Naked Eye 3D perception from conventional 2D displays.
by Ian Overington

Summary.
During the past few years there has been a major reawakening of interest in creation of imagery which provides perception of 3D depth. A number of methods of providing such perception have been devised, at least three types of twin lens 3D camera have been manufactured, various specialist 3D viewing facilities have been marketed and there has been considerable documentation produced discussing the pros & cons of various approaches.

In contrast to this there seems to have been little attention paid to the exact mechanisms within the human visual tract which lead to the actual perception of 3D. Having developed a wide knowledge of the fine details of early human vision during some 40 years of vision research, I have felt it desirable to attempt to answer this question. As a result of this I believe that I have developed an alternative method of handling input stereo pairs of images (or video streams) which permits substantial perception of 3D from what are essentially resultant 2D images or video streams without need for any form of visual aids. These forms of output imagery retain the full resolution, colour & brightness properties of the original individual input images, whilst also being capable of being presented in virtually any form which is normally available for 2D image presentation!

This report starts with a brief overview of the current general facilities for 3D viewing, together with a summary of some aspects of 3D image capture which are important to appreciate for anyone wishing to use the new generation of 3D cameras. It then discusses the visual knowledge background from which the new naked eye technique has been developed, together with the basics of the technique itself.

Introduction.
The possibilities of providing acceptable perception of 3D depth for home or cinema viewing has from time to time received considerable attention over some 150 years but usually each upsurge of interest has relatively quickly died down. However, of recent years there have been major developments in the fields of both digital image capture & handling, refinement of design & control of optical equipment and widespread use of (powerful) digital computers. These factors are believed (presumed) to be the main reasons for a major reawakening of interest in creation and presentation of imagery which provides perception of 3D depth. A number of methods of providing such perception have been devised - for both computer gaming, cinema presentation, home TV viewing and personal experimental activities.
Summaries of this historical background have been provided, amongst others, in Refs. 1 & 2. In addition a rather comprehensive coverage of the creation, handling & projection of such material as relevant to the modern cinema industry is provided in Ref. 3. This is a readily accessible & easily readable book primarily covering the commercial production of 3D films, but taking the reader comfortably from the basic considerations involved in starting to create 3D imagery right through to the presentation of the final product.

At the same time, particularly in the last year or two, there have been a number of articles in the Daily Press which have drawn attention to some problems claimed to have been experienced by observers, either at 3D cinemas, when using 3D TV’s or in 3D gaming (e.g.
Refs. 4 - 6). In particular these problems have apparently included development of headaches &/or dizziness after several minutes of concentrated viewing or a total inability to perceive any 3D at all! In one recent article the former of the problems has been suggested to be possibly due to requirement for unnatural (forced) differential eye movements to attempt to generate convergence relevant to a variety of real world viewing distances whilst actually accommodating for a fixed distance of the projection screen. This possible explanation (from a time some two years ago, before the majority of recent 3D movies & 3D TV presentations had become widespread) is given some expanded treatment & speculation in Ref. 3. Very recently some similar concerns have also been expressed with regard to the very new Nintendo 3DS gaming equipment (which employs a lenticular viewing screen). On the other hand, inability to achieve any 3D perception is apparently recognised to be a basic problem for some 10 to 15% of the population (which is presumably not realised in most cases until the person attempts to perceive the difference between a 2D & 3D cinema or TV presentation)! Such persons are accepted to be 'stereo blind' (Ref. 6).

Yet again, a recent article (Ref. 2) indicates that there are currently a considerable number of further research projects underway. These projects are apparently mainly aimed at finding ways to use newly evolving electro-optical techniques in order to provide displays which can present directly the necessary 3D imaging data to the observer's two eyes (without any form of glasses). The ultimate aim will then be to avoid need for an observer to use any type of (potentially restrictive) visual aid in order to perceive 3D in display imagery or to be restricted in viewing position.

Despite all the foregoing, as far as I am able to ascertain, there has been no overall assessment in open literature of how & why it is that any of the suggested approaches - existing or speculated - might (or should) work (other than the basic & obvious facts associated with the geometry of binocular vision). To attempt to redress this situation the present author has recently endeavoured to set down both a summary of the various approaches made over the years, the relative success & limitations of them and the relatively recent growing understanding of the functioning of the overall human visual system which must, after all, be the final stage in the perception process (Ref. 1).

This present report aims to take that general overview considerably further by considering in some depth the actual workings of the human visual system and how one can draw on some of these in order to simplify the requirements for satisfactory 3D perception from any form of conventional 2D display facility.

Some observations regarding the conventional 3D scene.

