# Space-Scale Analysis in the Human Primal Sketch

R.J. Watt
MRC Applied Psychology Unit
15 Chaucer Road,
Cambridge,
CB2 2EF, U.K.

## ABSTRACT

Recent psychophysical evidence suggests that the human visual system uses a Primal Sketch stage of processing that is radically different from the original proposal of Marr (1976). Spatial filters of different sizes are used by two distinct processes. One is concerned with statistical, "texture" properties of the image, the other with the spatial geometry of the image.

The statistical processor appears to have immediate access to all the appropriate spatial scales in an image. The subject is able to report texture qualities no more accurately after a 1000ms exposure than after a 10ms exposure. The geometric processor, however, appears to scan through the spatial scales in a coarse-to-fine direction, so that there are substantial improvements in visual performance over at least 1000ms.

In this paper I address the issue of why this division of labour should be found and consider the nature of control in the dynamic Primal Sketch representation. In other words, what are the advantages, to human vision, in this type of behaviour, and what constraints in the world do they reflect?

## INTRODUCTION

The existence of a multiresolution system of linear spatial filters in human vision has been acknowledged for some time (Campbell and Robson, 1968). There have been many speculations concerning the purpose of such a system; in this paper I shall make the assumption that they are used to create spatially differentiated images at different spatial scales so that important luminance changes in the image can be discovered and characterized. There is now a considerable accumulation of evidence to support this assumption (c.f. Watt and Morgan, 1985; Watt, 1987a).

Recent experiments (Watt, 1987b) have shown that for some tasks the visual system scans, over a period of at least 1000ms from a spatial scale with a standard deviation of approximately 2 arc deg. down to a scale of approximately 1 arc min. The experiments all involved judgements of simple attributes of single isolated lines, such as orientation, length or curvature. Take the orientation case as an example of the logic employed. Before spatial filtering, a vertical line (length $\underline{L}$ and negligible width) has a vertical luminance dispersion of standard deviation:

$$\sigma_v = \frac{L}{2\sqrt{3}}$$

and a horizontal dispersion of standard deviation:

$$\sigma_h = 0$$

Filtering at a spatial scale of standard deviation $\underline{s}$, leads to a more dispersed distribution with:

$$\sigma_v = \left(\frac{L^2}{12} + S^2\right)^{\frac{1}{2}}$$

$$\sigma_h = S$$

In judging the orientation of the line, it is the ratio of these two that will determine the level of performance:

$$\frac{\sigma v}{\sigma h} = \frac{1}{s}\left(\frac{L^2}{12} + S^2\right)^{\frac{1}{2}}$$

This factor ranges from 1, when L=0 to infinity when s=0. If sensitivity, $\underline{d\theta}$, is defined to lie in the range 0 to infinity, then:

$$d\theta = \frac{\sigma v - \sigma h}{h} = \frac{(L^2/12 + S^2)^{\frac{1}{2}} - S}{S}$$

This parameter is dependent on both $\underline{L}$ and $\underline{s}$. It follows that $\underline{s}$ can be estimated by measuring $\underline{d\theta}$ as a function of $\underline{L}$. This was done at a variety of different exposure durations and it was found that the spatial scale, $\underline{s}$, is reciprocally related to exposure duration, $\underline{t}$:

$$S = \frac{k}{t}$$



AVC 1987 doi:10.5244/C.1.22

A similar logic can be and was applied to other tasks, such as judging the length of the line, its curvature or its orientation in depth (with horizontal binocular disparity). Figure 1 shows how the spatial scale of visual analysis varies with exposure duration for each of these. The agreement between tasks is good.

