The Natural Flow of Perspective:
Reformulating Perspective Projection for Computer Animation.

E.H. Blake

Centre for Mathematics and Computer Science (CWI),
Department of Interactive Systems, Kruislaan 413,
1098 SJ Amsterdam, The Netherlands.
Email: edwin@cwi.nl

Abstract

"What essential information about natural scenes would have to be simulated to make convincing animated pictures?" This question can be answered by combining theories of perception, image analysis and computer graphics. The synthesis would form the foundation of realistic computer animation. Initial steps towards the answer are taken by exploring various aspects of *perspective*. Perspective refers to the appearance of things relative to the position and motion of the simulated observer. The term covers the change of spatial detail with simulated distance and the temporal flow of appearances due to relative movement —the optic flow.

It's shown how projection on a flat plane can fail to account for our intuitive expectation of how appearances change with distance. A way of describing truly moving pictures, not a sequence of still frames, is also derived. This accords better with the way things are seen: the eye has no shutter chopping up the continuous optic flow.

These ideas come from regarding image synthesis from the point of view of the observer of the simulated world rather than from analysing the physics of the simulation. This approach seems essential for further advances in computer graphics.

1. Transcribing a Simulation of Reality.

Perspective is the method for computing realistic images. Broadly interpreted, it covers numerous techniques for mimicking a viewer's subjective experience of an environment. Simple linear or artificial perspective can, however, produce a 'realism' that runs counter to intuition. Moreover, realism in a picture is not necessarily the same as simulating physics.

This renewed look at all forms of perspective also demonstrates a more general principle, which might be called the *viewer-centred approach*. This approach grew out of ascertaining that: (1) one models aspects of the environment which can be made visible, (2) the simulation is rendered on a display screen to satisfy a viewer and (3) the result need only be 'good enough'. Sight is a utilitarian sense and irrelevant detail, were it to survive the limitations of the display device, would be ignored.

A computer animation system therefore has to allow for effects in two distinct media: (1) The simulated medium: we must know what information would have been conveyed to an observer in nature so as to mimic the natural effects. (2) The physical medium: the viewer *in fact* looks at an artificial display device. We must know how the synthetic images are viewed to allow for the artefacts. The two descriptions have to be merged to yield a theory of how to simulate natural viewing conditions on a computer display unit (a Surface-Atmosphere-Camera-Raster-Observer Viewing theory!).

The viewer-centred approach has the following manifesto:

> The basis of a sound theory for realistic computer animation lies in appreciating what is visually important in the environment and integrating this with a

> theory of how artificial images are perceived.

This first section considers the need for such a principle. The next section examines what is seen in nature. The third section describes some mathematical tools for applying the principle and reviews some results. In the fourth section remaining components of a complete theory are brought together. Finally, in the conclusion, I summarize the results achieved and place them in the context of computer animation.

1.1 Making Realistic Images with a Machine.

Computer graphics and animation are governed by a pragmatic approach to producing realistic pictures. Little attention is given to any theoretical foundation for the techniques used. *Realism* is dismissively defined to be 'like a photograph'. This pragmatism is applied throughout computer graphics. However, the resulting disparate collection of techniques is becoming unmanageable. A foundation has to be sought, but it must not be too restrictive a framework.

The parallel quest for realism and improved theoretical standards has occasionally caused controversy. There are adherents of physical rectitude, like Greenberg [1]:

> Just as VLSI advances were made by material scientists and biological advances were made by cracking the genetic code and molecular modelling, so must computer graphics improvements be based on the laws of physics.

and there are the defenders of 'faking it,' like Reeves [2,3]:

> You do what you can, and then fake it. That's nothing to be ashamed of. I enjoy fooling you.

The more subtle version of the controversy is whether a clean break has to be made between the physical and perceptual stages of the image synthesis process [4]. I argue here that the ultimate act of perception should inform *all* stages of image synthesis. The problem with faking is of course its *ad hoc* nature. We need a broader basis for computer graphics than the laws of physics: a science which can incorporate 'faking' and provide an explanation of how it works. It should also be able to encompass principles of traditional animation, such as 'exaggeration' [5].

The world described by physics differs from the world of sensory experience. The purpose of computer graphics is not to *simulate* the former for its own sake but to *stimulate* the latter. The world of physics is the world of objective facts about what 'really' happens in the realms of energy and matter. The world of sensory experience is world of interesting or boring pictures, convincing or unconvincing images. It depends as much (more?) on the perceiver as on the objects actually perceived [6].