Image capture.
It seems (from Ref. 3 and also from the user manual of the Fuji REAL 3D camera which I am proud to possess) that the conventionally accepted closest foreground object for good general 3D scene capture should be at around 2.1 metres. This implies that the maximum generally acceptable parallax difference (the measure apparently normally used for 3D image assessment nowadays) relates to the change in parallax between a closest foreground at around 2.1 metres and infinity. In turn this equates to a change in angular subtense at the observer's eyes of approximately 2 degrees. Considering this in terms of the observer, it may be thought of such that an acceptable 3D scene generates differential angular subtenses between the two eyes of +/- 1 degree relative to the mean convergence. I believe that this concept of +/- 1 degree relative to the mean convergence should be considered as the basic

Naked Eye 3D perception from conventional 2D displays.
yardstick for assessing any form of 3D scene, since everything relating to imagery within the visual system must essentially be related to this.

From the foregoing it is readily possible to compute exactly what differential distances can be tolerated within the same eye-related angular limitations if one wishes to arrange for satisfactory 3D perception with objects closer that the foreground minimum of 2.1 metres. Some of the implications are rather dramatic. For instance, if the closest foreground is to be at 1 metre the maximum background distance should be no more than about 2 metres. Even more critical is where the closest foreground required is at 0.5 metres, when the maximum acceptable background distance will then be only about 0.75 metres! When planning to capture stereo pairs of images it is all too easy to forget this important fact - which is primarily controlled by differential angles subtended at the left & right eyes rather than any other factor! However, this specific subject is not the main purpose of this report, so it will not be developed further herein.

Some relevant observations concerning Human vision.
In order to appreciate the practicalities associated with 3D (stereo) viewing from the observer's standpoint it is useful to carry out a simple observational test.

Firstly find a scene which contains a suitable foreground object which can be viewed from around the closest recommended foreground distance (2.1 metres), together with some prominent object substantially in the background which is partially occluded by the foreground object. Then, whilst concentrating on the foreground object, alternately cover the left and right eyes. Try to note the relative relationships between the foreground & background object. Assuming that you have selected a background object substantially further away than the foreground object, you will observe that the relative apparent position of the background object moves very considerably sidewards. This is an indication of the visual parallax existing between the foreground & background objects. From this it follows that, whenever one is concentrating on an acceptable foreground object, the relative background presented to the left & right eyes is substantially different - by up to about +/-1 degree in angular subtense! But the foveal resolution of the human eye is of the order of 2 arc minutes under most ideal viewing conditions (i.e. around 1/30 degrees). Hence it is likely that parts of a typical visual 3D scene which we observe happily will have an angular disparity (equivalent to parallax) of at least 30 times the visual resolution for normal high resolution vision!

Now it has to be generally accepted that, for an observer with two good eyes, a high resolution image from each eye will continuously be transmitted along the optic nerve to the brain. However, it is argued by the present author that there is no way that the brain can sensibly fuse such disparate parts of images satisfactorily in terms of a high resolution image. Hence it must be accepted that we regularly live with a situation where our (3D) visual world, in terms of classical concepts of lines & edges, creates large parts of the scene which are effectively substantial double images. An attempt to illustrate this important fact is presented as figure 1. For this scene the original full colour images were first converted to 8-bit greyscale (i.e. monochrome) images, the resulting pair of greyscale images then being overlaid using red only for the left-hand image and cyan only for the right-hand image. In the resultant overlay the (very substantial) variation of parallax from extreme foreground to extreme background should be clearly seen, this being most clearly evident where dark objects are
viewed against a light background. In these cases the left-hand image is substantially to the left in the composite for the foreground (as evidenced by the red ribbons). Conversely the distant objects have the right-hand image substantially to the left (as evidenced by the cyan ribbons), whereas the 'top hat' topiary in the middle distance shows virtually no coloured ribbon (this being roughly the chosen fusion range).

In normal vision, we seem to have learned to suppress one or other of the two disparate images (presumably retaining the image created by the 'master' eye) in order to retain a sharp impression of the whole scene. I realise that one might argue that much of the scene will be some angular distance away from the immediate vicinity of the local object on which we are, at a given instant, concentrating - and hence will be receiving little conscious attention. However, I would counter with the fact that such an argument is inadequate to explain totally the general 'cleaness' of our visual imagery.

Let us now consider things a little deeper (for substantial discussion & illustration of what follows the reader is referred to the section concerning the hierarchical structure of the human visual system in Ref. 7 (pages 160 - 162) or to several Chapters of Ref. 8 - of which an electronic copy is available on Ref. 7).

