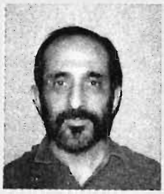
# Imaging Sensors: Artificial and Natural

*Vikram Dhar*



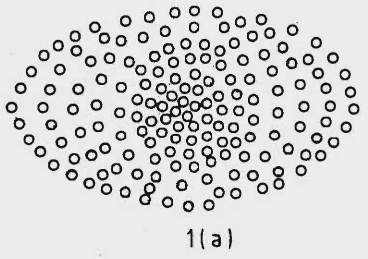
Vikram Dhar is a senior scientist at the Solid State Physics Laboratory, Delhi. He works on infrared imaging using linear and focal plane arrays. He is also interested in biological issues.

## Achilles is Still Trying to Catch up with the Tortoise

Often it seems that we humans, despite all our ingenuity, are only trying to catch up with *mother nature*. This should not be surprising – after all, she has had a head start of several million years. Soon after World War II, physical scientists invented the radar and sonar, but were shocked when it was found that bats have been using sonar all along. Not only that, some moths – which are preyed on by bats – are now known to be jamming the sonar of the bats, ('blinding' them at the eleventh hour!), as a counter-measure. In many fields, including the allied subjects of imaging and robotics, physical scientists today consciously try to emulate natural models.

## Natural and Manmade Sensors

A less well-known instance of how we have been anticipated by evolution refers to the compound eye of insects like bees, wasps, etc. and of arthropods like the horseshoe crab (*Figure 1*). The compound eye consists of several thousand ommatidia. Each ommatidium is a separate detector, conical in shape, arrayed on the surface of a sphere (of radius $r$). It is sensitive only to light coming from points on or near its axis (see *Figure 2*) which points in a radial direction. The compound eye presents to the insect a mosaic view consisting of many small spots due to each individual ommatidium. It turns out that the size of the ommatidium is just the value that gives the best possible resolution consistent with diffraction of visible light (see *Box 1*). That is, over evolutionary time scales, bees evolved to have ommatidia of just the optimum size (assuming the radius of the bee's head is fixed). Still, the resolution of the bee's compound eye is some fifty times less than that of a human – but then the



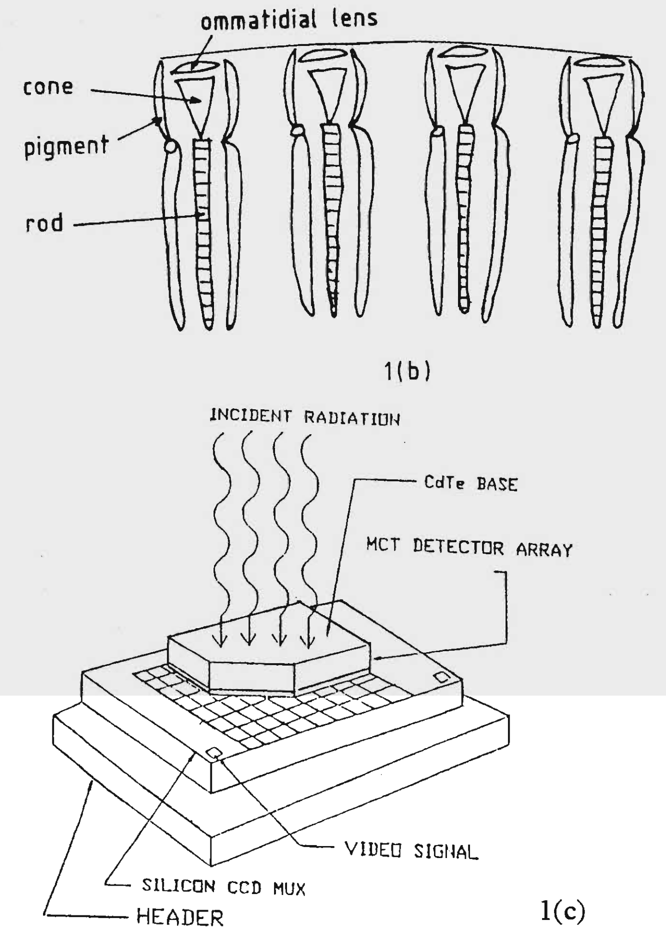




*Figure 1. The top and side views of the ommatidia in the compound eye of a horseshoe crab are shown in 1a and 1b, whereas the geometry of a manmade infrared detector is shown in 1c. The elements of the detector array are not visible in 1c, because they are on the lower side, but they are equal in number to the elements of the silicon CCD multiplexer (MUX) – to which they are connected one-to-one by indium bumps (flexible metal interconnect).*

bee's head is more than fifty times smaller than the human head! That is, the compound eye is the best way of packing an eye in a small volume and close to the surface; put differently: the size of the compound eye increases if higher resolution is demanded. For a particular radius, the bee can increase the resolution of its image by using lower wavelength light – and indeed the bee's eye is sensitive to the near ultraviolet, unlike our eyes. Some instances of creatures, which exploit infrared radiation for imaging is also known – as discussed below.

