

Is Pattern Recognition a Physical Science ?

Jean Serra
Ecole des Mines de Paris , France
Serra@cmm.ensmp.fr

Abstract

This paper deals with the roles of assumption, mathematical models, and experimental control in Pattern Recognition. In order to base the discussion on actual examples, three studies are first described. The first one presents filtering of bird singings via their sonograms. The second and the third ones deal with morphological operators on the unit circle, which are successively applied to hue gradient and top-hats in colour images, and to directional irregularities in wood textures.

The discussion holds on the necessity to bring implicit assumptions to light, to evaluate the consistency of the mathematical formalism with respect to these assumptions, and finally to control experimentally the domain of pertinence of the proposed approach.

1. Introduction

The topics I would like to develop here were suggested by the reading of about 400 abstracts, submitted to ICPR 2000, and of a number (fortunately smaller) of complete papers. In addition to which I should mention the survey of recent literature on pattern recognition.

From this inventory, one can draw some strong impressions. First, one can notice, that beside some intensely investigated areas (such as watermarks for example), a number of others are hardly present if at all.

For instance, it is quite difficult to find a study combining image and sound processing in the same methodology. This is surprising : How authors who are mostly neither blind nor deaf, who therefore constantly combine the two sources of information of sight and hearing, suddenly split up in two, as soon as they transpose their sensorial activities into intellectual ones ? Strange schizophrenia...

A second example, a little more complex, is that of data distributed on the unit circle. Although they appear in several different domains of applications such as directions, in textures, hue, in colour imagery, or phases

in sound processing, very little theory is currently developed in this area.

Beyond these two illustrations, I would like to comment on another aspect of the papers: it is sometimes admitted that the exposition of the goals, as well as the paths from the goals to the methods, and from the methods to the domains of validity, are superfluous and not worth quoting.

However, in other papers, the methodology rests on strict constraints. The data bases is well delineated, the key assumptions are clarified, and the results are verified by means of orthodox statistical tests. Such a hiatus confronts us with a real problem : I do not believe that the various levels of quality of the works explain completely so deep differences. Perhaps could we suspect the matter itself, which sometimes lend itself to “clean” paradigms, and sometimes not. This points will be discussed in the last sections of this paper.

2. Image and Sound

In sound processing, there exists a series of questions which are rarely tackled, because they do not stand inside the “good” scales for signal processing methods. They concern phenomena that last several seconds, such as :

- urban (eg cars) or natural (eg waves) background noises ;
- sonic style of a speaker (accent, cough, long (or short) sentences, etc..) ;
- cry , or song of an animal .

Now, even if we cannot identify a speaker, nor recognise his words, but if we are able to extract the segments of time where a same person is speaking, we generate a nice piece of information for content based labelling. More generally, in this sort of questions, one has to define what is foreground and background, and then to separate them.

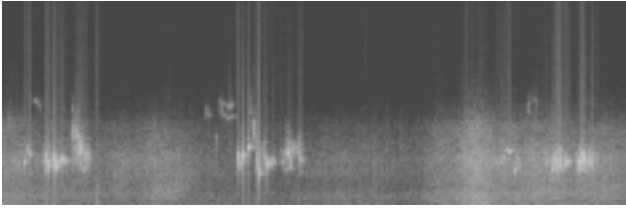


Figure 1. intensity sonogram of a robin singing above the sea (20 seconds).

Fourier transforms based on sliding windows provide two (time x frequency) representations $a(t,v)$, for the amplitude, and $\phi(t,v)$, for the phases, of the phenomenon under study. Should it be possible to treat them as images, in order to obtain two pairs of images for foreground and background respectively, and finally go back, for each of them, to a sound wave by means of inverse Fourier transform ?

In the state of the art, the question is correctly solved as soon as the background is stationary for the time dimension (but not necessarily for the frequencies). It is not the case for the robin song that we would like to extract from the sea background (fig. 1).



Figure 2. intensity sonogram of the robin without the noise of waves (20 seconds).

But we observe that the robin song covers completely the sea, so that we apply a morphological filter that extracts the zones of fig1a above a certain threshold t and marked by points above a second threshold t' (fig2). This new intensity sonogram, associated with the (unchanged) phase sonogram, allows to restore the song of the robin with a totally silent background [1].