It must first be realised that our high resolution vision is controlled only by largely opponent interactions between the R & G retinal cones (i.e. primarily by the red through green portions of the visible spectrum). The primary result of this (driving our high resolution vision) is the creation of two 2-dimensional arrays, one of local R inputs & one of local G inputs. However, each of these inputs is inhibited by a surrounding mixture of (R + G) inputs - simplistically red inhibited by yellow and green inhibited by yellow (as opposed to a true opponent inhibition between R & G). Then in turn the local (R + G) groupings are pooled and themselves inhibited by larger groupings, whilst pools specifically consisting of R or G inputs are also created. The inhibitory interactions between these various pools of R, G and (R + G) appear to yield secondary & tertiary channels of progressively lower resolution, each still basically driven from R & G inputs. On the other hand, it is now well known that there is a further opponent mechanism at work in the visual system - generally known as the B / Y mechanism. In turn, this is driven by very large pools of mixed (R + G) retinal outputs (equating to yellow or Y) interacting with very sparse B retinal receptors (Ref. 9). These B receptors receive very out of focus, predominantly blue light. This grossly blurred input imagery falling on sparse B receptors is seen by the present author as nature's way of producing a very low resolution blue imagery which obeys the rules for optimum signal / noise transmission for a sampling system! In support of this observation it has long been known that our perception of very small patches of predominantly blue light are not seen at all under controlled experimental conditions (foveal tritanopia; small field tritanopia - e.g. Ref. 10). Yet
our overall impression of the visual world in normal viewing is of quite a rich blue content (even for small objects)!

For many years I had some difficulty in imagining exactly what purpose this very low resolution channel could have, other than providing some form of very coarse source of reference levels and a probable ready means of providing coarse approximate initial fusion of a new area of visual interest (see Chapters 9.9.3 & 10.2 of Ref. 8). It is now my firm belief that it is this very low resolution channel which also controls our awareness & perception of 3D depth at a cortical level. However, in normal vision we are normally hardly aware of the mechanism! It is extremely difficult to prove anything about this situation but, since involving myself deeply in 3D experimental studies over the past year or so, I believe that I have developed a personal ability to 'feel' this third dimension consciously (in natural viewing). I say 'feel' because the physical effect is little more than a feeling of gentle 'pulling' of my eyes to nearer or more distant objects. Certainly I have now become acutely aware of the difference in perception of any natural 3D scene if I alternately cover & uncover one of my eyes!

A section in the book '3D Movie Making' (Ref. 3) encourages anyone wishing to become deeply involved in 3D image handling to experiment in an endeavour to develop what I interpret as just this personal 'feel' for 3D. One matter relevant to this observation concerns a statement in Ref. 2. There that author makes a statement which infers that, in natural 3D viewing, one of the factors concerns a subconscious (automatic) focussing of the eyes as we change our gaze. However, I would argue that, particularly in good light where the eye pupils are very small, at the viewing distances associated with the majority of natural real world viewing (beyond a few metres) there will be little if any need for any refocusing - conventional depth of field will cater for any theoretical adjustments! Rather, I suggest that the concept of focussing is most probably being confused with my description of 'pulling' - this effect then being associated with relative convergence demands on the very low resolution system at some deep neural level (most probably the Optic Chiasma) rather than any effect on the high resolution system!

Currently available commercial viewing systems.
At this point it is appropriate to consider how well currently available systems for viewing of 3D fit in with the foregoing facts about human vision. It would seem that the main aim must be to find a means of presenting 3D image data such that it is possible to develop what I have described as the 'feel' for 3D without providing obvious dual high resolution imagery. This aim appears to be fulfilled by the old (19th century) side by side presentation of stereo pairs (since there the two true single images of the pair are provided separately). But this is only achieved if a suitable viewing facility is available to produce comfortable overlay of the image pairs for other than short time viewing. Such a viewing facility, as far as I can conceive, must essentially contain (in its optical viewing paths) something in the nature of a pair of narrow wedge prisms (essentially as in the original Holmes Stereoscope - see Ref. 1), together with a suitable means of at least encouraging the correct viewing alignment in the vertical plane. I have given some considerable thought to this subject for a variety of electronic viewing situations (and carried out practical viewing experiments where appropriate) and have concluded that the viewing optics necessary are not all that easy to arrange (although presumably not impossible). Indeed, I now believe that the total construction of the Holmes Stereoscope (with its central baffle and the 'goggle' like eyepiece) is of great importance in readily achieving the necessary overall alignment necessary for comfortable viewing of such side by side stereo pairs! It is also pertinent to comment that, if suitably scaled side by side
image pairs are created from any modern high resolution, full colour 3D image pairs (such as can be captured using the Fuji REAL 3D camera), the overall 3D viewing experience is superb! In fact, the said experience is, to me, superior to any other currently commercially available 3D viewing facilities (for viewing still images as hard copy, of course).

