A new imaging architecture and an alternative interpretation of the structure of the human retina

Maria Petrou
Imperial College London, SW7 2AZ, UK

This paper was published, as key-note speech, in Proceedings of the International Symposium on Information Technology, Kuala Lumpur Convention Centre, Malaysia, 26–29 August, H B Zaman, T M T Sembok, K van Rijsbergen, L Zadeh, P Bruza, T Shih and M N Taib (eds), IEEE Cat. No CFP0833E-PRT, ISBN 978-1-4244-2327-9, Vol 1, pp 9–17.

Abstract

An architecture is proposed where the layers of cells that lie in front of the photosensitive cells of the human retina are interpreted to be estimators of the first and second order derivatives of the brightness of the imaged scene. Such an architecture bypasses the problems of estimating derivatives from sampled and digitised data, as they are estimated directly from the scene. It also offers an explanation on why the photosensitive sensors of the human eye are placed at the back of the eye, behind four other layers of cells.

1. Introduction

In many image processing and computer vision operations we have to use the brightness derivatives of the observed scene. Examples of such operations include all methods that rely on gradient estimation, Laplacian estimation or estimation of higher order fluctuations. Applications include edge detection, multiresolution image representation using the Laplacian pyramid, all methods that rely on anisotropic (or isotropic) diffusion approaches and even wavelet-based methods. In all these cases the derivatives needed are calculated from the discrete data. Discretisation, however, introduces significant errors in the calculation of differentials. An example is shown in figure 1, taken from [17]. The signal shown in 1a lasts 5.12sec. Its first half is the sum of 10 cosines with frequencies 1,2,...,10 Hz. The second half is the sum of 5 cosines with frequencies 11,...,15 Hz. The amplitudes of the cosines are normalised so that the two halves have the same energy. The sampling frequency is 50 Hz and the number of samples is 256. The discrete wavelet transform analysis of the signal is shown in 1b. We can hardly identify the dominant frequencies present in the signal. The continuous wavelet transform of the signal is shown in 1c. This is a much more accurate representation of the frequency contents of the signal, as judged from 1a. The mother wavelet used in both cases was the quadratic biorthogonal spline with 7 vanishing moments and the discrete wavelet transform was applied up to the fifth level of decomposition. The reader is reminded that the wavelet transform is computed using signal/image convolutions with high-pass differentiating filters.

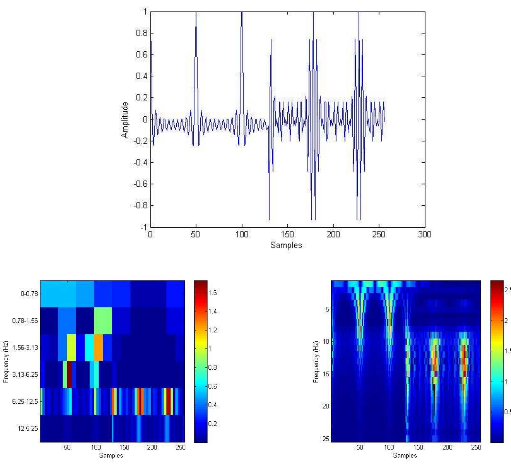

Figure 1. A signal and its discrete (left) and continuous (right) wavelet transform.

In recent years a realisation has started to emerge about the importance of the continuous signal we are dealing with, through the sampling data we have. Going back to the continuous scene to perform our processing, rather than forgetting it once the digital data have been gathered, has been shown to have many advantages. For example, Joshi [6] has shown that much better histogram estimates of the data may be obtained by upsampling and interpolating the data before calculating the histograms. Thus, the probability density functions of the various distributions needed for further image processing may be estimated much more accurately this way. Splines [19] is a very versatile and powerful tool for representing the discrete data in the continuous domain. However, there have also been cases where people try to go back to the continuous domain by emulating "continuous" sensors. In [18] virtual cameras were introduced, with spectral responses in between the discrete spectral responses of actual cameras, in order to improve colour segmentation. In [4] virtual sensors measuring the potential at border points of a 2D vector field were introduced in order to improve the vector field reconstruction using the inverse Radon transform.