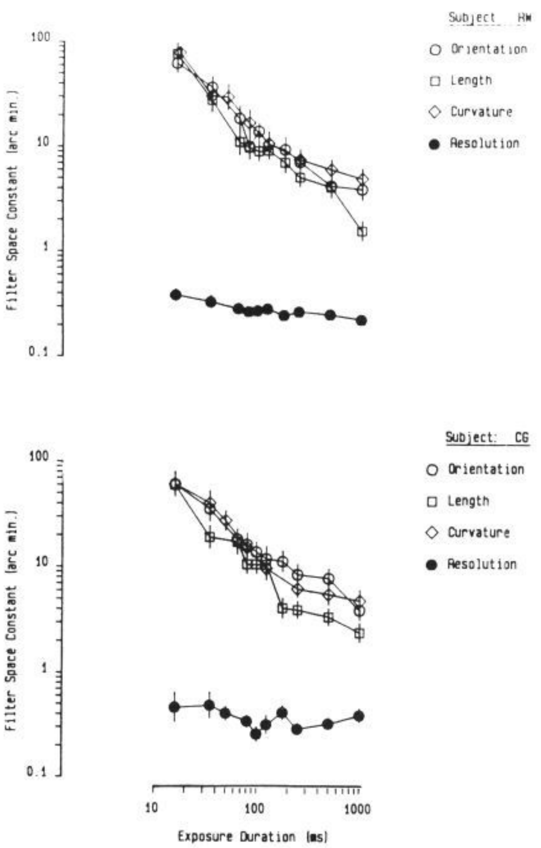


Figure 1: shows the variation in spatial scale of analysis in human vision with exposure time for a variety of different tasks.

The figure also shows the derived spatial scale of analysis for a visual resolution task where a dotted line was to be discriminated from a continuous line. In this instance there is no variation in spatial scale with exposure duration.

These experiments reveal two distinct patterns of change in the spatial scale of visual analysis with time. In this paper, I wish to address the computational reasons for this.

## I: What is Being Computed at the Two Extremes of Spatial Scale?

There are several possibilities for the two distinct modes of spatial computation that have been described in the introduction. In this section, I will list the possibilities and describe a series of experiments that decide between them. The answer to the question remains descriptive at present, but I shall speculate why the distinction exists.

i) The attributes of line orientation, length, curvature and stereo-disparity are all geometric properties. The continuity or dottedness of a line is a topological property.

ii) The continuous lines all show a single spatial organization at all scales of analysis; this is not true of a dotted line.

iii) The dots on the dotted line are "textons" (Julesz, 1981). The visual resolution task is, in effect, a texture discrimination task and may be done pre-attentively, whatever that means. The others, require attention, whatever that is.

iv) The local statistics of the dotted line are different to those of a continuous line. Local statistics are those defined only with respect to the single stimulus and without reference to the global position, size, or orientation of the pattern. It is the calculation of these external references that requires scanning from coarse-to-fine spatial scales.

Experiments in which subjects examined three patches of random dots to detect the odd-patch-out, throw light on these hypotheses. Basically, there are two types of effect of exposure duration: no effect and a reciprocal effect on detectability of the odd-patch-out depending on how the odd patch differed from the other two. There was no effect when the cue was a doubling of each dot as in a glass pattern either coherently in direction and distance or randomly (resolve cue). There was no effect when the cue was the direction of the glass pattern (dipole cue). There was a reciprocal effect when the cue was a random perturbation in the position of each dot (jitter cue). There was a reciprocal effect when the cue was a rotation of the entire pattern of dots (rotate cue). These tasks are illustrated in Figures 2 and 3, and the results in Figures 4 and 5.

With the results it is clear that the distinction between the two processes is determined by the task not the stimulus. The distinction is not determined by a topological versus geometric task. Simplistically at least, the distinction is not determined by the number of textons. It is likely that the first, high resolution process concerns local statistics of patterns, and the slower scanning process concerns precise geometric measures of individual dot positions.

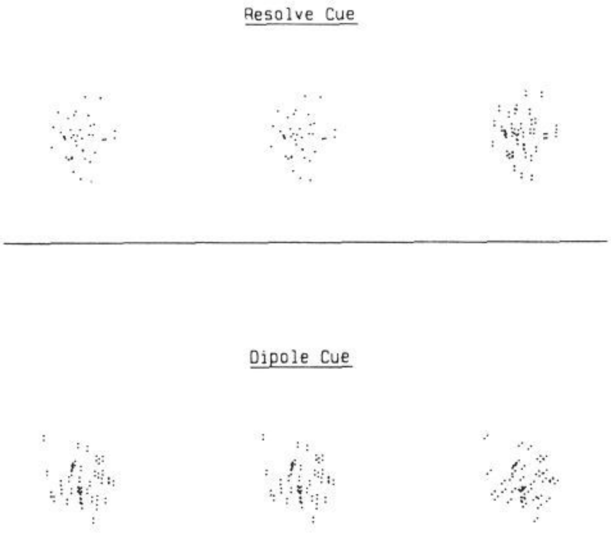


Figure 2: illustrates the stimuli for experiments in which subjects were required to detect the presence of a glass pattern doubling (top) or to judge the direction of the glass pattern (bottom).