What does that have to do with the laggard physical scientist? Well, 'pixelated' arrays (see *Box 2*) of visible light detectors

**Box 1**

The resolution is defined as the minimum distance or angle that the system can distinguish. The resolution of the eye is determined by the angular size of the ommatidium (see *Figure* 1). The smaller the angle, the better the resolution until one reaches the diffraction limit, whereupon decreasing the angle will only degrade the resolution.

$$\Delta\theta = (\delta/r) - (\lambda/\delta)$$

where $\delta$ is the size of the ommatidium, $\lambda$ is the wavelength of light, the first term is the angle defined by the ommatidium, and the second refers to the diffraction term.The angle giving best resolution is obtained for $\delta = (\lambda r)^{1/2}$, by differentiation. From the above expression, for $r = 3$ mm and $\lambda = 4 \times 10^{-4}$ mm, the $\delta = 0.035$ mm (observed is 0.030 mm).

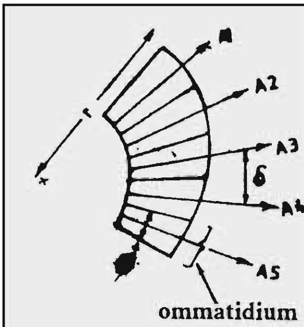



*Figure 2. Shows the geometry of a section of a compound eye with 5 ommatidia. The angle of the cone of an ommatidium is Θ, the radius of the insect's head on which the ommatidia are laid out is r, and the arc of the ommatidium at the surface is δ (and the axes of the ommatidia are A1–A5). Decreasing the angle increases the resolution, but increases the diffraction. Thus, beyond a point, decreasing the angle further is not advantageous, because diffraction dominates.*

resembling the compound eye have been manufactured in silicon (apart from the more conventional vidicon tubes, where the pixels are less well defined), as the so-called charge coupled devices (CCDs, in which charge packets corresponding to the signals are transferred from the points of generation in a controlled fashion across the array to the output, and then passed on to the signal amplifier). Apart from visible detectors, even infrared detectors have been made, with materials like indium antimonide (that detect 3–5 micrometer radiation) and mercury cadmium telluride (that can detect either/both 3–5 and 8–14 micrometer radiations, which correspond to 'windows' in the atmosphere, of minimum attenuation of infrared). The detectors and the optics that are used (see *Box 3*) are subject to the same limitations as those of the insects.

We can take some solace in the fact that we are able to make infrared detectors with as many as 1024 × 1024 pixels in the 3–5 micrometer range and 256 × 256 in the 8–12 micrometer range – whereas our only competitor in the natural world, the reticulated python (which responds to 8–12 micrometer infrared radiation) has a measly 2 × 13 infrared detector! It is interesting to note that all the animals, which respond to infrared radiation, are cold-blooded. Perhaps this is because warmer animals

**Box 2.**

The concept of pixel – or 'picture element' – comes from the standard television format: it consists of an array of 512 × 384 dots (i.e. a 4:3 aspect ratio). These dots are graded in intensity according to the number of gray levels. In CCD arrays or in infrared focal plane arrays, the pixel is much more spatially well-defined, because each pixel corresponds to an individual detector, separated from the next by 'dead' space which does not respond to radiation.

A CCD camera contains optics that focuses visible radiation on a CCD array, followed by amplifying electronics. The CCD array, fabricated in silicon, serves both as a detector of visible radiation, as well as a multiplexer (i.e. it performs the serial readout of signals from all pixels of the array). For infrared radiation, the silicon in the CCD is not a sensitive detector, so here the CCD only acts as a multiplexer, while the detection is done by an array made of another material – typically indium antimonide (InSb) (3–5 μ m) or mercury cadmium telluride (MCT) (3–5 and 8–12 μ m). The two arrays of Si and InSb/MCT(making an IR-CCD focal plane array) are mated together by an array of indium bumps, which joins them both mechanically and electrically, on a one-to-one basis.