It is not necessary, however, for the extra sensors introduced to measure the same value as the existing sensors. It is true, that a much denser array of CCDs will sample the brightness of the scene much better than a not so dense array, and it will make the estimated quantities approach more the true ones that refer to the continuous scene. There is another, option, however: it may be possible to use extra sensors that measure the desired quantities directly from the continuous scene. I started by mentioning the significant role differentiation plays in image processing. I would suggest that we may incorporate sensors in our imaging devices that measure the first and second derivatives of the scene directly, as they measure the brightness of the scene. This may be done densely and for different colour bands. The information from the derivative sensors may be used for subsequent processing as extra information alongside the brightness information, for example in the form of extra image bands, or it may be used to construct a very accurate representation of the scene, much more accurate than a single layer of brightness sensors may do on their own. This interlacing of sensors of different types and different sensitivities appears to sound too complicated, but it may be what nature has implemented for us.

It is well known that the retina responds to changes of light and not directly to light [3, 20]. Apparently this is based on studies which showed that when the tremor of the eye was switched off, the person went blind. This indicates that the eye works like a spatial differentiator of the scene.

So, one may speculate that what the eye does is not (or at least not only) to measure the brightness of a scene, but the derivatives of the scene too. Apart from the derivative information being valuable in its own right, integration may be used to work out a much more accurate representation of the signal than it can be recorded by the brightness sensors themselves. Integration requires values for the constants of integration. I will analyse next the roles of the constants of integration and show how they may lead to an imaging architecture that resembles very much the human retina, at least in appearance and topological structure.

2. Constants of integration

Assume that we know the second derivative $d^2f(x)/dx^2$ of a function $f(x)$. What are the values of function $f(x)$? We have to integrate $d^2f(x)/dx^2$ twice: first we find the first derivative of the function

$$\frac{df(x)}{dx} = \int \frac{d^2f(x)}{dx^2}dx + c_1 \tag{1}$$

where c_1 is a constant of integration. Then we have to integrate $df(x)/dx$ once more to derive function $f(x)$

$$\begin{aligned} f(x) &= \int \left(\int \frac{d^2f(x)}{dx^2}dx + c_1 \right) dx + c_2 \\ &= \int \left(\int \frac{d^2f(x)}{dx^2}dx \right) dx + c_1 x + c_2 \end{aligned} \tag{2}$$

where c_2 is a constant of integration.

Note that the constants of integration appear because we perform an indefinite integral. When we perform definite integrals between pre-specified lower and upper limits, say a and b, the result we get is a numerical value of the area under the curve of the integrand between these two limits.

Now, let us consider digital integration. In the digital domain, differentiation is replaced by differencing and integration by summation. The summation, however, is between specific values of the summation index, and so it really corresponds to the *definite* integration of the continuous domain. What in the digital domain corresponds to the *indefinite* integration of the continuous domain is the recovery of the values of the differenced function at all sample positions, by propagating a starting value. I will explain this with a specific example.

Assume that the true values of a function in a succession of sampling points are:

$$x_1, \quad x_2, \quad x_3, \quad x_4, \ldots, x_N \tag{3}$$

Assume that I am given only the first difference values at each of the sampling points, defined as $d_i \equiv x_i - x_{i-1}$:

$$\begin{aligned} ?, d_2, d_3, d_4, \ldots, d_N \quad &\equiv \\ ?, x_2 - x_1, x_3 - x_2, x_4 - x_3, \ldots, x_N - x_{N-1} & \end{aligned} \tag{4}$$

Here the question mark means that I do not have the value at the first point due to the definition I used for d_i. Note that if I had used the alternative definition, $d_i \equiv x_{i+1} - x_i$, I would still have to have a question mark, but this time referring to the last point of the sampling sequence. Also note, that usually, in image processing problems, we ignore this fact by assuming that the signal repeats itself ad infinitum, and so the first point becomes the last plus 1 point and the last point becomes the first minus 1 point. When such an assumption is made, there are no points for which we do not have a next door neighbour to compute the first difference, whichever definition we use. This of course is a trick of convenience in computing and it has nothing to do with reality. In reality we are always going to have a sample (either the first or the last) with no difference value.

Let us stick to the case depicted by equation (4), and let us try to recover the values of the original sequence, from the knowledge of the d values. To do that we hypothesise that the first value of the sequence is c_1. Then, the recovered values are:

$$c_1, \quad c_1 + d_1, \quad c_1 + d_1 + d_2, \quad c_1 + d_1 + d_2 + d_3, \\ c_1 + d_1 + d_2 + d_3 + d_4, \ldots, c_1 + d_1 + d_2 + \ldots + d_N \quad (5)$$

This process corresponds to the indefinite integration of the continuous case, with constant of integration the guessed original value c_1.