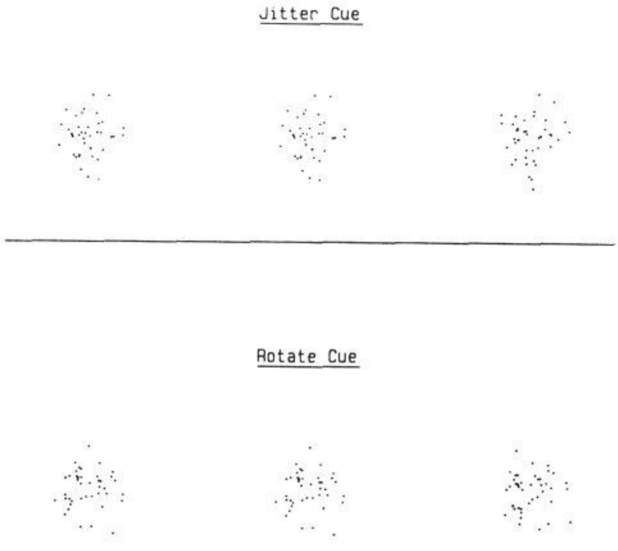


Figure 3: illustrates the stimuli for experiments in which subjects were required to detect a random perturbation in the position of each dot (top) or judge the orientation of the pattern (bottom).

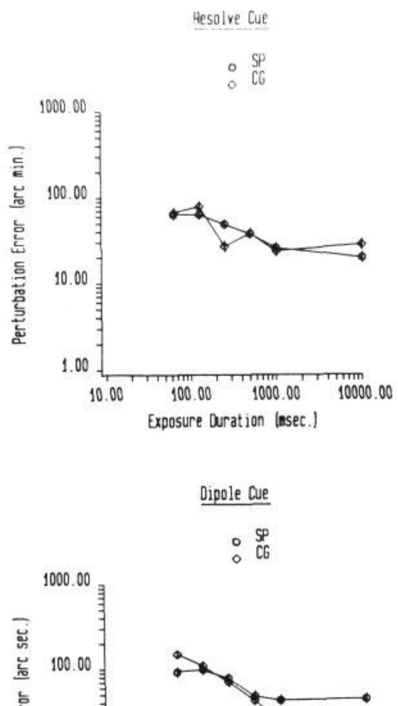


Figure 4: shows the effect of the exposure duration on the glass pattern tasks.

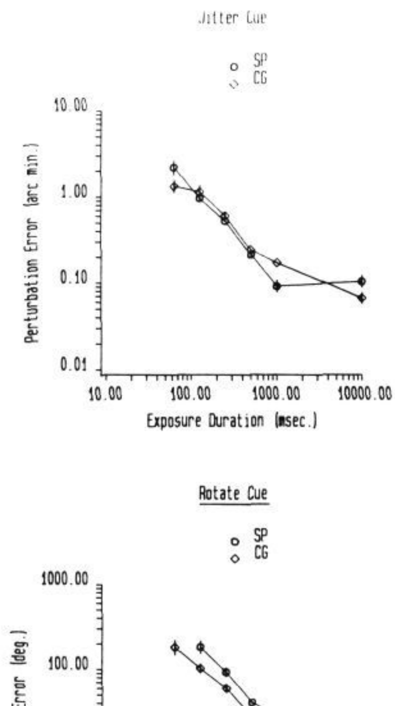


Figure 5: shows the effect of the exposure duration on the local position tasks.

## II: Why this Distinction?

It has been found that it takes time to compute the exact positions of individual dots (or line ends, corners, or intersections). The time is spent in scanning from a coarse to a fine scale. Why? I now offer a speculation, which follows the argument of Watt (1987).

i) The human visual system has many sources of geometric distortion: imperfect optics; hemispherical retina; relatively scrambled connections in the optic nerve; and many others.

ii) These distortions can be compensated for by the use of error-correcting metrics or codes. (c.f. Andrews, 1964).

iii) Such corrections lead to a non-Euclidean representation of the retinal image.

iv) In order to satisfy the constraint of Euclidean geometry, an iterative constraint relaxation type process is invoked.

v) The number of iterations increases with the number of dots (or line ends, corners, intersections) to be computed.

vi) The number of dots (etc.) decreases with increasing scale. Therefore a coarse-to-fine strategy is advantageous.