3. Operators on the Unit Circle

In this section, two approaches for morphological operators on the unit circle are proposed, via two examples. Both deal with the processing of the hue in colour imagery. The reader will find a more developed approach of operators on the unit circle, with in particular

studies on directional textures, in recent work with A.G. Hanbury [2].

3.1 Hue processing

The spaces designed for colour imaging are "bad" vector spaces. To explain what is meant by "bad", we can take the counter-example of multi-spectral data for satellite imaging. Twelve infrared channels are easily reduced to one or two using the appropriate statistical method (principal component analysis, correspondence analysis...), and the few synthetic images thus obtained are then treated separately.

Things are very different with colour, probably because colour is much closer to human perception. Whereas retina cells get only three types of spectral responses (three "colours"), there exist more than twenty representations of colour in electronic imaging, with two, three, or four components, depending on the case. The founding experiments of 1931 made by the CIE, which enabled to establish the chromacity diagram, played an important role in generating an abundance of colour spaces, whose existence is often justified. The RGB of computers, which was not well suited for Hertzian transmission, has been replaced by YUY in TV transmission standards in Europe, and YIQ in the United States. However, the controls available to the TV-user, in Europe or in the United States, are of the HLS or HSV types,... which already makes seven different systems.

When we consider colour image processing, things get even worse. The (RGB) triplet, which is more or less used as a common denominator to other representations, remains arbitrary. Consequently, colour processing based on RGB, if aiming at being intrinsic, should commute with the rotations of the (R,G,B) vector cone. But, among all algorithms that process RGB colour images, how many do satisfy this requirement ? One ? None ? (NB : in grey scale, convolution and all morphological operators commute with the affinities and changes of origin).

In fact, human physiology is much closer to (H,S,V) or (H,L,S) representations than to those combining luminance and some chrominances, such as (R,G,B), (Y,U,V) or (Y,I,Q). This point is corroborated by some diseases such as colour-blindness or retina hypo-sensitivity to saturation. Now, (H,S,V) type representations mix heterogeneous dimensions, since an intensity (L or V) is grouped with the hue, defined on the unit circle. From the point of view of image processing, this constitutes a real challenge, which, I think, partly explains the lack of methodological advances in the field of colour.

3. 2 Increment based algorithms

To overcome this drawback, we can focus on the class of operators which bring into play a difference, such as gradients or top-hats. Fix an arbitrary origin a_0 on the unit circle C with centre o by, for example, choosing the highest point, and indicate the points a_i on the circle by their curvilinear co-ordinate in the trigonometric sense between 0 and 2π from a_0 . Given two points a and a' , denote by $a \div a'$ the value of the acute angle aoa' [3], i.e.

$$\begin{aligned} a \div a' &= |a - a'| & \text{if } |a - a'| \leq \pi \\ a \div a' &= 2\pi - |a - a'| & \text{if } |a - a'| \geq \pi \end{aligned} \quad (1)$$

It is known that in the Euclidean space \mathbb{R}^d , to determine the modulus of the gradient, at point x , of a numerical differentiable function f , one considers a small $S(x,r)$ centred on x with radius r . Then one takes the sup minus the inf of the increments $|f(x) - f(y)|$, where y describes the small sphere $S(x,r)$, i.e.

$$\begin{aligned} g(x, r) &= \bigvee \{ |f(x) - f(y)|, y \in S(x,r) \} \\ &- \bigwedge \{ |f(x) - f(y)|, y \in S(x,r) \} \end{aligned} \quad (2).$$

Finally, one determines the limit of the function $g(x, r)$ as $r \rightarrow 0$. This limit exists as the function f is differentiable in x . In the two-dimensional digital case, it is sufficient to apply (2), taking for $S(x,r)$ the unit circle centred on x (square or hexagon). This yields the classic Beucher algorithm for the gradient. As the previous development only involves increments, we can transpose (2) to the circular function a by replacing all the $|a(x) - a(y)|$ by $a(x) \div a(y)$. This transposition then defines the modulus of the gradient of the circular distribution. For example, in \mathbb{Z}^d , $K(x)$ indicates the set of neighbours at distance one from point x , hence

$$\begin{aligned} (\text{grad } a)(x) &= \bigvee \{ a(x) \div a(y), y \in K(x) \} \\ &- \bigwedge \{ a(x) \div a(y), y \in K(x) \} \end{aligned} \quad (3).$$

As an illustration, consider the hue component of the Virgin portrait, fig 3, whose histogram is shown in fig. 4.