There are three important observations to make.

- Without the knowledge of c_1 it is impossible to reconstruct the sequence.
- To recover the value at a single point we need to add the values of several input points.
- As the sequence is built sample by sample, any error in any of the samples is carried forward and is accumulated to the subsequent samples, so the Nth sample will be the one with the most erroneous value.

There are two conclusions that can be drawn from the above observations.

- Such reconstructions cannot be too long, as very quickly the error of reconstruction accumulates and the reconstruction becomes useless. So, for the reconstruction of a long sequence, one has to consider many small sequences in succession, and possibly with overlapping parts.
- If one has a series of sensors that return the local difference value of the observed scene, one needs another series of sensors that return the value of c_1 every so often in the sequence, ie at the beginning of every small reconstruction sequence.

Next, suppose that the array of sensors we have does not measure the first difference of the sequence, but the second difference, $dd_i \equiv d_i - d_{i-1}$. Then we must apply the above process of reconstruction once in order to get the sequence d_i and then once more to get the x_i values. Note that this implies that we must have a series of sensors that every so often in the long sequence of dd_i will supply the starting constant we need, which in this case is denoted by c_2. This constant is actually a first difference, so these sensors should measure the first difference at several locations.

3. A novel imaging architecture

A device that functions according to the principles discussed above, has to consist of five layers, as shown in figure 2.

Now, suppose that the sensors that detect the constants c_2 at the same time act as adders, ie they add what they detect to the signal they receive from the dd sensors. Then the device will look like the one in figure 3. This structure may be re-modelled as shown in figure 4. The function of this structure effectively repeats twice: below the dashed line we have the first integration, outputting above the dashed line the values of the first difference it computes, and above the dashed line we have the second integration, integrating the first differences it receives and outputting the signal values. Figure 5 shows for comparison a cross section of the human retina. The topology of the structure of figure 4 is strikingly similar to the observed structure of the human retina!

4. Extension to 2D

The analysis done in the previous two sections is in 1D. However, images are 2D. This has some serious implications, particularly for the c_2 sensors.

From the mathematical point of view, once we move to 2D, we are dealing with 2D integrals, not 1D. A 2D integration implies spatially dependant constants of integration. For a start, a 2D function $f(x, y)$ has two spatial derivatives, $\partial f/\partial x$ and $\partial f/\partial y$. Let us assume that we know both of them and we wish to recover function $f(x, y)$ by integration. Integrating the first one of them will yield

$$f(x, y) = \int \frac{\partial f}{\partial x} dx + c_x(y) \quad (6)$$

where $c_x(y)$ is a function of y, which, as far as integration over x is concerned, is a constant. Differentiating result (6) with respect to y should yield $\partial f/\partial y$, which is known, and this can help us work out constant $c_x(y)$ as a function of y.

There is an alternative route to work out $f(x, y)$. Integrating the partial derivative with respect to y we get

$$f(x, y) = \int \frac{\partial f}{\partial y} dy + c_y(x) \quad (7)$$

where $c_y(x)$ is a function of x, which as far as integration over y is concerned, is a constant. Differentiating result (7) with respect to x should yield $\partial f/\partial x$, which is known, and this can help us work out constant $c_y(x)$ as a function of x.

Obviously, both routes should yield the same answer. In the digital domain, this corresponds to reconstruction of the 2D signal either line by line or column by column. So, let us assume that the true values of the 2D digital signal are g_{ij}. However, we do not have these values, but we are given instead the first differences of the digital signal along both directions. So, we assume that we have $dx_{ij} \equiv g_{ij} - g_{i-1,j}$ and $dy_{ij} \equiv g_{ij} - g_{i,j-1}$. We can construct the signal column by column as follows. First column:

$$g_{12} = dy_{12} + c_y(1)$$
$$g_{13} = dy_{13} + dy_{12} + c_y(1)$$
$$\ldots$$

Second column:

$$g_{22} = dy_{22} + c_y(2)$$
$$g_{23} = dy_{23} + dy_{22} + c_y(2)$$
$$\ldots$$

And similarly for the rest of the columns. This is shown in figure 6a. In a similar way, the signal may be reconstructed along rows. First row:

$$g_{21} = dx_{21} + c_x(1)$$
$$g_{31} = dx_{31} + dx_{21} + c_x(1)$$
$$\ldots$$

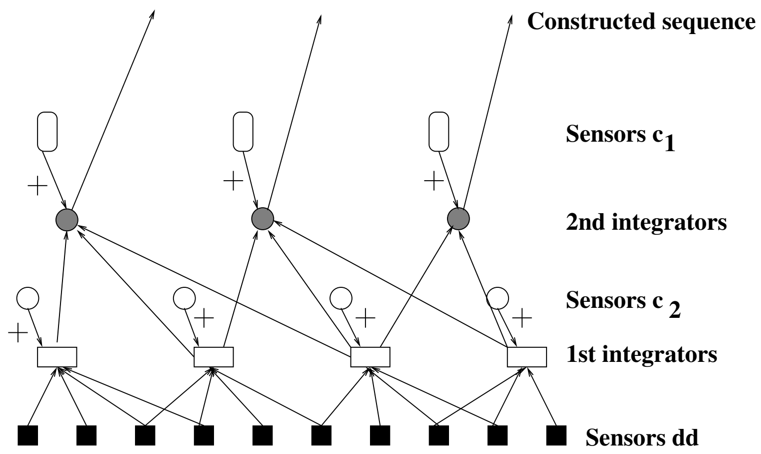

Figure 2. A five layer sensor device

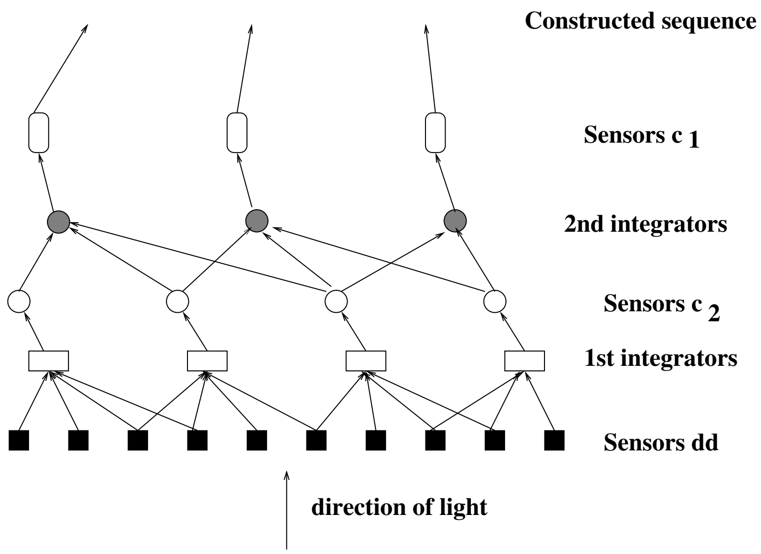

Figure 3. When sensors for c_2 also act as adders

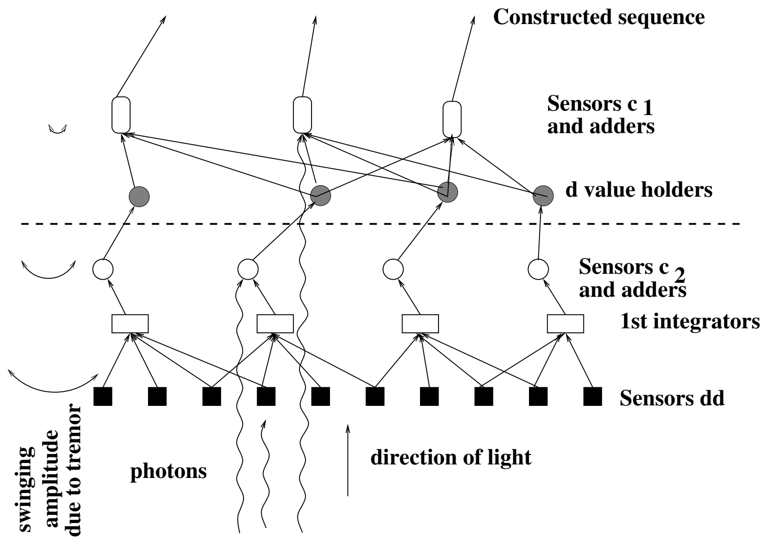

Figure 4. When sensors for c_1 also act as adders

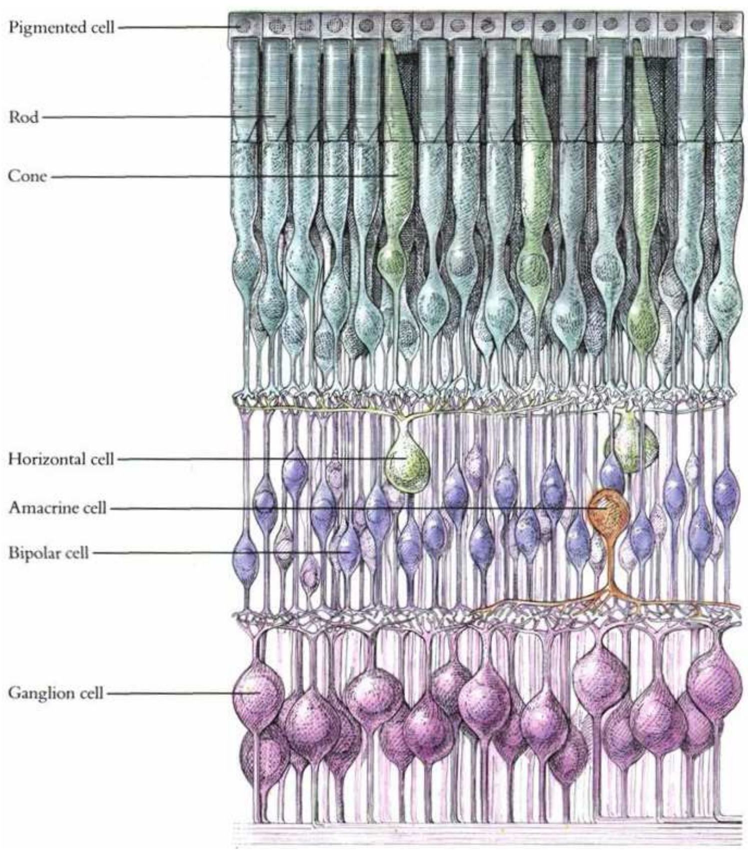

Figure 5. A cross section of the human retina.

Second row:

$$g_{22} = dx_{22} + c_x(2)$$
$$g_{32} = dx_{32} + dx_{22} + c_x(2)$$
$$\ldots$$

And similarly for the rest of the rows. This is shown in figure 6b.

Of course, these reconstructions should be equivalent, ie one expects that $g_{22} = dy_{22} + c_y(2) = dx_{22} + c_x(2)$. One may also reconstruct the signal by using a combination of rows and columns, and again, the reconstruction should be the same irrespective of the path followed. This is shown in figure 6c.

There are two problems with the above analysis: in practise the alternative reconstructions are never identical due to noise. This is something well known from digital image processing. The other problem is the use of two directions which creates an anisotropic grid, as there are two preferred orientations. Along these two orientations, the samples used are at a fixed distance from each other. However, if we consider samples that are aligned along the diagonal of these two orientations, their distance is $\sqrt{2}$ times that of the samples along the preferred orientations. This anisotropy is not desirable. The third problem we can see is that such a reconstruction has to proceed along parallel lines where we first reconstruct the $c_y(i)$ values, for example, and then use them to propagate the solution in the orthogonal direction. Not desirable logistics.