The contrast between physical representation and experiential modelling may be illustrated by a simple experiment (first described by Otto von Guericke in 1672 [7]): If on a white wall both a red light and a white light is shone and we put our hand in front of the white light we get a shadow surrounded by a pink background. The shadowed area is of course reddish. But if we obscure the red light instead we do not get a whitish shadow. It is aquamarine!

No amount of ray-tracing colour components could ever give that colour. This does not mean that there is no real world, nor that studying tri-stimulus colour theory is worthless. But the only test for pictures is the conviction they carry. Remarks concerning physical faithlessness are irrelevant. A corollary is that a new technique should not be accepted only because it models physical reality more accurately.

The problem is also one of levels of description. The 'hard' sciences (e.g., radiation transfer, neurophysiology) are at too low a level. They will form components of an integrated theory, which also has to take into account the practical investigation into human visual experience conducted by artists over the centuries.

1.2 Perspective: Natural, Artificial and Otherwise.

The relevance of artistic experience to computer animation is well demonstrated by that old topic—the *sine qua non* of realism—perspective projection. The knowledge that the eye perceives only the solid angle subtended by an object, and that more distant objects subtend smaller angles, is ancient. It can be found in Euclid's *Optics* of the third century B.C. This is commonly called *natural perspective*. Natural perspective is closely related to the projection of images onto a sphere surrounding the viewer. The dimensions are usually normalized to yield a unit sphere.

Linear or *artificial perspective* is the mathematically accurate perspective projection of three-dimensional scenes onto a two-dimensional plane. The first exponent was probably Filippo Brunelleschi in the early fifteenth century. On the plane, artificial perspective will reproduce the natural perspective solid angle correctly for *only one* particular viewing point.

For Renaissance painters the term perspective covered a broad range of effects: perspective projection on a flat plane (artificial perspective or linear perspective), natural perspective (Euclid's viewer-centred pyramids; cf. Gibson's optic array discussed later), motion transformations and atmospheric effects.

1.2.1 The Realism of Linear Perspective.

Perspective pictures do *not* absolutely reproduce physical reality since pictures are normally viewed without any regard for the *single* point at which the planar projection re-creates the visual solid angles of the scene. We have all learned to unscramble off-centre views to make them look realistic [8]. Artists also tend to avoid wide angles where artificial perspective diverges very much from natural perspective (unless anamorphic images are sought).

Leonardo da Vinci [9] gave a good example of this distortion, or divergence, of artificial perspective from natural perspective: If a row of circular columns parallel to the picture plane are projected, then those further away from the observer will be *bigger* on the picture, as indicated in Figure 1. This figure also clearly shows the difference between artificial and natural perspective [10].

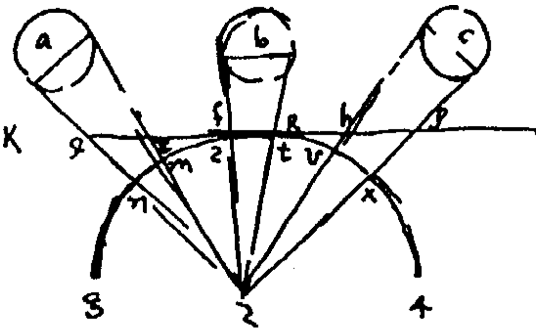

Figure 1. Leonardo da Vinci's illustration of perspective distortion.

This plan shows how more distant columns have larger images than closer columns with planar projection, while on a unit sphere the images are smaller. Circles a, b and c are the cross-sections of the columns; the observer is at point 2. The line K represents the picture plane of artificial perspective. The arc 3–4 is the unit sphere of natural perspective. Size on the picture plane depends on the distance from the plane. *Size on the sphere depends on distance from the* observer. *(Diagram from Bibliothèque de l'Institut de France, Léonard de Vinci, ms. A, fol.* 38^{r}*)*

In Figure 1 column b is closer to the observer than column c, yet segment fr, the projection of b, is shorter than segment hp, the projection of c. The angles of natural perspective (or arcs in the figure) do behave in the way we would expect. Various curvilinear perspectives have been proposed as alternatives to artificial perspective but ultimately the advice has been to avoid situations yielding gross distortions. For example, segment fr has much the same length as arc zt, so if wide angle views are avoided then artificial perspective and natural perspective nearly coincide. Synthetic perspective refers to various techniques that reduce the counter-intuitive distortions of artificial perspective, usually by combining a number of projections (vanishing points) in a single image.