A CCD or IR-CCD camera does not employ a mechanical scan to cover the complete field of view (FOV) – unlike the human eye which always scans the scene – since each pixel corresponds (one-to-one) to an element of the scene. The signal leaving the CCD and containing the scene information is amplified by electronics, and then subjected to signal processing. The scene may then be displayed on a monitor, and/or it may be further processed by some means such as an artificial neural network.

*Figure 3* shows six 5 × 5 arrays of pixels, each focal plane array covering a FOV of 60°, in order to cover all directions. Such a configuration may be advantageous in a highly hostile or dangerous environment.

**themselves emit more infrared radiation, which being much closer to the infrared sensor than any target to be detected, prevents them from efficiently detecting targets (too large a**

**Box 3.**

Analogously to the compound eye, the minimum dimension $d$ of each visible – or infrared-detector is limited by diffraction, determined by the wavelength $\lambda$ of radiation and the $f/\#$ (pronounced as '$f$-number', and defined as the ratio of the focal length to the diameter of the lens) of the optics in front of the detector:

$$d = 1.22\, \lambda f / D$$

where $f$ is the focal length and $D$ is the diameter of the optics. For example, for 10 μ m radiation and $f/2$ optics, the minimum dimension is about 25 microns. This quantity is much less in the visible region, so that for the same total size of the detector, the number of elements can be more than 10 times higher for visible detectors than for infrared detectors.

background component will override the signal from the target). Analogously, the moment that the bat uses its very sensitive sonar detector, its transmitter must be switched off, or else it will be blinded. This technique is used in radar as the send/ receive method – but it cannot be used in the IR unless the detector – or the animal! – is cooled (since the amount of IR radiated is proportional to the fourth power of the temperature – Stefan's law; for details see *Box 4*). This may be the reason that the animals, which use infrared for imaging, have to be cold-blooded.

At the input end, where light is received, it has been known for some time that insects use ommatidial lenses (see *Figure 1b*) which have graded refractive indexes (GRIN), that couple incident light more efficiently to the rest of the ommatidium.

**Box 4.**

According to Stefan's law, the total amount of radiation emitted by any object at temperature $T$ is proportional to the fourth power of the temperature, in kelvin:

$$P = \varepsilon \sigma T^4$$

where the constant $\sigma$ equals $5.67 \times 10^{-8}$ Watts/m$^2$/K$^4$, and the emissivity $\varepsilon$ is one for a perfect emitter – a blackbody.( Stefan's law is obtained from Planck's law by integration over the wavelength, i.e. it refers to the area under the $P_\lambda$ vs.$\lambda$ curve). For an object ( a black body) at room temperature (300 K), and an emitting area of 1 sq metre, the power is about 450 Watts, whereas it drops to 2 watts if it is cooled to the temperature of liquid nitrogen (77 K).

Even though a snake is cold blooded, it will not decrease the power much – but it will improve the contrast.

Wien's law (obtained from Planck's law by differentiation with respect to wavelength) states that the peak of the emission shifts to lower wavelengths for higher temperature bodies:

$$\lambda_{pk} T = 2898 \ (\mu\text{mK})$$

Most objects which are at ambient temperatures ($T$ about 300 K), therefore have their peak emission at about 10 μm – which happens to fall in the 8–12 μm 'window' in atmospheric transmission. The surface of the sun is at about 6000 K, because of which its peak emission is at about 0.5 μm, which also falls in a window, the visible range 0.3–0.7 μm, that we are accustomed to. There is another window in the 3–5 μm region, which is used to image aircraft.

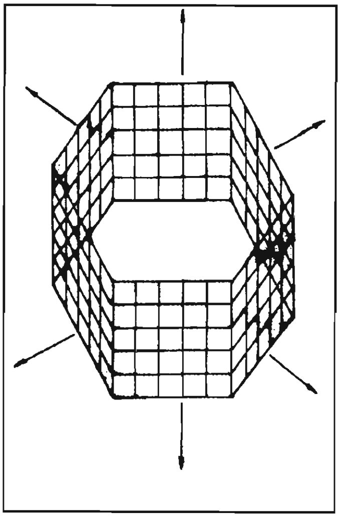


*Figure 3. Indicates one way of realising a large field of view (FOV) – in fact, 360°. Six 5 × 5 focal plane arrays (FPAs) are arranged in a hexagonal configuration so that they can look in any direction. This FOV is higher than that attained by any living creature, and it is a clever way of utilising a flat FPA (with its limited FOV), to do the job of a curved compound insect eye. It may be noted that Figure 3 is schematic since it does not show the CCD or IR-CCD arrays separately, nor does it show the dead spaces (which do not respond to radiation between detectors), nor is the post-processing electronics shown, for simplicity.*