As the dominant colour in the image is red, most of the values in the hue band are clustered at the top and bottom end of the histogram. A classical gradient on this hue band produces a large number of spurious high-valued pixels, as shown in fig. 5. The gradient calculated using rel. (3) shown in fig. 6, overcomes this problem.



Figure 3. P. Serra : the Virgin (detail), Monastery of St Cugat, Barcelona.

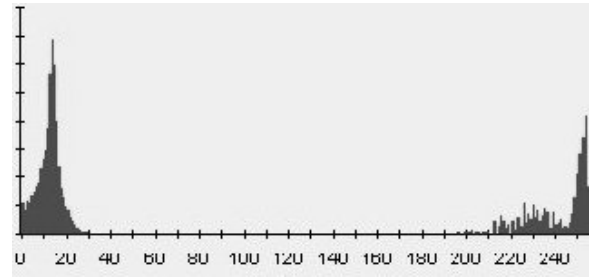


Figure 4. Hue histogram associated with fig. 3.



Figure 5. Classical gradient for the hue of fig. 3.

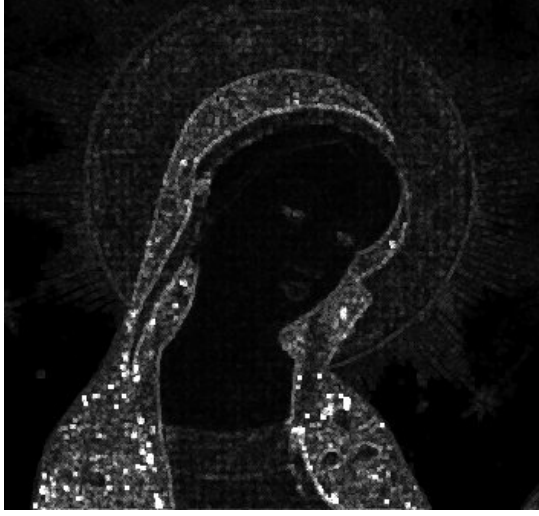


Figure 6. Circular gradient for the hue of fig. 3.

Note that if we rotate the hue band pixel values by π , the classical gradient will be the same as the angular gradient. The angular gradient is, however, invariant to rotations of the pixel values.

The above approach easily extends to *any grey-tone operator* that can be written by means of increments. Consider, for example, the notion of the "top-hat", in the sense of F. Meyer, i.e. the residue between a numerical function and an open version of it. Clearly, it only involves increments, and hence can be transposed to functions with values in C . We explain below the algorithm in case of an adjunction opening γ_B (i.e. the composition product of an erosion of structuring element B by the adjunct dilation). If $\{B_i, i \in I\}$ stands the family of structuring elements which contain a given point x , we have

$$\gamma_B(x) = \sup \{ \inf [f(y), y \in B_i], i \in I \}.$$

For the top-hat $f(x) - \gamma_B(x)$ we therefore write

$$f(x) - \gamma_B(x) = \bigvee \{ [\wedge [f(x) - f(y)], y \in B_i \setminus \{x\}] \vee 0, i \in I \},$$

which involves increments only of function f . Therefore it can be transposed to circular function exactly as we did for the gradient, which results in

$$a(x) - \gamma_B(x) = \bigvee \{ [\wedge [a(x) \div a(y)], y \in B_i \setminus \{x\}] \vee 0, i \in I \} \quad (4).$$



Figure 7 still life: (a) luminance, and (b) hue, images.