One final point can be made regarding realism: what seems very realistic today becomes dreadfully artificial tomorrow, what is realistic in one culture is stilted and artificial in another.

> When you look at a bed sideways, or in front, or from any other position whatever, does it alter its identity at all, or does it continue really the same, though it appears changed? ... Does painting study to imitate the real nature of real objects, or the apparent nature of appearances? ... Painting therefore is busy about a work, which is far removed from the truth [11].

Plato would never have accepted that perspective views could be compatible with physical laws—it's fakery!

1.2.2 The Realism of Photographic Perspective.

A radical solution to the problems of artificial perspective is to return to the source of information about the environment: natural perspective. The problem of rendering natural perspective becomes that of drawing a sphere on a flat surface, a familiar problem to map makers. Any form of projection has to distort the appearance of things; it is thus a question of finding the less objectionable distortion for the purpose of the picture.

The photographic collages of David Hockney are a critique of many aspects of photography. His work, the *joiners*, "were much closer to the way we actually look at things, closer to the truth of experience" [12], however "you don't just dump one point perspective, ... it is part of a more complex perspective, which we must move on to" [13]. Here I concentrate on perspective projection, but the further critique of photography apparent in the joiners relates to the underlying aim of this paper: the introduction of a viewer-centred approach to image synthesis.

Figure 2. David Hockney. "Photographing Annie Leibowitz while she photographs me, Mojave Desert, February 1983".
Photographic Collage, 25 × 60 inch © David Hockney, 1983.

Figure 2 is a joiner by David Hockney which "mocked" the realism and paraphernalia of his photographic portrait [14]. What is interesting about many of Hockney's joiners is the curved perspective so apparent in the road and motor car on the left of Figure 2. Each component photograph is a small sample of the sphere of natural perspective. When they are joined up the artist has to choose the rule for their composition and in most joiners Hockney opted to sacrifice straight lines. Hockney's joiners show how easily we can overcome the conventions of artificial perspective—decoding a joiner is no more difficult than decoding an off-centre view of a picture on a wall. Gombrich briefly discusses composite photographs but denies that they show "the conventionality of the single perspective view of a snapshot" [15].

Hockney's photographs demonstrate an alternative perspective which comes very close to that of Stark [16]. This perspective was inspired by the 'retinal image' and is designed to reproduce the visual solid angles on the picture plane [17]. The construction involves placing a curved surface between the plane K and arc 3–4 of Figure 1. All visual rays intersect this surface and a perpendicular is dropped from each intersection onto the plane K. The curvature of the constructed surface is such that the resulting image on K is exactly proportional to the angle of the natural perspective on the arc 3-4. In the same way the small photographs of Hockney are each individually proportional to that angle of natural perspective (as shown in the previous section). The "scientifically exact" image for wide viewing angles (e.g. 120°) drawn with Stark's method shows the familiar hyperbolic curves also seen in Hockney's work [18]. For slightly smaller angles approximations are possible which produce straight lines with multiple vanishing points.

Hockney says about the viewer-centeredness of his joiners: "Look at those Grand Canyons I did, I though that the horizon was horizontal but, look, it's becoming a curve. The curve is about you, not the horizon. Its related back to your body" [19]. In Hockney's later photographic work he chose to reproduce the straight lines again. Doing this involves increasing the samples of the natural perspective so as to expand the inward curving edges. There is no longer the direct correspondence between natural perspective and picture area.

This whole discussion of realism has attempted to show how unsure our footing is in even that exemplar of realistic picture production: perspective projection. Hockney shows that defining realism to be 'like a photograph' does not even settle the question of perspective projection. A working definition of realism could now be: what is realistic is whatever the viewer will accept as such—whatever convinces the viewer enough to suspend disbelief [20].

2. Describing Natural Scenes

Natural scenes are the obvious proving ground for realism. The psychologist Gibson's research into ecological optics can be usefully applied in computer graphics [21]. He gives a very detailed description of the complex stimuli available in the environment, while avoiding any theory of the processing involved in perception.

Gibson insists on the distinction between the world of physics and the environment as perceived by animals. He maintains that the observers and their environment are complementary (cf. §1). A major feature of Gibson's work is his insistence that visual perception operates properly only if the observer is free to move about, use both eyes and observe rich, changing scenes. Clearly a video display unit is limited in this respect, but computer-generated environments, where the viewer is immersed in a scene and where there is freedom to move about, could recreate such a situation.