The above approach is compatible with the conventional CCD sensors that consist of rectangular cells, ie rectangular pixels. The cones in the fovea region of the retina, however, have a hexagonal structure, as shown in figure 7a. At first sight this does not look very useful. However, instead of considering the cells, consider their centres as the sampling points of a grid. The nodes in the grid shown in figure 7b are the points where the reconstruction has to take place.

This sampling grid at first sight does not appear hexagonal, but rather based on equilateral triangles. However, several hexagons of various scales can be perceived here.

Imagine now, that we have a device centred at the centre of one of these hexagons. Imagine that the device vibrates along the paths shown. Imagine that this device hangs from a vertical nail above the centre of the hexagon, and consists of three types of sensor hanging from the same string: the bottom one measures second differences, the middle one first differences, and the top one just values. As the string swings like a pendulum, the bottom sensor swings more, the middle less and the top not at all (see left of figure 4). This will be consistent with the notion that the second difference sensor needs to see larger part of the scene to do its job than the first difference sensor, while the fixed sensor does not need to swing at all to do its job. Note: it is mathematically impossible to calculate any derivative if you consider only a single sample. So, a device like the one shown in figure 4 swinging along one direction, will allow the reconstruction of the signal along that direction for several sampling points. The amplitude of the swing and the range of reconstruction performed by each single set of sensors are two different things. The amplitude of the swing is for measuring locally what is needed for the reconstruction. Swinging along another direction, will measure the first and second differences along that direction, and the signal will be reconstructed along that direction, by using the propagation techniques we discussed in the 1D case.