The various types of anaglyph presentation (see Ref. 1) largely overcome the dual high resolution image problem as well as the orientational alignment requirements - by having the pair of images overlaid on top of one another whilst attempting to present only one high resolution image to a particular eye. They also in principle essentially overcome the limitations of only hard copy still imagery of the Holmes Stereoscope. However, I believe that it is only by effectively suppressing one of the two high resolution images as transmitted to the brain that a satisfactory fusion will be obtained (effectively similar to what I have already discussed regarding natural binocular vision). The fusion capability itself must then be somehow controlled effectively by a low resolution visual channel. In the case of the red / green or red / cyan anaglyphs I now believe that it will be the low resolution pools of R & G receptor outputs which will control the 3D depth perception. On the other hand, in the case of the yellow / blue anaglyphs essentially only a single high resolution image will be transmitted to the brain - usually that received by the left eye. This is because in this case the image actually perceived by the right eye through the deep blue filter should only essentially comprise of the very low resolution image processed by the sparse B cones. So in this case the actual 3D depth perception will be achieved from the interplay of the low resolution (R + G) pools from the left eye and the B pools from the right eye. In all anaglyph style 3D presentations one other important fact must be realised - that the effective impact of the 3D (including such effects as the foreground appearing to come out of the screen) are totally dependent on the choice of range of fusion selected. This itself must be determined during the original creation of the anaglyphs - basically by adjusting the mean parallax. Once this is set, any subsequent handling & viewing of the resulting anaglyph will always be constrained by this range of fusion - but very effective changes in 3D scene appearance when viewed can be created by changing the mean parallax setting during anaglyph creation (or by creating additional anaglyphs from the same original image pair). Such manipulations prove to be exceedingly simple for both still image pairs and stereo video clips using two pieces of free software available for download from the Internet (Refs. 12 & 13).

Other currently available viewing methods (see Ref. 1) - such as the recent 3DTV offerings and the lenticular screen displays of the Fuji 3D camera monitor and the new Nintendo 3DS gaming equipment do appear to have found a way of avoiding any form of viewing aid at the observer's eyes - but only if one has available the dedicated viewing screen and even then only if viewed at a fairly critical angle! As with the anaglyphs, all these methods are also totally dependent on a selected mean fusion setting. In the case of the Fuji camera monitor the built in software makes its own attempt to optimise this setting. However, this setting is not carried through to the saved image files - but the software of Refs. 12 & 13 provide ready facilities for both automatic & manual adjustment.

All the foregoing leads straight on to my own considerations - based on the assumption that the Colorcode3D approach is essentially based on the yellow / blue opponent channel (see Ref. 1, also Ref. 11). Wouldn't it be nice if one could find a way of starting with the approach followed by them, but somehow 'hiding' the additional information necessary for the low resolution signalling of 3D depth, in order to avoid the need for coloured filter glasses.
The proposed optimum solution.
Initial considerations.
Starting from the Colorcode3D approach of working with the yellow / blue opponent colour concept, I initially asked myself the following question:-
"Since the yellow / blue opponent system is of very low resolution, instead of generating yellow / blue anaglyphs which exhibit sharply defined yellow & blue bands, would it be possible to extract the necessary twin image data, blur them very strongly and then add them back into a single high resolution image?".

In order to answer this question I considered it first necessary to attempt to estimate what the actual properties of the yellow / blue low resolution neural channel are likely to be. As a starting point I went back to my considerations of many years ago (as published in Chapter 2.5 & 2.6 of Ref. 8). There I have discussed at some length the fact that there are four concentric neural layers created from the original receptor matrix in each retina, these having measured properties (particularly Fig. 2.7 of Ref. 8) which roughly conform to a scaling of 1:3:9:27. Let us now invoke the concept that, for optimum signal / noise transmission, all these neural layers should have effectively similar blur properties with appropriate scaling. But the high resolution channels have long been considered to have a blur of roughly Gaussian form and with a standard deviation (SD) of between 1.3 & 1.7 receptor spacings. So for similarity the lowest resolution channel should have a blur SD 27 times as large or around 40 receptor spacings. What this means (roughly) is illustrated in graphical form as figures 2 & 3. Bearing in mind that the foveal receptor spacing is of the order of 0.7 arc minutes, it can be seen on these figures that the assumed blur is indeed very large - amounting to well over 1 degree!

Consider now a typical modern digital image resolution. It is quite common for a modern camera resolution to be in the region of 3600 to 4000 pixels for a standard field of view of the order of 60 degrees. Hence each pixel will subtend around 1 arc minute in angular terms - that is, very close to even the (highest) foveal resolution of the human eye, let alone more peripheral areas. It did therefore seem reasonable to assume that my premise of superposing some very blurred difference data from stereo image pairs might well work. However, how blurred should 'very blurred' be?

The input image / observer's eye interface.
There are usually a number of steps between the original capture of digital images and the final presentation of processed imagery as falling on an observer's retinae. For conventional forms of image processing it is usual to pay little concern to the intermediate processes, as long as

![Figure 2. Estimated potential spread function of the very low resolution vision channel.](image1)

![Figure 3. Estimated potential edge blur due to the very low resolution vision channel.](image2)
the best (or adequate) resolution is maintained for the final image. However, as soon as any suggestion of introducing substantial blur is considered, it becomes important to assess to what extent this blur will interact with the (fixed) neural blur present in the observer's low resolution visual channel. A little thought should permit it to be realised that this interaction of the two blur functions must be directly dependent on the relationship between the field of view (FOV) of the camera lens used for the original image capture and the FOV of the final processed image as projected onto the observer's retina. If the presented FOV is small, then the imposed blur will be relatively small as viewed. However, if the presented FOV is large the imposed blur at the retina may be rather large. This imposed blur may then be expected to interact with the (fixed) neural blur in a complex way. On the other hand, if the imposed blur is only small, not only may significant residual ghost edges still be perceived but also for large amounts of parallax the left & right images of a given edge may not be merged together. It is thus highly desirable for both reasons to employ the highest level of blur which will work adequately. In order to attempt to explore the potential of the proposed blurring, it was therefore necessary to make some assumptions concerning what might be acceptable viewing situations.