2.1 Natural Surfaces: Appearance and Simulation.

A most important feature of the natural environment (especially for computer graphics) is that the components of nature exist at many levels of detail. These levels of detail are simultaneously perceived as *nested* within one another, for example, grains of sand on beaches in bays along the coast and fine fuzz on the leaves on trees in forests. At every scale there are forms within forms. Objects consist of components, and these components are composed of smaller components, and so on.

There is no fixed scale with which to measure things, rather scale adapts to the situation: "... no atomic units of the world considered as an environment" [22]. Equally, there is no absolute flow of time. Instead of time there is change and sequences of events.

The environment is characterized by persistence of solid substances and their layout. There are semi-solids which change shape and liquids which are contained by solids. The air is not a substance in our view but a medium permitting vision and locomotion. Solids are perceived by the layout of their textured surfaces.

Gibson argues that in ecological optics it is more important to distinguish between animate and inanimate objects than between living and non-living (as biologists do). This is because observers generally treat plants as part of the background.

Mandelbrot [23] has advocated the use of fractals to model many natural phenomena. The name fractal is meant to evoke the idea of broken, irregular objects. Fractals have (statistically) self similar detail at all scales. They are very compact procedural models from which convincing images of rivers, coastlines and mountains can be generated. Pentland [24] presents some evidence to indicate that fractals capture what naive observers mean by surface roughness.

In contrast, Marr and Nishihara argue that we can recognize animals quite well without having to reproduce their surfaces [25]. The success of stick figures and pipe cleaner animals bear this out. The essential feature of such figures is a hierarchy of coordinate systems arranged along the natural axes of the parts of the figures. Stick figures serve for many natural shapes whose form was achieved by growth. Other natural shapes, perhaps because they were produced by random weathering, can be described by fractals. Both Mandelbrot ("to see is to believe" [26]) and Marr and Nishihara ("as we see ... animal shapes are portrayed quite effectively" [27]) appeal to the convincing images presented to justify their models: this is a very good basis for a representation to use in computer graphics.

2.2 The Natural Perspective of Surfaces: the Optic Array.

The natural environment contains many surfaces and a great many textures. All of them reflecting light in their varied ways. Light is not only transmitted by air, but also rebounded between all the surfaces to reach an equilibrium. This ambient illumination is densely structured with information about all the nested surfaces of the environment.

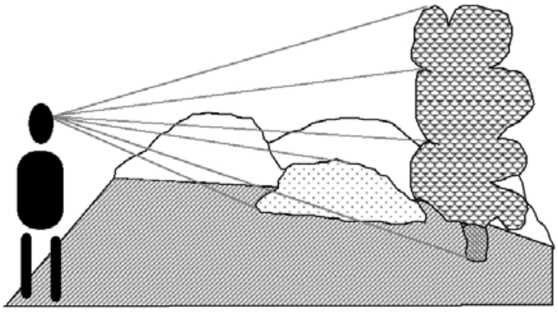

Figure 3. A small part of the ambient optic array.

The optic array is the array of nested solid angles which radiate from the viewpoint to all the surface elements of the environment. These solid angles correspond to the natural perspective of the surfaces. The drawing indicates only a very few of the myriad of solid angles which converge at the eye.

Gazing in any direction, an observer is apparently at the convergence point of a dense structure of intersecting visual pyramids. There is a pyramid for every discernible feature in the scene—it is the solid angle that light from the outline of the object subtends at the eye of the observer. Together the visual pyramids from all objects in the scene form the *optic array* (Figure 3).

Some of these interlocking pyramids belong to animals and have their own independent motion, but superimposed on these independent proper movements are the global effects of the observer's own translations and rotations. These movements of the optic array constitute the *optic flow field*—to which we return below. The notion of optic flow originated with Gibson in the 1950s and has been in wide use in computer vision research [28,29,30]. In the computer graphics literature there is a brief note by Neumann [31] and an introduction by van de Grind [32].