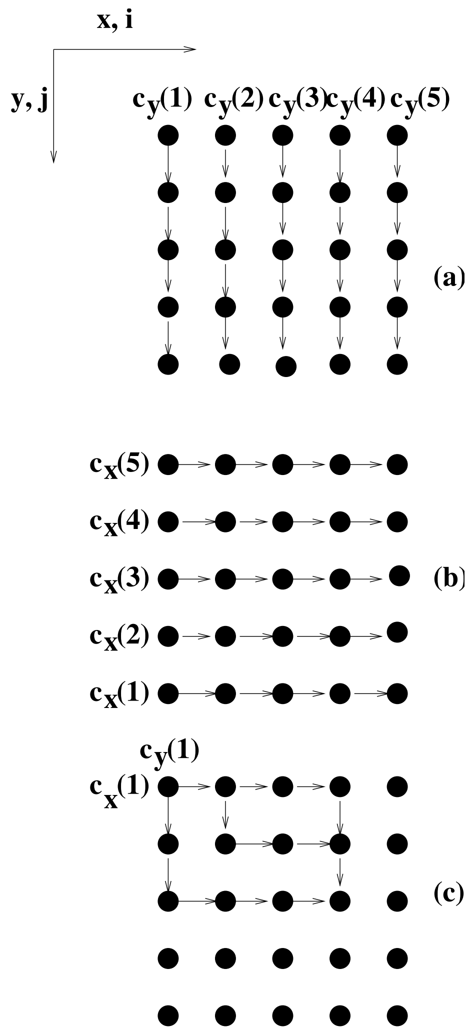

Figure 6. Reconstruction from first difference values in 2D can proceed along columns (a), or rows (b), or along any path (c). The answers should be equivalent.

There are many advantages of this approach: the reconstruction grid is isotropic; we have no preferred directions; the hexagons fit nicely with each other at all scales; the reconstruction along the lines of one hexagon can be complemented by the reconstruction along the lines of other hexagons that may be directly underneath other sets of sensors hanging from our swinging strings; overlapping reconstructions are expected to add robustness and super-acuity (ie resolution higher than the sampling distance as determined by the spacing of the sensors); the reconstruction is expected to be complete and very accurate.

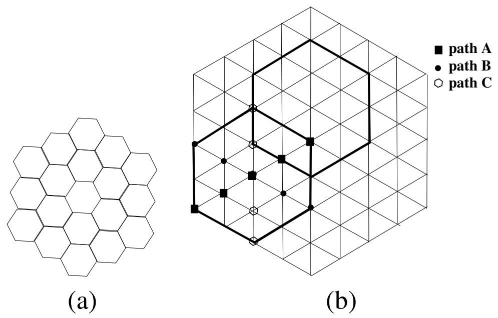

Figure 7. (a) The arrangement of cells in the mammalian retina. (b) The hexagonal sampling grid.

5. Implications for the physiology of the retina

The proposed imaging structure has several implications.

- The light meets first the dd sensors (see figure 4) because the dd signal has to be known first before the constant c_2 is added by the c_2 sensors. So, the observed layering of the sensors in the retina is in agreement with this theory (see figure 5).
- The dd sensors should have receptive fields that look like $(-1, 2, -1)$ (or $(1, -2, 1)$) so that they sense the local second derivative. Ganglion cells (see figure 5) are known to be spot detectors [5], i.e. they do have such receptive fields.
- The sensors that sense the c_1 constant do not have to move. The sensors that sense the c_2 constant have to move less than the dd sensors, because the c_2 constants are first order derivatives. Mathematically you need to use a smaller patch to compute the first derivative than the second derivative. Now, this says that the dd sensors are further away from the wall of the retina, so they swing more than the c_2 sensors which are closer to the wall, and which swing more than the c_1 sensors. The c_1 sensors are fixed on the retina wall, and most likely they are not moving at all (figure 4). Probably tightly packed as well, so they are not swinging.
- The c_2 sensors have to estimate the first difference at the points they see. This means that their receptive fields have to be $(-1, 1)$ or $(1, -1)$, so when they are shifted over an area, they get the first difference at the central point.