The still life, figure 7a, will illustrate the circular top-hat algorithm (4). As previously, we are facing with a colour image whose hue component, fig. 7b, has a strong contribution of the purple-to-red band, i.e. a zone around the hue origin. This distribution can be observed on the hue values of the grapes, for example, in fig. 7b. A direct top-hat of image 7b considered as a grey tone function, yields fig. 8a. The alternative circular top-hat of rel. (4), when applied to the same image, turns out to provide a completely different result (see fig. 8b). First, the latter is invariant under hue rotation, and second it actually extracts what is expected from a top-hat residual of size three, i.e. the small or narrow zones where the hue is changing.



Figure 8 : a- grey tone top-hat of size 3x3 of the hue image 7b ;
b- circular top-hat of size 3x3 of the same image.

3. 2 Angular openings

Circular data may be treated in another manner, which is more set-oriented, and where is no obligation to work on increments only. This second approach is based on the idea of first *labelling* the points of the working space according to the local hue, then processing the obtained sets, and finally combining the results in a isotropic way. Denote by $A(\alpha, \omega)$ the set of those points $x \in E$ whose circular value $a(x)$ lies in acute sector $[\alpha, \alpha + \omega]$.

$$A(\alpha, \omega) = \{x : x \in E ; a(x) \in [\alpha, \alpha + \omega]\}.$$

Let now $\{\gamma_\lambda, \lambda > 0\}$ be an arbitrary granulometry on $\mathcal{P}(E)$. Take the union of all transforms $\gamma_\lambda[A(\alpha, \omega)]$, as α describes the unit circle, i.e.

$$\gamma(A/\lambda, \omega) = \bigcup \{ \gamma_\lambda[A(\alpha, \omega)] , 0 \leq \alpha \leq 2\pi \} \quad (5)$$

Such an isotropisation generates a new family of operators $\{\gamma(*/\lambda, \omega), \lambda > 0, \pi > \omega > 0\}$, which turns out to be a *two parameters granulometry*. Note that the

γ 's do not depend on the choice of an origin of the directions.

Here a question arises. Since algorithms (4) and (5) rest on radically distinct assumptions, and manage different pieces of information, it should be instructive to control their differences in practice. We may have an idea of it by taking again the same still life fig. 7a.

To perform the angular opening of rel.(5), an angle $\omega = 45^\circ$ and a square structuring element with side of length $\lambda = 3$ were used. figure 9a shows the residual of this opening, i.e. the corresponding binary top-hat. For comparing this result with the (grey) circular top-hat of fig. 8b, we must apply a threshold to the latter. By taking all pixels of fig. 8b whose value is > 50 we obtain figure 9b. Strangely, the two sets are almost the same. Thus, the circular top-hat (4) and the angular opening (5), although they are conceptually independent, and they involve different parameters, yield finally, for some values of these parameters, results which are practically identical.

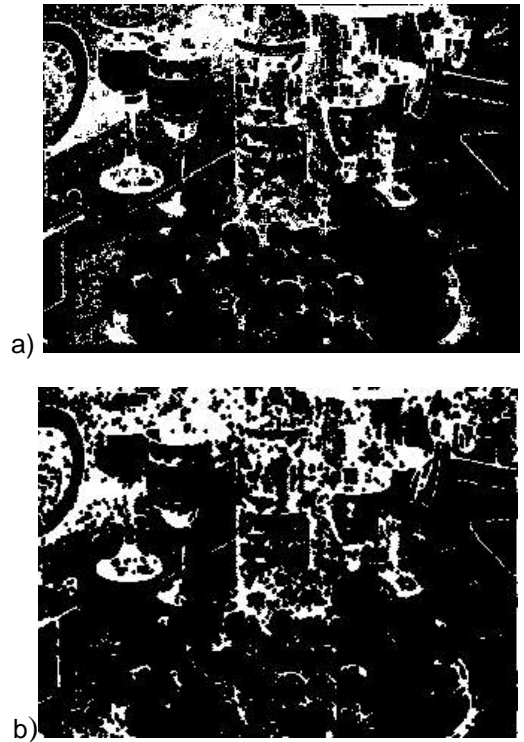


Figure 9 : a- residuals of opening γ of rel.(5), when applied to the hue image fig.7b. The parameters are $\lambda = 3$ and $\omega = 45^\circ$ in fig. 8a;
b- threshold of the hues > 50 in the circular top-hat of fig. 8b. Note the strong similarity of the results.