3. The Two Domains: Frequency *versus* Features.

To deal with the broad range of effects of Renaissance perspective we need two kinds of tools: coordinate geometry and Fourier analysis [33]. Coordinate geometry deals directly with the shapes of objects. Fourier analysis is an harmonic analysis of their spatial frequency. A full analysis of perspective in computer graphics needs to jump continually between the spatial domain and the frequency domain (see §4 for an example by Leonardo da Vinci of frequency domain effects). With computer animation, space becomes space-time. The changing, distorting, geometry is dealt with by differential geometry and temporal frequencies appear as well as the spatial frequencies.

The question of when to use which set of tools, geometry or Fourier analysis, leads to debates between 'Fourier freaks' and 'feature creatures'. Fourier analysis often tends towards a study of image formation in terms of transfer functions that state overall relations between object and image; in computational terms we may loosely call this a declarative approach. I use the Fourier or convolution method for reasoning about the process of image formation, but I generally use an explicitly simulated geometrical or procedural approach when producing pictures.

3.1 The Geometry of Moving Images.

The following analysis of optic flow is an extension of our example of perspective projection to motion perspective. I provide only the most basic results (see [34] for a detailed treatment).

Consider an ideal pin-hole camera. A set of three-dimensional observer coordinates are chosen with the origin at the pin-hole and the z-axis pointing along the direction of view. The image plane is fixed at $z = -1$ with normal vector **k** (**k** being a unit vector in the z-axis direction, perpendicular to the image plane). Let us indicate an object point with the position vector $\mathbf{R} = (X, Y, Z)$, and an image point with the position vector on the image plane of $\mathbf{r} = (u, v, -1)$ (Figure 4).

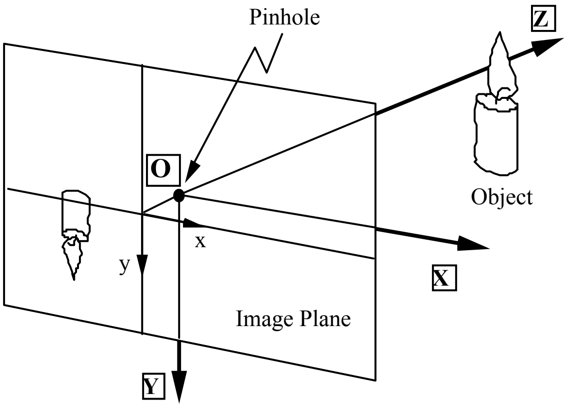

Figure 4. Perspective projection of a pinhole camera.

The pinhole is at the origin, O-Z is the optic axis. The image plane is to the left and the object viewed on right. In computer graphics the image plane is often on the same side as the object: 're-inverted'.

The familiar perspective projection equation is then:

$$\mathbf{r} = -\frac{\mathbf{R}}{Z} \qquad (1)$$

Equation 1 states that the appearance, **r**, of objects, **R**, is smaller by their distance from the eye perpendicular to the image plane, Z.

If **R** is a function of time then:

$$\frac{d\mathbf{r}}{dt} = \dot{\mathbf{r}} = \frac{\dot{\mathbf{R}}}{Z} + \frac{\mathbf{R}\dot{Z}}{Z^2} \qquad (2)$$

Equation 2 provides an expression for the changes in the image, $\dot{\mathbf{r}}$, in terms of the movement of the object $\dot{\mathbf{R}}$.

If we deal only with rigid bodies moving and rotating at a constant rate then the object velocity is $\dot{\mathbf{R}} = \mathbf{\Omega} \times \mathbf{R} + \mathbf{V}$, where $\mathbf{\Omega}$ is the angular velocity of the object rotating about the origin (eye point) and **V** is its velocity due to translation (i.e., straight line motion). The symbol '×' indicates the vector cross product, it expresses the fact that rotation involves a movement which is at right angles both to the axis of rotation and position relative to that axis.

The component of motion along the view direction is $\dot{Z} = \mathbf{k} \cdot \mathbf{\Omega} \times \mathbf{R} + \mathbf{k} \cdot \mathbf{V}$, which is the projection of $\dot{\mathbf{R}}$ on the z-axis. The symbol '•' indicates the scalar or dot product that projects one vector onto another.

Now, eliminating **R** with Equation 1, we can write the *equation of the optic flow field*:

$$\dot{\mathbf{r}} = -\frac{\mathbf{V} + (\mathbf{k} \cdot \mathbf{V})\mathbf{r}}{Z} + \mathbf{\Omega} \times \mathbf{r} + (\mathbf{k} \cdot \mathbf{\Omega} \times \mathbf{r})\, \mathbf{r} \qquad (3)$$

Here the changes in the image are written directly in terms of the translation and rotation of the object. One feature of this result is that it can be split into a (first) translational part that depends on Z, the depth, and a rotational part that is independent of depth.