The reason for this high efficiency follows from Fresnel's laws of refraction. If the incident ray of refractive index one (approximately) originates in air, and then passes successively through two media of refractive indexes $n_1$ and $n_2$, the transmittance is maximised if the refractive index $n_1 = n_2^{1/2}$. In general, for more layers, $n_i = (n_{i-1} n_{i+1})^{1/2}$ Humans have also recently mastered this graded refractive index technology, and it is now possible to make GRIN microlens arrays by micromachining techniques.

One may note that the large visible and infrared detector arrays that we are able to make are all flat. These are called focal plane arrays (FPAs) and involve optics that focus the incident radiation on to the plane which is at the focus of the optics, with a one-to-one correspondence between the scene (or target) element and the pixels of the FPA (i.e. you can divide the scene within the field of view into, say 1024 × 1024, elements and the FPA has correspondingly the same number of pixels). For extremely large arrays, which seek to look at a larger field of view (FOV) (shall we call it a bird's eye view?) – as would be needed for satellite imaging of the earth – we will have to make curved focal plane arrays, similar to the bee's eye, which is mounted on a roughly spherical head. Recently, detectors have been made in a geometry such that the imager can simultaneously look in all directions (see *Figure 3*), which is better than the imaging done by any form of life (even birds have a maximum FOV of only about 270° )!

## Image Processing

Needless to say that is not the end of the story. After the detector array has received the signals from the scene/target (i.e.the image has been 'captured'), it must be processed by electronics (i.e. image processing), and finally the image must be 'recognised' (i.e. compared with some pre-existing, stored images or 'templates'). We can pick up quite a few clues from the natural world. Studies of the eyes of bees, crabs and frogs, have shown that the signals from several ommatidia/cells are combined (enhanced or inhibited) in such a way as to enhance features of interest such as edges and contours (curves across which the contrast changes), and target motion. A frog will remain blissfully at rest with the prey right under its nose as long as the prey remains motionless! The field of image processing seeks to achieve what the visual systems of insects/amphibians carry out routinely – image capture, processing and recognition. One approach to image processing, implying the 'recognition' of the image, that is closest to being biologically inspired, is that of artificial neural networks (see *Resonance*, Feburary 1996), in which the network is trained to respond so as to recognise a particular image/target. It is clear that what is needed is a synthesis of focal plane arrays and neural networks, for post-processing of the scenes or targets that are being looked at. This is a rapidly changing field, but one example can serve to illustrate the approach. A 16 × 16 focal plane array was coupled to a neural network with 256 neurons in the input layer, 50 neurons in the hidden layer and 9 neurons in the output layer. The neural network was trained with 27 IR targets (IR images of vehicles, chosen for military applications) to classify them into three types, in three orientations and at three different temperatures. The 27 targets were also rotated 5°, both clockwise and anticlockwise. The neural network identified the targets in 78 out of the 81 trials, wrongly identifying the target type twice, and the temperature, once. Although this is still somewhat simplified in terms of robustness – a larger variety of target types and temperatures and larger rotational angles need to be demonstrated – the approach is definitely promising.

## Advanced Concepts

A question may arise: where exactly is the neural network or other image processor going to be located in the FPA? At present, the infrared focal plane array ( see *Figure 1c* ) consists of a detector array (e.g.16 × 16) which is connected one-to-one with a 16 × 16 CCD multiplexer by means of an equal number (256) of indium bumps. In more advanced concepts, additional layers of circuitry are sandwiched in between the two (detector and multiplexer) layers in the so-called $z$-plane technology. It is called by this name because the previous technology involved arrays fabricated on layers, utilising only the $x$- and $y$-dimensions ($x-y$ array). The more ambitious $z$-plane technology adds all the further processing functions between the detector and muliplexer arrays in the form of additional circuitry mounted on several ceramic boards. It may be noted here that the $z$-plane technology is the most challenging application of processing technology in semiconductors today that seeks to use all three dimensions as real estate. Again, it is clear that such post-detection processing – which is routinely done by living creatures – is being perfected by using semiconductor technology.