4. Modelling in Physics

The scientific activity that a conference such as ICPR 2000 reflects is partly a matter of physics. The various headings under which it comes, such as Pattern Recognition, Electronic Engineering, Signal of Image Processing, or even Applied Mathematics, though of easy use, must not make us forget the epistemology we depend on is that of modelling. This means that :

1 - The objects we study, namely image and sound, belong to the *physical* world. This sets us apart from pure mathematics whose achievements do not require an outside world. This also distinguishes us from other points of view on the outside world : when we describe cells becoming cancerous, we do not consider them as zones undergoing specific chemical reactions, but as objects undergoing specific deformations. As concerns the medical end, the two (chemical or image) points of view can be useful or not, but in any case they remain distinct.

2. We reduce the object under study to certain aspects we consider to be essential, at least for the specific questions we have in mind. This step, inescapable, is indeed the very hypothesis that justifies the development of an existing mathematical model, or the creation of a new one.

3. The model that has been selected, or elaborated, is all the more powerful, also, as it links, via characteristic equations, the various different pieces of information that can be measured on experimental objects. Note here that in a mathematical model, everything does not lend itself to a given purpose, as observed by G. Matheron [4] : one *chooses* to apply probabilistic methods rather than deterministic ones; one *chooses* to represent shapes by Euclidean closed sets. It is only the second consequences of these choices that lend themselves to experimental control. The characteristic function of a Boolean random closed set may be tested; two compact granulometries may be identified, for example.

4. The verifiable part of the model has a meaning only if it is *falsifiable*, in K. Popper's sense [5], or if one can exhibit situations where it is wrong. Conversely, we thus delimit the field of applicability of the model, hence its descriptive, or predictive, value.

5. Modelling in Pattern Recognition

One could argue that the epistemology of physical sciences is not relevant here, since its object is different ;

and one could say that Pattern Recognition basically aims at a quantitative description of visual perception, and not at possible laws of the material world.

5. 1 Several objects for a same model

This argument would be valid if every recognition method only applied to a unique object of study, which is obviously not the case. For example, a same texture model can represent a tangle of microscopic crystals as well as the spatial distribution of the trees in a forest. From one application field to the other, the *means* chosen to assess the models are different : in image compression (forecast), the important criterion is visual quality ; in the detection of pre-cancerous cells (medicine) the results are compared with physiological analyses whereas in material sciences they are matched with physical or chemical properties. But in all cases the criterion appears at the ultimate step and plays a little role in the mathematical formulation of the model and in its logical derivations (i.e. in the obtaining of the algorithms).

From the point of view of the mathematical model, this means that each physical domain where it is significantly used provides it with some meaning, some semantics. For example, Matheron's axiomatic for granulometry, which models the notions of size and of size distribution, has been introduced in the seventies as an abstraction of sieving techniques (for particles), and of mercury injections (for porous media). Thirty years later, the same axiomatic applies in a completely different domain, since it is currently used to model pyramids of operators in image segmentation [6].

Conversely, the mathematical model, as soon as it is comprehensive enough, looks at the real word with a certain style. Mathematical morphology, wavelets or fractals introduce in the world their own intuitions, and tend to reduce it to the articulations of their own thought. For the morphologist, everything will become matter for granulometry, filtering or connection, because he can then contemplate nice constructions of his method.

If the reader still doubts that Pattern Recognition shares common features with Physics, because of its methods as well as a part of its subject, here is an example. In 1983, Y. Pomeau and myself were asked two similar questions. They concerned stacks of near spherical particles, whose conductivity (electric in one case, thermic in the other) was sought. Starting from the assumption that the transfer only occurred at contact points (insulating exterior medium), we elaborated a model capable of estimating the three dimensional number of the contact points per volume unit, from the histogram of the distances between neighbouring circles seen in plane sections [7]. The model was tested,

independently, and proved its validity [8]. The law we had reached (rel (3) in [7]) is then as physical as the law of Ohm ; simply, the resistance is no longer proportional to the length/diameter ratio of a metallic wire, but to another, more complex but as much geometrical "pattern".