To aid with this I drew rather broadly on some important findings during the 1970's regarding optimum information transfer from raster sampled imagery. These concluded that, for optimal information transfer from sampled images, the individual pixels in the displayed image should subtend between 2 & 4 retinal receptors at the observer's retinae (Ref. 14 - electronic copy available on Ref. 7). I therefore made an initial assumption that, for optimum conventional viewing of typical high resolution (2D single) sampled images captured by modern digital cameras, the presented FOV should be around 120 degrees (i.e. such that individual pixels would subtend between around 3 retinal receptor spacings). However, typical viewing is likely to be with FOV substantially less than this. In fact, I have noted that, for typical presentation of modern digital images on relatively modern computer screens the zoom level for complete image presentation is around 25%. From this I argued that it was possibly reasonable to introduce blur which, at 1 : 1 pixel / receptor spacing, would be of the same order of magnitude as the assumed blur existing in the low resolution visual channel. Working on from this I came up with a need for blurring of the left / right difference images from a stereo image pair having a Standard Deviation of around 28 pixels. A check on the effects of this using a geometrical image yielded a comparison of pixel-related blur for a 40 pixel parallax and retinal receptor blur for the low resolution visual channel as shown in Figure 4. This was considered satisfactory for further practical studies.

On initially trying out this sort of approach on realistic images I found that I did seem to obtain a perception of 3D without the need for any sort of glasses, but there was still some (albeit grossly blurred) semblance of local blue & yellow ghosting. Whilst this was far less objectionable for viewing without glasses than the clearly defined blue & yellow bands.
characteristic of the yellow / blue anaglyphs when viewed without glasses (and may not even be noticed by the untrained observer), it was nevertheless not really acceptable.

I therefore next questioned as to whether I really needed to separate the yellow & blue parts of the spectrum at all. Was it not possible to extract the entire spectral data from each image and generate some strongly blurred differential data from this to add back in? The objective here was to attempt to provide each eye with the same composite image, whilst at the same time presenting both eyes with an effectively subliminal representation of the complete (multispectral) low resolution differences between the left & right views. The basic image would then have all the high resolution properties (resolution, brightness, colour etc.) of one of the original images of the stereo pair (conventionally the left view). However, the subliminal difference image must not be of such extremely low resolution as to significantly affect the perceptual appearance of this difference image as received at the cortex. This might sound to be an impossible requirement. However, as previously discussed, the low resolution blue / yellow opponent channels of the human visual system are of the order of 30 times lower than the high resolution channel. Therefore it was believed that, considering the mathematical properties of convolution of roughly Gaussian envelopes and the really very extreme blur, it was possible that such a pair of requirements could both be met with reasonable comfort!

In practice it was found possible to achieve the desired results by the following sequence of processes (it being assumed that, as with anaglyphs etc., the necessary stereo pair have been adjusted using facilities such as Ref. 12 in order to create a satisfactory mean fusion setting).

- Take copies of the left & right pair of input images;
- Subtract the right from the left, pixel by pixel;
- **Split** the resultant difference image into positive & negative components, convert the negative going data to positive (scalar) and subsequently save the two components as 24-bit positive distributions (RGB levels of 0 to 255). This concept is effectively the same as that used by me in my simulation of conventional, high resolution vision (and discussed at some length in Chapter 3.7 of Ref. 8). As in that case, I would argue that such a splitting of positive & negative portions of difference signals is a necessary requirement in practical vision in order that the difference data can be successfully transmitted along the optic nerve! Again as in that case, it would seem that the act of splitting into the two components itself leads to subtle differences from the conventional handling of vector data. At the same time, splitting into two components permits more sensitive and simple handling of the difference data for the necessary processing;
- Scale down the resultants by a linear factor of 4 (for easier processing), then blur strongly in order to mask the discontinuities associated with left & right locations of individual edges. It has been found that a Gaussian blur of effective standard deviation (SD) of approximately 28 pixels at this stage seems to provide a good tolerance of local parallax (i.e. practical handling of substantial local parallax between left & right images of individual edges) whilst still yielding a satisfactory level of 3D perception. I would, however, hasten to add that it might well be possible to optimise this by substantial further experimental studies. [N.B. It is considered that, in order to achieve a good approximation to true Gaussian blur with such an extreme SD in practice, it is desirable to run a much lower SD blur function several times. SD=2.3 run 12 times in sequence has been found to yield a good approximation];
- Rescale both outputs from the latter process. Then **add** the output for the original positive partial difference to the original (sharp) left image (as a true pixel by pixel
summation) and **subtract** the output for the original **negative** partial difference (also as a true pixel by pixel process);

- Save the resulting composite image, appropriately named.