If we write $\mathbf{r}_0$ for the image of the point towards which the observer is moving then:

$$\mathbf{r}_0 = \frac{-\mathbf{V}}{\mathbf{k} \cdot \mathbf{V}} \qquad (4)$$

Substituting Equation 4 into 3 and setting $\mathbf{\Omega} = 0$ (pure translation) we get:

$$\dot{\mathbf{r}} = -(\mathbf{r} - \mathbf{r}_0)\frac{\mathbf{k} \cdot \mathbf{V}}{Z} \qquad (5)$$

If **V** is fixed then image points (texture elements) move away from $\mathbf{r}_0$, which is called the *focus of expansion* for that reason. The speed of the texture elements is proportional to their distance on the image from the focus of expansion and inversely proportional to their depth (Figure 5).

Features of the Flow.

The optic flow field $\dot{\mathbf{r}}$ (Equation 3, Figure 5) is a smooth function over **r** and *t* only if **V**, $\mathbf{\Omega}$ and Z are smoothly varying. If we assume we are dealing with freely moving rigid bodies then the flow field will be segmented into smoothly varying patches. The boundaries of these patches occur where there are depth discontinuities or where one object occludes another or moves faster than another.

Rotations around an axis through the vantage point (or eye movements) involve no information about spatial layout or depth. Translations of the vantage point always yield the same pattern of field lines. The shape, but not the value, of the flow field is independent of spatial layout. For example, when moving forward the general trend will be for objects to move backward. Optic flow consists of piece-wise smooth regions within which the flow is smoothly varying, separated by discontinuities. The local optic flow field can be analysed into components. There is the average flow velocity at a point in the field and the motion parallax field is the structure of the local variation of velocity in the immediate neighbourhood of the point.

This analysis can be taken much further by analysing the local distorting effects of the optic flow. These distortions can be organized into hierarchies of optic flow effects. This in turn leads to hierarchies of frame-to-frame coherence effects.

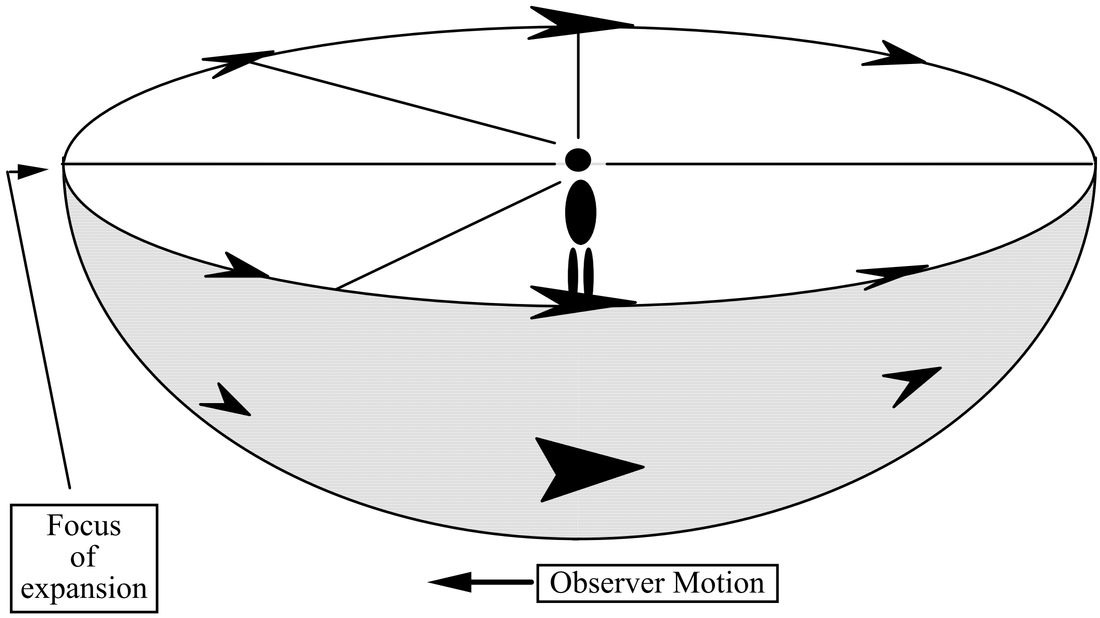

Figure 5. Motion perspective or optic flow on the unit sphere.