5. 2 Several models for a same object

However, Pattern Recognition, differs more radically from usual Physics because in our case both mathematical models *and* actual situations turn out to be generally polysemic. Two contradictory theories may equivalently be relevant to the interpretation of a same phenomenon. In the colour example on the still life, one of the two approaches, but not the other, was binary and depending on the choice of an angle range. However, they resulted in the same filtering.

Here, a pertinent epistemology seems to be that of W.V.O. Quine [9]. According to him, physical theories are *under-determined* with respect to even all possible observations. Therefore, prior to any test, we must concentrate on what we wish to obtain from the theory. There is no sophism here, since we can always confront different theories from the angle of view of a given goal. There is no sophism also in that some assessments are relevant to testing and other not.

The first example, on birds songs, illustrates quite well these various aspects. We have *chosen* to pass from the sound to the "sonogram" image (and back) to treat this problem. When we respect the rules of sampling, the operator for this passage is reversible : there is no actual hypothesis at this point. The choice to formulate the problem in terms of images is suggested by the duration (several seconds) of the patterns to extract, but this does not enter the Popperian logic of falsification. As to the hypothesis, here it is : if, as air vibrations, sounds are additive, as perceptive events, they follow other laws ; among other things, when a structured signal is significantly higher than a permanent (but variable) background sound, the ear does not perceive the background noise any more as soon as the higher signal occurs. In other words, in the intensity sonogram, the sum can be replaced by the supremum inside the zones corresponding to the dominant signal. Hence, the use of morphological filtering is justified. Second hypothesis : we assume that when the intensities in a region Z of the sonogram are set to zero, that does not modify the phases of the signal inside complement region Z^c .

Here the experiments are clearly falsifiable (Popper's epistemology) : it suffices to extend the same approach to the background, by taking the difference between initial sound and extracted bird singing to produce audible deformations. We reach the limits of the model.

On the other hand, by assuming that the audition was rather that of "birds \vee sea" than "birds + sea" we have replaced a deconvolution by a supremum. The alternative linear approach can perfectly be implemented and its results confronted with the current ones (Quine's epistemology).

6. Textures and random models

Although the real word under-determines the possible mathematical models, as soon as we penetrate inside a given model, we must stay consistent with ourselves and accept the constraints of its framework. At this level, there appears a real danger of confusing the two epistemologies.

For example, when one surveys the literature on Pattern Recognition, it soon appears that the expression "heuristic use of " more than once serves as an alibi to inconsistent treatments. And this is particularly true when the adjective "heuristic" is close to the word "texture". The text usually begin by the ritual incantation repeated over and over from one paper to the other, and according to which the "*notion of a texture is not defined, and nobody knows exactly what it means*". Then, curiously, the rest of the study definitely refutes these incantatory preliminaries.

Considering a zone Z on an image as texture, is simply *choosing* to model it by a periodic or stationary random function of small range with respect to the zone Z (the range in one direction is the distance from which two point can be considered independent). It is absolutely amazing that those prone to the said incantations never notice that *each time* one speaks of textures, a probabilistic approach comes immediately after...

However, a probabilistic approach has its own constraints. Its hypotheses have to be explicit (by the choice of a model), but also it imposes a certain way of proceeding. It is at this point that the "heuristic" talisman appears. If this word means that a random function may suggest algorithms, provided that we do not have to test it, by some strict statistics, in order to validate its use, then I agree with the "heuristics". But this word can also mean "just doing anything under cover of the talisman".

We will illustrate this point with the uses of the stationary Boolean functions. This very simple texture model is obtained by taking various realisations of a compact random function, implanting them in Poisson points, and taking the supremum. Boolean sets were invented by G. Matheron [10] [11], their extension to numerical functions is mainly due to J. Serra [12, founder

text] [13, tutorial]. They allow to model and simulate a huge variety of textures such as the crystals of fig.10a.