Other new concepts include explicitly the notion of 'region of interest'. At present, all the detectors (each corresponding to a different pixel) are treated on an equal footing. But once a target has been identified, some pixels are more important than others! With all multiplexers, one can go for 'electronic zooming' in which one can select only those pixels that contain the target. Similarly, one might want to have greater resolution at the centre of the field of view, and lower resolution towards the edge. With a proviso: One needs a motion detector at the edge of one's field of view. You may notice that in your eye – anything (even an insect) moving at the edge of your FOV causes you to immediately look in that direction: obviously a hard-wired survival mechanism! This occurs because the region of vision that is clear and distinct (focused) in the human eye is restricted to about a minute of arc (related to the fovea), whereas the total FOV of peripheral vision is much larger. This obliges one to continually scan across the scene (your eyes are almost perpetually in motion); whereas a focal plane array 'stares' at the scene – no mechanical scanning is needed – with all pixels being equally distinct. A human who is prevented from moving eyes and head (i.e. without scanning) is functionally blind.

There a few other concepts that emerge from everyday experience. One is the concept of 'sensor fusion' – we integrate the information from our eyes and ears: a fusion of sight and sound.

The pit viper has both visible and infrared sensors and has enabled fusion of the two: in the form of visible signals enhancing or inhibiting the infrared signals. This enables the viper to detect motion and to discriminate the target from background clutter. Similar concepts are being used in military applications to fuse the information from laser radar and from IR imaging. Another is that of stereo: we have two eyes and two ears, needed because we live in a three-dimensional world, and we require information about directions and about three-dimensional objects. Obviously, if we are going to build robots that survive in the real world we have to provide them with all that Mother Nature has designed – and more. One technique that is already being used in satellite imaging is multispectral scanning: The same scene is imaged using detectors that respond to visible, infrared (3–5 and 8–12 microns) and even microwave (millimetre) radiations. In satellite imaging, each band is divided into sub-bands and information collected separately from each.

## Speculations

Let us speculate a bit. If we provide our robots with senses that no living organism possesses, we may send them where no living being has gone: the other planets. Scientists hope to learn from nature in designing an autonomous robot, as will be needed for a robot to explore Mars ('Pathfinder' was provided with two cameras for stereo vision) or descend into Mt Vesuvius (the robot 'Dante' has already attempted to descend into this inferno!). Some engineers are trying to copy the gait of the cockroach – surely an excellent design as anyone who has tried to pursue a cockroach will testify (except when it overturns!). Another topic of research, which will see applications in robotics, and mimics nature, is that of 'smart' skins. These consist of semiconductor sensors embedded in the skin of the robot, vehicle or aircraft which will measure pressure, temperature, pH etc., monitoring its 'health' and enhancing its survivability. Obviously, all these sensory data have to be fused in the robot's brain (without overloading it!) in order for it to accurately respond to its environment.

So if we are to design survivable robots ,we will probably need multispectral stereo focal plane arrays, with sensor fusion of the different spectral regions as well as of additional sensors for sound, touch, smell, X-rays etc, processed by neural networks (for target/scene recognition), and mounted on a (possibly cockroach-style) locomotion system. Of course, there is no necessity that the neural networks will be implemented in silicon as our electronics has been so far – biopolymers or something may be more suitable!

We already can do comparative testing of different schemes that may work. For instance, we already have exhibitions in which different robots compete. The next question is: Can we go the whole hog? If – and this is a big IF – we can figure out how to make our robots design/reproduce other robots, then we could just sit back and let evolution take its course: survival of the fittest. So far we have referred to ways in which human design compares with those perfected by evolutionary processes. What comes next, at the appropriate level of complexity, is to use evolutionary processes to advance our designs of robots, and place them in unusual environments. This may open up new vistas: That we may, finally, come up with something different from what nature has already achieved. Of course, this evolution may converge with that of living organisms, but it may also diverge – especially if we are able to explore and colonise environments, which life is yet to encounter.

*Address for Correspondence*
Vikram Dhar
Solid State Physics Laboratory
New Delhi 110 054, India.

## Suggested Reading

[1] **Richard Dawkins. *The Blind Watchmaker*. Longman. Essex. 1986.**

[2] **R Feynman, R B Leighton and M Sands. *The Feynman Lectures in Physics*. Addison Wesley. Palo Alto. Chapter 35–36, 1965.**

[3] **P Kunze and K Hausen. *Nature*. 231. 392–393, 1971.**

The majority of educated persons are not interested in science, and are not aware that scientific knowledge forms part of the idealistic background of human life

*Erwin Schrödinger*