Predictions that may be tested:

- The bipolar cells as well as the ganglion cells are also light sensitive with receptive fields that look like $(-1, 1)$ or $(1, -1)$.
- Tremor may have larger amplitude as we move away from the retinal wall.

At the moment it is believed that the ganglion cells receive input from the cones in order to detect the 2nd difference. This would be absurd: why discretise first and compute the derivative afterwards, if you can compute the derivative directly from the continuous data? Discretisation introduces errors. It is possible that the sensitivity of the ganglion cells to light has not been observed yet because the electrodes used in order to study them stabilise them. If you stop the tremor, it is obvious that the 1st and the 2nd order difference filters will give 0 response. The theory presented in this paper contradicts that: ganglion cells are spot detectors directly from the scene, with no need of the cones to send them information. It also predicts that bipolar cells are photosensitive. If such radical ideas are correct, and what we see is built up from its first and second derivatives, then large regions with no sensory information may take up longer time to be built up than small regions. This is in agreement with several experiments where the so called "in-filling" of large regions takes longer to be accomplished [11, 14].

6. Implications for the philosophy of mind

The fundamental question that pre-occupies cognitive philosophers on vision is whether a perfect image of the viewed scene exists somewhere inside the human brain. The idea that such an image exists is known as Cartesian materialism [21], and the place where it exists is known as Cartesian theatre. Most cognitive philosophers nowadays reject this idea as absurd [1]. There are several philosophers, however, who support instead the idea of analytic isomorphism: in simple words, they do not accept that an image of the scene is constructed by the brain, but rather that there is a one-to-one correspondence between the state of the neurons and the perception created [16]. The basic argument is: why should the brain create a perfect picture if nobody needs it? All the brain needs is to be able to deal with the information received and translate it into actions. There are also philosophers who even dispute this isomorphism, and accept that it is possible the mapping from sensors to perceptions and actions not to be one-to-one at all [12]. The ideas presented here imply that the human retina may act as a device that creates an accurate and complete image of the viewed scene. This assertion may be interpreted as Cartesian materialism. However, the ideas presented do not enter into these arguments. For example, the information gathered by the retina may not at all be used to build a 2D image. The vibrating sensors that measure first and second derivatives as well as the level of brightness at a point, may only do local reconstructions along 1D lines about their central position. It may be that it is only these local reconstructions that are mapped to perceptions and actions, with the global picture being an "illusion" of the high level processing of the brain-a matter of which combination of such local reconstructions/sensations we call, for example, a "lake" as opposed to reconstructing actually the 2D image of a lake in our brain.

The ideas presented here can further explain the time reversal effect [22] (the so called Colour Phi phenomenon): when two stimuli are rapidly flashed on the screen, in some cases, the second stimulus is perceived before the first! Two explanations have been proposed for this. The so called Stalinisque explanation (according to which there is buffer in the brain where information is kept and the incoming second stimulus affects the stored information from the first stimulus during the time lag of perception) and the Orwellian explanation (according to which the error happens in the recall rather than during reception/perception). Both names of the explanations are due to Dennett who rejects both of them [1, 21]. This phenomenon may be explained if when the integration of the stimuli takes place, the constants of integration change and affect the result. This explanation is similar to the Stalinisque explanation, except the interference happens when the sensors are building the incoming information rather than due to the existence of a buffer, (which buffer may be identified with the Cartesian theatre, the existence of which is mostly disputed).

7. Implications for the construction of novel imaging devices

The use of a hexagonal grids as a better way of performing image processing has been known for some time, and several authors have done a lot of research on them [10, 2]. Also, the use of vibration as a way of improving image quality has been investigated both theoretically and implemented in practice in imaging devices [7, 9, 13]. There have also been developed devices that can measure the first scene derivative [8, 15]. However, no device has been constructed yet for measuring the second image derivatives directly from the scene. Such a device, when constructed, it will revolutionarise the way we perform image processing.