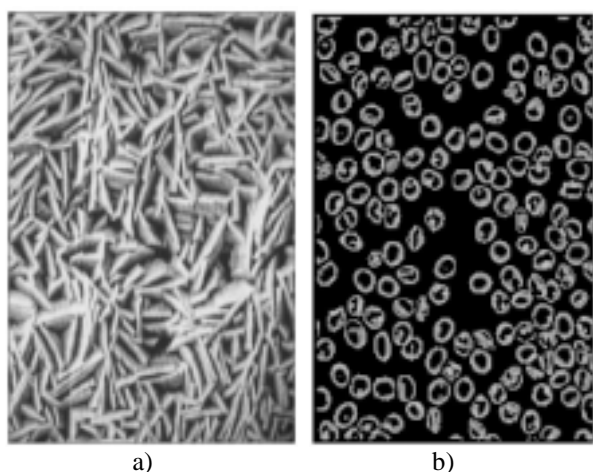


Figure 10 Two “boolean” textures, the first (ferrite crystals, a)) in the statistical sense, the second (red blood cells, b)) in the heuristic one.

From a heuristic point of view, we owe boolean functions, one of the most performing algorithms for counting. Suppose for example that we want to count the red blood cells of figure 10b. If n stands for the number of red blood cells, Z for the area of the field, v' for the average grey-volume of a cell and $q(t)$ for the area proportion of the white cells at threshold t , then we have

$$n = (Z/v') \int q(t) dt \quad (6)$$

if the space distribution is Boolean. Then, we just have to put the background t_0 at zero, to extract a few isolated cells (in order to estimate v') and to compute rel (6). Here, by placing the background level at $t_0 = 90$, we find $n = 197$ and with $t_0 = 100$, $n = 189$. A manual count yields $n = 211$. Algorithm (6) is robust and precise enough for a correct medical use. Now clearly, all statistical tests should reject fig. 10b as a Boolean realisation, because of its lack of overlapping.

Imagine now that, instead of integrating over the grey tones, in rel (6), we integrate over the bit planes of the digital image. Obviously, the result become disastrous (indeed, it results in $n = 42$). Now, it is not difficult to find in recent literature on Pattern Recognition, that after having assumed a Boolean function for the texture under study, the authors study it via its bin-planes, by arguing that their approach is heuristic.

Another variant of such "heuristics" sometimes appears in computing chord length distribution of a binary texture. The said texture is first supposed to realise, say, a

Boolean set. Then, as soon as they have exhibited the “heuristic” talisman, the authors consider themselves to be free enough to replace the usual scanning by a Hilbert one (i.e. a digital peano curve). Why not, since they are “heuristicians” ? Unfortunately, the only interest of the Boolean chord distributions is to estimate, via correlation, some parameters of the models, and irregular scanning patterns (typically Hilbert ones) are purposefully constructed to destroy all possible space correlation...

8. References

- [1] Serra J., Azcaraté F., Bachelart L., de Frescheville F. Musical Dictations 19p. *workshop on Frontiers of Mathematical Morphology* Univ. of Strasbourg, April 17-20, 2000.
- [2] Hanbury A., Serra J., *Morphological Operators on the Unit Circle*, 14Pp. March 2000 (submitted to IEEE Transactions on Image Processing.
- [3] Peters II, R. A. *Mathematical Morphology for angle-valued images*. In *Non-Linear Image Processing VIII*. SPIE volume 3026, 1997.
- [4] Matheron, G., *Estimating and choosing*, Springer-Verlag, 1989.
- [5] Popper, K., *The logic of Scientific discovery* 1959. (first edition, *Logik der Forschung*, Vienna, 1934)
- [6] Meyer F. *Morphological Multiscale and Interactive Segmentation* NSIP99, June 1999
- [7] Pomeau Y., Serra J. *Contacts in random packing of spheres*. J. of Micr., Vol. 138, N°2, 1985, pp.179-185
- [8] Chermant J.L., Coster M. Giraud G., Jernot J.P. Robine H. *Détermination du nombre de coordination dans un empilement de sphères à partir de sections planes aléatoires*. Journal de Physique, 1986, vol.47, 829-837.
- [9] Quine W.V.O. *On the reason of indeterminacy of translation* J.of Philosophy., 1970.
- [10] Matheron G. *Eléments pour une théorie des milieux poreux*. Masson, Paris, 1967.
- [11] Matheron G. *Random Sets and Integral Geometry*. Wiley, New York, 1975.
- [12] Serra J. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1982.
- [13] Serra J. *Boolean random functions*. J. of Micr., Vol. 156, N°1, October 1989, pp. 41-63.