The resulting image visually **looks** identical (in full colour, brightness & sharpness) to the original - and **is** identical **geometrically**! However, superimposed on this genuine geometrical image is a grossly blurred subliminal representation of the local differences between the left & right images. This will be displaced either to the left or right of the geometrical image by small amounts depending on whether the local edges are beyond or closer to the observer than the selected plane of fusion (i.e. zero parallax) - by effectively the half width of the local coloured bands in an anaglyph representation! Although not at present thought to be necessarily relevant to the current topic, it is considered worth drawing attention to a substantial body of experimental data concerning local subliminal effects on visual thresholds which were discussed at some length in an earlier book by the present author (chapter 13 of Ref. 15 - available as an electronic copy on Ref. 7). It seems to be just possible that some of this earlier work may help to explain more fully why the present approach to naked eye perception from 2D (flat) composite images seems to work!

An attempt to illustrate the practical effects of interactions of local parallax between left & right image at individual edges and the level of gross blur introduced as proposed above is provided in figures 5 to 9. Figures 5 & 6 illustrate the practical distribution of the blurred difference trends for a high / low brightness edge and a low / high brightness edge as a function of magnitude of parallax for edges beyond the range of fusion, whilst Figures 7 & 8 illustrate similar situations for edges closer than the range of fusion. It will be seen that the general characteristics are similar. However, there are two important differences:

- for figures 5 & 6 show progressive shifts to the right with increasing parallax, whereas in figures 7 & 8 the corresponding shifts are to the left;
- although not obvious from a cursory look, the trends shown for a high / low brightness edge originate from the opposite partial differences in figures 5 & 7, whilst the reverse situation exists in figures 6 & 8. This inversion of characteristics is believed to be part of the in built coding relating to polarity of brightness change at an edge and the 'polarity' of the parallax.

**Figure 5.** Blurred difference trends for a high / low brightness edge at long ranges.

**Figure 6.** Blurred difference trends for a low / high brightness edge at long ranges.
Following on from these illustrations, an attempt to provide an illustration of a composite effect from compounding the difference trends with an original input for a simple geometrical block image exhibiting various levels of parallax is presented as figures 9 & 10. There it will be seen that (in this case for an object at greater range than the range of fusion) the effects are:

- at a high / low brightness edge to lower the overall brightness locally;
- at a low / high brightness edge to raise the overall brightness locally;
- both these effects being asymmetric (shifted to the right) with relation to the physical edge location. For objects closer than the range of fusion a similar asymmetry will occur, but shifted to the left rather than the right! Additionally, in the situation illustrated in figure 10, there is a flattening of the top of the overshoot, this being due to limiting at maximum grey level. Such flattening would not be a characteristic for lower contrast edges and lower maximum basic brightness levels (which tend to be by far the more common practical situations)! It is, however, of some importance to note, since human vision is very acutely aware of both sudden local changes of brightness and local sudden changes of rate of change of brightness! Thus any tendency to 'clipping' due to overshooting of maximum grey level will tend to create some measure of residual ghosting!

I believe it is the subliminal components which provide the evident 'pulling' perceptual sensation which I have come to recognise as the main signal of 3D depth. In other words, when the composite (essentially 2D image) is viewed there is a strong perceptual sensation of
3D depth (much as when viewing a real 3D scene) when viewed without visual aids of any kind, whilst for a wide range of situations there is no obvious evidence of either double imagery or ghosting! Of course, if the parallax situation is too extreme there is then evidence (to the trained eye) of a form of very blurred ghosting at the extremes of parallax, but such ghosting is of local 'natural' colouring and, as such, is hardly evident if one doesn't know exactly what the scene should look like in 2D. Also in such circumstances one is dealing with a situation where, in natural viewing, one would equally be potentially aware of some form of defocused double imagery if concentrating sufficiently.

As already stated, the composite images created by the foregoing method do not differ geometrically in any way from the original left-hand image on which they are built. That is to say, the geometrical layout of edge & line locations, if carefully measured in a graphics program, will be identical (and, of course, are totally free from double imagery). So the only differences are in subtle brightness distributions in the areas close to strong edge contrasts. As such, whilst one gets what I consider to be the perceptual 'pull' as one's gaze moves to different 3D depths within the scene, it is felt that there should be little chance of unpleasant side effects due to prolonged viewing of such imagery! After all, both eyes are receiving the same (essentially 2D) geometrical image and therefore there should be no binocular rivalry in the accepted meaning of the term - that is, both eyes looking at a flat screen or other flat display and never themselves requiring to attempt to readjust optical convergence for different parts of the scene! At this time there is no significant evidence to suggest that there just may be some residual cognitive conflict, but at least it is considered that any such conflict should be minimal.