## III: Why a Reciprocal Relationship Between Time and Spatial Scale?

The experimental evidence is quite unequivocal that a reciprocal relationship is applied between time and spatial scale. There must be a good computational-theoretic reason for this behaviour.

To investigate this question I have examined synthentic fractal, one-dimensional functions. Natural images are thought to be fractal, at least over a wide range of spatial scales (Pentland, 1984). It would have been impractical to examine several thousand real images. In some respects it would be more generally interesting to use surface fractals, but there are three comments to make concerning the use of line fractals. First, line fractals are computationally cheaper. Second, the interest lies in the number of intervals found at each scale: it is plausible that dividing a line signal into distances is analogous to dividing a surface signal into areas (blobs). Third, so far as human vision is concerned, it remains an open question whether isotropic filters or one-dimensional filters are the more appropriate analysis.

Fractal patterns vary in their fractional dimension, which corresponds to the jaggedness of the pattern. This covaries exactly with the slope of the power spectrum, and is a measure of the relative energies at large and fine spatial scales.

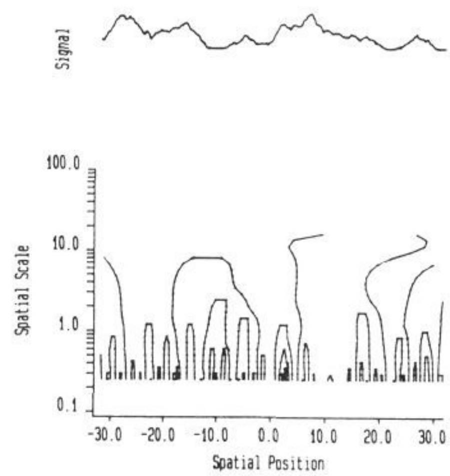


Figure 6: shows the space-scale diagram for a synthetic fractal function.

For each function, a space-scale diagram (c.f. Witkin, 1986) was computed (see Figure 6). At each spatial scale the number of zero-crossings was then counted. This is plotted as a function of spatial scale in Figure 7. Note that both axes are logarithmic. The panel in Figure 7 labelled "Log.change in zero-crossings" is a plot of derivative of the log/log plot of the number of zero-crossings. The mean of this derivative is approximately -1.0, indicating an exponent of -1 in the spatial scale versus number of zero-crossings function.

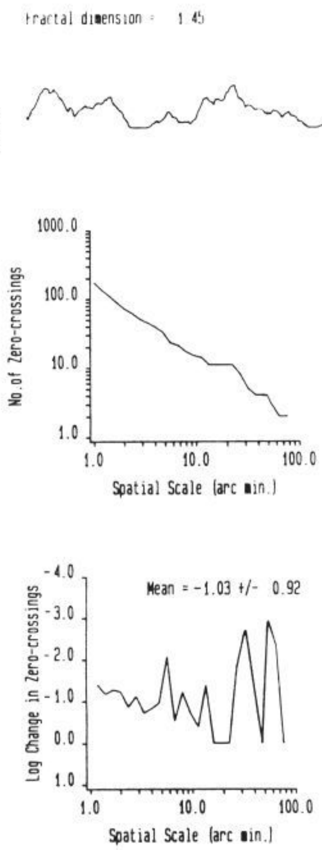


Figure 7: shows the variation in the number of zero-crossings with spatial scale (center panel) and the exponent of this function (bottom right hand panel).

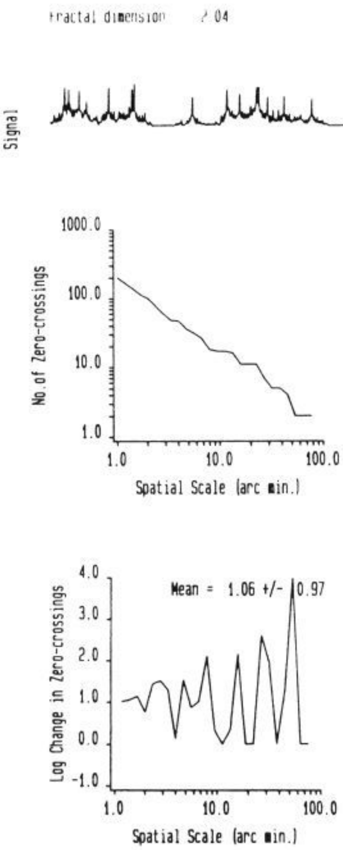


Figure 8: is the same as Figure 7, but for a function of higher fractal dimension.

