{s_{uv}}{s_{vv}} \tag{15}$$

Collectively, equations (1) to (15) formally define the proposed *two-dimensional Hurst operator* associated with a given mask and tile size.

The shape of the mask shall be fixed here to be octagonal and the limiting radii denoted by ρ_1 and ρ_n, the latter being the larger. We then refer to the two-dimensional Hurst operator as $\mathcal{H}$ and denote its operation on the square tile T_t of side t by

$$\mathcal{H}_{\rho_1,\rho_n}(T_t) = \begin{pmatrix} m \\ c \\ \eta^2 \end{pmatrix}_{T_t} \triangleq \varphi_{T_t} \tag{16}$$

The two-dimensional Hurst operator introduced herein maps a *tile* T_t in the image to a *three-component feature vector*, φ_{T_t}. Each of the m, c and η^2 values from each tile in the image could be scaled and displayed as an image to yield respectively the m, c and η^2 *images* resulting from $\mathcal{H}$. The two-dimensional Hurst operator may then be said to produce three *tile-to-pixel* mappings.

We shall, for the purpose of computation, centre the mask on every pixel within the tile and consider every pixel within the mask as long as it lies within the image, even if it lies outside the tile in question.

3. EXPERIMENTAL RESULTS

Mammogram *mdb244rm* from the MIAS database [8] was analyzed using $\mathcal{H}_{2,3}$, and square tiles of 8 pixels. The input image had a resolution of 400 μm per pixel (obtained by successive averaging of the 50 μm per pixel original image over square regions) and is shown in Figure 3(b). Due to space constraints, the m, c and η^2 images are not shown here, but only briefly described. The m image varies somewhat with tissue type and the c image appears edge-sensitive. The η^2 image is almost featureless; the range of η^2 is 0.82 to 0.98, indicating good fit of the straight lines to data across all tiles in the image.

A more interesting result is obtained from a scatter-plot of the (m, c) ordered pairs from each tile, as shown in Figure 3(a). Here, the scatter plot has been hand-segmented into a dense cluster, indicated by dots, at the top and a less dense, diffuse class of outliers indicated by circles. If the tiles in the image are binarized, with white indicating the outlier class and black, the tight single cluster, the image that results is shown in Figure 3(c). It is interesting to note that the circled lesion in Figure 3(b), which is not visually discernible from its surroundings, is readily apparent in Figure 3(c) as white patch within a dark region.

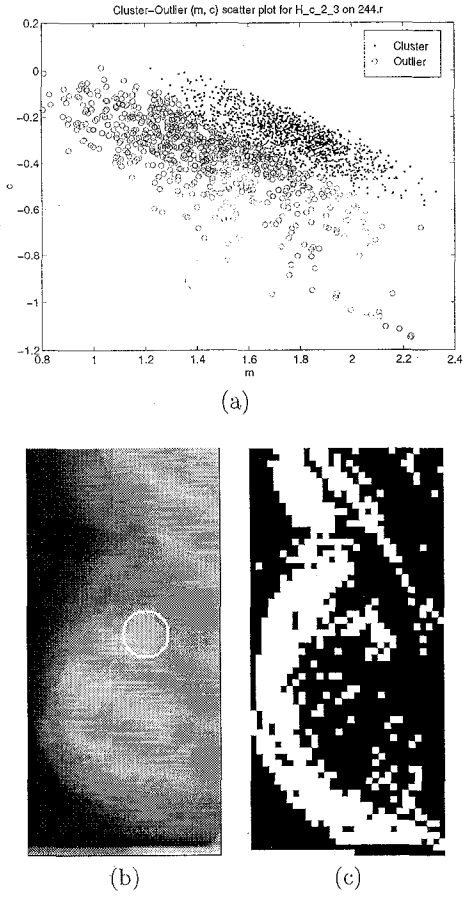

Fig. 3: This figure shows: (a) the single cluster and outlier group segmented manually from the (m, c) scatter plot resulting from applying $\mathcal{H}_{2,3}$ to image *mdb244rm*. The original mammogram is shown in (b) with the circumscribed mass lesion circled. Note that the glandular tissue is dense and makes the lesion difficult to discern. The binary labelled image from the scatter plot is shown in (c). Note that the outlier class in (a) includes the fat and muscle margin in (c). The lesion is also readily apparent as a patch of white in a black region in (c).

4. DISCUSSION AND CONCLUSIONS

The results of experiments on mammograms using $\mathcal{H}$ suggest that the operator discriminates between regions that are predominantly rough or jagged and those that are predominantly smooth. Thus, the largely smooth "fatty" regions on a mammogram appear white and the more jagged "fibroglandular" regions appear black on Figure 3(c). Such discrimination is valuable in detecting lesions that are texturally smooth, such as circumscribed masses, occurring in relatively rough "fibroglandular" regions. However, smooth areas in the "fibroglandular" region of mammograms do not always correspond to lesions, and additional features are necessary to increase specificity of lesion detection.

The nature of the information extracted from the data by $\mathcal{H}$ is also worth further investigation in the light of statistical, information or fractal theory.

5. ACKNOWLEDGEMENTS

RC gratefully acknowledges financial support through an OPRS Scholarship from the Australian Commonwealth Government, and support from the University of Western Australia through UPA awards and CIIPS Scholarships.

References

[1] A. R. Rao, *A Taxonomy for Texture Description and Identification*. Springer Series in Perception Engineering, New York: Springer-Verlag, 1990.

[2] T.-K. Lau and W. F. Bischof, "Automated Detection of Breast Tumours Using the Asymmetry Approach," *Computers and Biomedical Research*, vol. 24, pp. 273–295, June 1991.

[3] R. Gupta and P. E. Undrill, "The use of texture analysis to delineate suspicious masses in mammography," *Physics in Medicine and Biology*, vol. 40, pp. 835–855, 1995.

[4] H. E. Hurst, "Long-Term Storage Capacity of Reservoirs," *Transactions of the American Society of Civil Engineers*, vol. 116, pp. 770–808, 1951.

[5] H. E. Hurst, R. P. Black, and Y. M. Simaika, *Long-Term Storage: An Experimental Study*. London, UK: Constable & Co. Ltd., 1965.

[6] B. B. Mandelbrot and J. W. van Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, vol. 10, pp. 422–437, Oct. 1968.

[7] B. B. Mandelbrot and J. R. Wallis, "Robustness of Rescaled Range R/S in the Measurement of Noncyclic Long Run Statistical Dependence," *Water Resources Research*, vol. 5, pp. 967–988, Oct. 1969.

[8] Mammographic Image Analysis Society, (MIAS), "Digital Mammogram Database." Published electronically, 1994. Electronic contact addresses as of April 1996 are mias@sv1.smb.man.ac.uk for e-mail and http://s10d.smb.man.ac.uk/MIAScom.html for WorldWideWeb.