8. Conclusions

The ideas presented here challenge our understanding on the way retina gathers information rather than our understanding on what happens to this information from there on. The fundamental understanding this paper questions is that the only photosensitive cells in the retina are the cones and the rodes, and that ganglion cells receive information from the cones in order to respond to spatial variations and thus act as spot detectors. The predictions of this theory may or may not be proven right by the appropriate physiological experiments. However, either the human retina is constructed as discussed here, or as it is thought by current conventional thinking, this is irrelevant to whether we should construct such imaging devices or not. The basic questions that will have to be addressed by physicists and engineers are: 1) can we develop sensors that can estimate the first and second derivatives directly from the scene? 2) will the outputs of these sensors be more accurate and resilient to noise than the calculations of the derivatives from the sampled data? These questions have to be answered by sensor scientists, as they cannot be answered theoretically. There is no doubt that if the answer is "yes" to both these questions, the image processing that we shall be able to do with such devices will be much more reliable and accurate than the image processing we are doing now.

Acknowledgements: This work was supported by a UK Research Councils translation grant.

References

[1] D. C. Dennett, *Consciousness Explained*, Little, Brown and Co, USA, ISBN: 0-316-18065-3.

[2] F. Faille and M. Petrou, *Bio-inspired, Invariant Image Reconstruction from Irregularly Placed Samples*, BMVC 2008, Leeds, September 1–4, 2008.

[3] J. T. Fulton, *Processes in Animal Vision*, (online) Corona Del Mar, CA. USA, Vision Concepts, URL:http://www.4colorvision.com

[4] A. Giannakidis, L. Kotoulas and M. Petrou, *Improved 2D Vector Field Reconstruction using Virtual Sensors and the Radon Transform*, International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, Canada, 20–24 August, 2008.

[5] D. H. Hubel, *Eye, Brain and Vision*, Scientific American Library, ISBN 0-7167-5020-1, 1988. http://neuro.med.harvard.edu/site/dh/.

[6] N. B. Joshi, *Non-parametric probability density function estimation for medical images*, PhD thesis, University of Oxford, 2007.

[7] A. Kadyrov and M. Petrou, *Reverse Engineering the human vision system: a possible explanation for the role of microsaccades*, in Proceedings of ICPR, 2004.

[8] J. Kramer, *An integrated optical transient sensor*, IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, 49(9):612–628, 2002.

[9] O. Landolt and A. Mitros, *Visual sensor with resolution enhancement by mechanical vibrations*, Autonomous Robots, 11:233–239, 2001.

[10] L. Middleton and J. Sivaswamy, *Hexagonal Image processing*, Springer, ISBN-13: 978-1-85233-914-2, 2004.

[11] M. A. Paradiso, S. Shimojo and K. Nakayama, *Subjective contours, tilt-aftereffects and visual cortical organisation*, Vision Research, 29:1205–1213.

[12] L. Pessoa, E. Thompson and A. Noe, *Finding out about filling in: a guide to perceptual completion for visual science and the philosophy of perception*, Brain Science, http://www.bbsonline.org/Preprints/OldArchive/bbs.pessoa.html, 2004.

[13] Remote Group Ltd, eXsYght camera, http://www.exsyght.com/

[14] W. D. Rosi and M. A. Paradiso, *Temporal limits of brightness induction and mechanisms of brightness perception*, Vision Research, 36:1391–1398.

[15] A. A. Stocker, *Compact integrated transconductance amplifier circuit for temporal differentiation*, Proceedings of the International Symposium on Circuits and Systems, I:25–28, 2003.

[16] D. Todorovic, *The Craik-O'Brien-Cornsweet effect: new varieties and theoretical implications*, Perception and Psychophysics, 42:545–600.

[17] A. Varnavas, *Signal processing methods for EEG data classification*, PhD thesis, Imperial College London, 2008.

[18] J Verges-Llahi, *Colour Constancy and Image Segmentation Techniques for Applications to Mobile Robotics*, PhD thesis, University Politecnica de Catalunya, Barcelona, Spain, 2005.

[19] M. Unser, *A Guided Tour of Splines for Medical Imaging*, Plenary talk, Twelfth Annual Meeting on Medical Image Understanding and Analysis (MIUA'08), Dundee UK, Scotland, July 2-3, 2008.

[20] A. Yarbus A, *The role of eye movements in vision*, Plenum Press, New York, 1967.

[21] http://en.wikipedia.org/wiki/Cartesian_materialism

[22] http://en.wikipedia.org/wiki/Color_Phi_phenomenon