One might question just how such a processing can work - bearing in mind that both eyes are now receiving identical imagery and there appears to be no tangible differential information between the outputs of the two eyes for the brain to interpret. I believe that the answer to this most probably lies in the fact that the stereo and motion sensing of the visual system are extremely similar and closely related (see Chapters 5, 9 & 10 of Ref. 8), such that it seems perfectly reasonable, given that sufficient monocular information is provided, for the high level processes associated with each eye to be able interpret such 'monocular' 3D depth data (possibly with minimal training?). Furthermore there is ample evidence from modern optometry that the optical parts of the brain are readily able to adapt to much more severe changes than this - for instance when considering the substantial distortions of the binocular image fields associated with wearing varifocal spectacles (very widely worn by older people and worn continuously by me personally for some 30 years with no apparent ill effects!).

Should the reader wish to understand the overall processes associated with stereo (and motion) more fully, the subject has been documented in great detail in Ref. 8 (Chapters 5 & 9 for motion & 10 for stereo). I believe that virtually the entire treatments developed there for the high resolution system (including sensitivity) can be applied directly to the lower resolution channels just by applying a simple appropriate scaling factor. This factor is believed to be approximately 9 for the R / G channel and approximately 27 for the B / Y channel.

**Practical considerations.**
In order to carry out the processing described in the previous Section it is necessary to employ some processing steps which seem not to be available in any known commercial image processing suites. Hence batch processes have been designed to run a series of personal component bits of software to yield the desired end product for both stereo pairs of still
images and stereo video clips. A large number of stereo pairs of stills (in excess of 200) and a small number of short video clips have been run through these batch processes. However, the said processing is at present by no means turned into a commercially viable & available facility. When and whether I shall find myself able personally to take this extra step is by no means clear. I would, however, be delighted to hear from anyone who might see themselves able to undertake some form of commercialisation in association with me. E-mail communication can, if desired, be achieved through my website, any such communication being forwarded directly to my routine personal e-mail facility.

All the foregoing - and all illustrations to be provided later - are concerned with pairs of still images. It is perhaps relevant, at this point, to add just an observation relating to stereo video clips. For the latter, one will typically be working with input (capture) resolutions of roughly 25% of those for modern still images. It has been found that, by ignoring the 4:1 scaling down for blurring processes, the general results are very similar (and acceptable) for the limited number of stereo video which have bee processed & viewed. There is therefore no reason to assume that, by using appropriate processing sequences, the procedures described herein will not work equally successfully for video clips (and, by inference, for films and TV presentations!).

Some practical examples.
It is time now to present a small subset of examples of composite 3D images created as described earlier. In making this small selection the reader must realise that each and every individual real world image tends to be relatively unique in its subtle details. Therefore it is quite difficult to choose a small group of images which can between them illustrate a wide variety of situations. Because of this I have tried to pick a few versed types of original image which hopefully will serve to demonstrate a few of the main points.

A first observation which must be made is that a great deal of 3D in the real world is so minimal in extent that it is very difficult to be sure of how much one perceives is real and how much is intuitive. After all, everyone who has been used to looking at 2D images to any extent will have a built in concept of what should be 3D in a given scene. The only main difference in many situation is the perceptual 'pull' which I have tried to describe - and in many cases even this is difficult to be sure of. Hence it is necessary, in order to provide any sort of credible demonstration of my efforts at generating images with inbuilt subliminal perceptual 3D, that I try to find the relatively rare, reasonably pronounced 3D effects. Even then, I have no personal knowledge whether any observers with other than perfect binocular 3D perception will be able to perceptive what I can perceive personally (after considerable training).

Figure 11. A 3D image for naked eye viewing generated from a stereo pair of images captured by a Fuji 3D camera.
The first image which I hope will provide a significant 3D experience is shown as figure 11, this being derived from a stereo pair of full colour originals similar to those from which figure 1 was derived. This stereo image pair has a wealth of substantially high contrast, cleanly defined structure at a wide range of distances and the original matching of parallax was selected such that the range of fusion was in the middle distance. It is hoped, therefore, that this will provide a fairly good perception of depth. Where this report is being viewed electronically, for full appreciation of the subtleties of this and succeeding images, benefit will be gained by zooming to around 200%.

The next image (figure 12) is from a set of stereo images captured at a miniature village, again yielding a variety of substantially contrasty objects at a variety of distances.

Figure 12. A typical 3D image generated from a stereo pair of images of a miniature lakeland village.

Figure 13 is of a natural outdoor scene which contains a deep valley in the middle distance. It is considered that the depth contouring of this valley and the rising ground beyond serve to demonstrate a subtly different type of 3D perception.