After many such calculations, each with a fresh fractal function, the grand mean exponent was found to be -1.032. Interestingly, the correlation between fractal dimension and exponent was nearly zero at 0.106.

The consequence of this for a visual system which exhibits the scanning behaviour illustrated in Figure 1 is straightforward. As the visual system reduces spatial scale by a factor of two (for example) the number of elements in the representation is approximately doubled. There is not of course a particular scale at which all the number of elements suddenly doubles. As scale is increased an adjacent pair of zero crossings appear: this corresponds to the creation of a new zero-bounded distribution of response. The original has been replaced by three. Analysis of the space-scale diagram suggests that this tends to occur at a spatial scale that is one third of the scale at which the original distribution itself appeared.

Since we have the number of zero-crossings inversely proportional to spatial scale, which is itself inversely proportional to time (in the human visual system) we then have the relationship that number of zero-crossings is proportional to time. It then follows that the rate of increase of number of zero-crossings does not alter with time. If the visual system were limited in speed at any one instant by the number of zero-crossings, then the scanning behaviour would not help; if it is limited by the increase in zero-crossings then the scanning behaviour is a sound and rational strategy. This latter is believed to be the case.

For an interesting contrast to this behaviour of fractal patterns, Figure 9 shows the one-dimensional luminance pattern down a page of text. Notice that number of zero-crossings alternates between being independent of spatial scale and being a very steep function of spatial scale. The page of text is curious in that the zero-bounded distributions of response all divide at very nearly the same spatial scale.

I conclude that the representation of position is hierarchical, with position for zero-bounded distributions being calculated when they appear and added to an already existent Primal Sketch representation.

The Primal Sketch in human vision is dynamic, structured and has memory. This stands in contrast to most machine vision approaches to edge-finding.

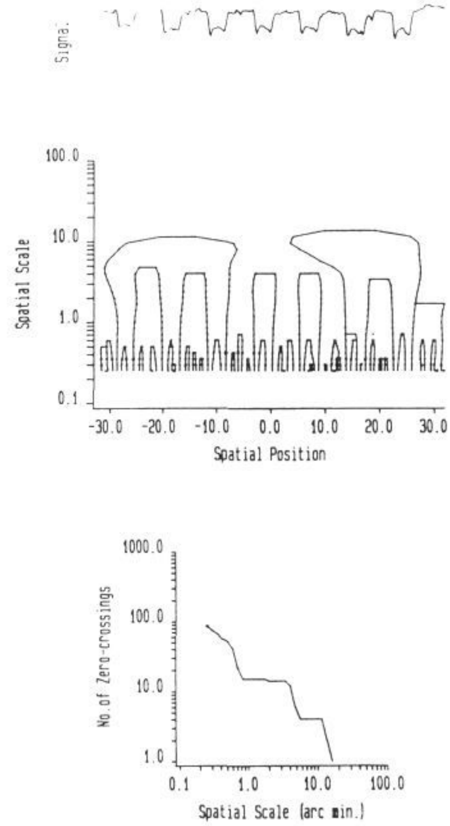


Figure 9: shows the space-scale diagram for a page of printed text and the variation in number of zero-crossings with spatial scale. This is quite different from the fractal functions.

## ACKNOWLEDGEMENTS

The support of the CEC COST 13 "Low-Level Vision and Image Processing" project is gratefully acknowledged.

## REFERENCES

1. Andrews, D.P. (1964). Error-correcting perceptual mechanisms. Quart. J. Expt. Psychol., 16, 104-115.

2. Campbell, F.W. and Robson, J. (1968). Application of Fourier analysis to the visibility of gratings. J. Physiol., 197, 551-566.

3. Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. Nature, 290, 91-97.

4. Marr, D.C. (1976). Early processing of visual information. Phil. Trans. R. Soc. B., 275, 483-524.

5. Pentland, A. (1984). Fractal-based description of natural scenes. IEEE PAMI, 6, 661-675.

6. Watt, R.J. (1987). An outline of the primal sketch in human vision. Pattern Recognition Letters, 5, 139-150.

7. Watt, R.J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. J. Opt. Soc. Amer., in press.

8. Watt, R.J. and Morgan, M.J. (1985). A theory of the primitive spatial code in human vision. Vis. Research, 25, 1661-1674.

9. Witkin, A.P. (1986). Scale Space Filtering in "From Pixels to Predicates". Ed. A. Pentland. Ablex, Norwood, N.J.