Figure 13. A 3D scene containing a deep valley in the middle distance and rising ground in the far distance.
Yet another type of general scene is shown as figure 14, this being looking along a typical central old town street. In addition to the natural receding trends due to the two rows of old buildings, added depth interest is provided by the arch in the middle distance, by which the more distant parts of the street are framed.

Figure 15 is included to illustrate that it is possible to generate successful 3D images in portrait format by capturing stereo pairs of images with a twin lens camera such as the Fuji REAL 3D camera held vertically. On the face of it such a process would seem to conflict with the normal concept of stereo. That is, one assumes that the observer will have his head roughly erect (his eyes therefore being horizontally displaced) when viewing, whereas the stereo created by a twin lens camera on its side will be along a vertical axis!

Nevertheless, it was found that, when using the StereoPhotoMaker software to separate the left & right images from the composite MPO file created by the camera, it was possible to create at least red / cyan or yellow / blue anaglyphs. In these, despite the parallax in portrait form being vertical, when viewed through the appropriate filtered glasses, the 3D was correct! Subsequently it was found that, by careful attention to the exact sequence of processes and the point at which a 90 degree rotation is carried out, it is also possible to create a successful 3D image in portrait format for naked eye viewing.

On thinking about this, it was no great surprise to me, bearing in mind that, as discussed at some length in Chapters 5, 9 & 10 of Ref. 8, the human visual system appears to have extremely similar processing relating to both motion (which can be in any direction) and stereo. Additionally, as has been discussed in Chapter 2.5 of Ref. 8, the entire midget processing system of primates in general appears to be

Figure 14. 3D image of an central old town street.

Figure 15. A 3D image in portrait orientation created from a stereo pair captured with a twin lens camera held with the lenses displaced vertically.
'built onto' a more basic back-up system which is extremely similar to that of many lower animal species - and which is primarily sensitive to motion rather than form!

Returning to observations relating to figure 15, the sort of view presented there calls out for a portrait format in order to provide the imagery of the full extent of the stream without a lot of spurious adjacent clutter. I have found that very effective results can be achieved with several similar scenarios.

One final aspect of practical 3D photography of which I wish to demonstrate the practicality is the situation discussed much earlier where one wishes to capture depth detail on subject matter which is at much closer proximity than the recognised conventional standards for 3D image capture. In figures 16 & 17 I have attempted to illustrate one such situation.

In this case I wished to present the rather fine, subtle depth within a spectacular epiphyllum flower, where the main subject needed to be at about 30 cm from the camera. In this case the basic pair of images captured by my Fuji 3D camera were wildly mismatched - well beyond the automatic parallax matching provided by the built-in software. However, by careful use of SPM (Ref. 12) I was able to adjust the parallax such that the central portion of the flower was fused. This still left very substantial parallax over the majority of the image - as I have attempted to illustrate by a monochrome montage as figure 16 (where the left image is depicted in red and the right image in cyan). In this montage both the fusing of the centre of the flower and the large residual parallax over much of the image should be clearly seen. Nevertheless, when the optimally matched image pair was processed through my batch processes the results were as shown in figure 17. I think that it should be agreed that there is substantial evidence of 3D there, whilst even the (very mismatched) background has become acceptable (in fact, much as in real life viewing when concentrating on the flower itself). I consider that this is an ample demonstration not only of the powerful capabilities of my approach to 3D presentation, but also a confirmation that a basic 3D camera such as the Fuji REAL 3D can be comfortably used for reasonably close-up 3D image capture (say as close as a foreground at about 50 cms) even without recourse to the clever ADV 3D macro facility!
A bonus which has been found with virtually all the 3D images created by my processes is that the 3D effects are robust when the images are zoomed (or scaled), it being possible within broad ranges to retain the (appropriately scaled) amounts of perceived 3D!

Conclusions.
It is hoped that the foregoing dissertation is both adequately clear and useful in attempting to take forward practical 3D viewing to a new level. This new level of 3D viewing is such that, once appropriate output image files have been created, viewing can be readily arranged on virtually any 2D electronic display or as any form of standard hard copy - i.e handling of the final images (both stills & video clips) is essentially the same as any 2D imagery! So output material could be readily displayed on any standard computer, via any conventional 2D TV system, conventional postcards, brochures etc. where appropriate. At this time I have shown personally that all the necessary components for practical 3D viewing without any visual aids are now available. All that remains is for someone to pick up the concepts and convert the necessary software routines into a viable commercial (essentially automatic) procedure - possibly somewhat along the lines which have been incorporated into the freeware 'StereoPhotoMaker' (Ref. 1) and 'StereoMovieMaker' (Ref. 2). Sadly the present author has been retired from active paid work for some 20 years and it is considered that, apart from the concepts and any possible guidance necessary, further commercial progress must be left to others.

References.
7. Website 'www.simulatedvision.co.uk'.