The flow of the optic array resulting from observer motion. The arrows represent the angular velocities of texture elements in the scene. The flow lines radiate from the focus of expansion. The effects of natural perspective are preserved by this projection on the unit sphere surrounding the observer.

3.2 Image × Object Relationships in Frequency Space.

This section outlines some concepts from Fourier analysis [35,36]. The treatment is very brief and is only intended to acquaint the reader with some very useful insights of the theory.

Changing images are functions of two space variables x , y and of time t : $f(x, y, t)$, where f is the varying illuminance at the image plane. The image can be analysed as the sum, or integral, of a large number of space and time frequencies (its spectrum) by taking its Fourier transform, $F(\xi, \eta, \nu)$, where ν represents temporal frequency and ξ and η represent the spatial frequencies. In optics F invariably has an upper limit on the frequencies it contains, that is, f is a band-limited function. Motion induces a shear in the temporal frequency dimension. The spectrum of a stationary image lies in the ξ, η plane, when the image moves the spectrum is sheared into an oblique plane through the origin. (Details are beyond the scope of this article, see [37,38]).

3.2.1 Convolutions and Transfer Functions.

The components of an optical system can be characterized by their point spread function. This function tells how every point of incoming light energy is dissipated or spread out. We can assume that the spread function is (approximately) the same all over. The resulting image is then the sum of the spread functions of all the points in the source scaled by their brightness (Figure 6).

This weighted sum (actually an integral since it sums over infinitesimal points) is what is called a *convolution* of the source with the point spread function. So the output of an optical system is (approximately) related to the source by a convolution. Frequently, a convolution will have a blurring effect on sharp changes in the image.

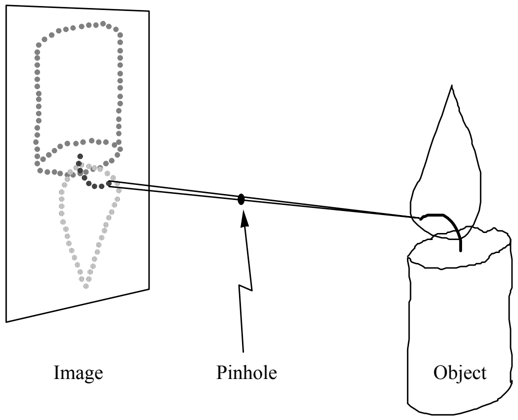

Figure 6. The image of a pinhole camera as a convolution.

Every point in the source is spread out by the point spread function that characterizes the pinhole camera. The sum of all these spread points (and not all have been drawn) makes the resulting image. It is then said to be the convolution of the source with the point spread function [based on 39].

Let us write f for the image that a ‘perfect’ optical system would give, and d for the actual observed image, which is f altered by the optics. We write g for the point spread function, a characterization of the optical system. If we use the symbol * to indicate the convolution operation., then the relation is:

$$d = f * g \qquad (6)$$

$$= \int_{-\infty}^{\infty} f(\mathbf{r}')\, g(\mathbf{r} - \mathbf{r}')\, d\mathbf{r}'$$

where the integral is over all dimensions, here indicated by **r**'.

If we regard information as being transmitted in stages, as a series of images, from the object, f_0, to a final image, f_n, then for each intermediate stage $f_1, f_2, \ldots f_{n-1}$ there is a corresponding convolution, and we can write:

$$f_0 * g_1 * g_2 * \ldots * g_n \qquad = f_n$$

Each successive convolution is a frequency filter. The $g_1 \ldots g_{n-1}$ are the intermediate point spread functions.

A very important feature of convolutions is that the Fourier transforms of the terms, written below in the corresponding capitals, are related by multiplications.

$$F_0 \; G_1 \; G_2 \; \ldots G_{n-1} \qquad = F_n$$

So instead of convolutions in the spatial domain we have simple multiplications in the frequency domain.

The Fourier transforms, $G_1 \ldots G_{n-1}$, are called the transfer functions of the filters. This is then one reason for the usefulness of Fourier analysis: The response of an imaging system to incident radiation can be described in terms of the optical transfer function (OTF). The OTF is simply the Fourier transform, G, of the impulse response or point spread function, g, of that system. The magnitude or modulus of G is called the modulation transfer function (MTF):

$$M = |G| \qquad (9)$$

3.2.2 Sampling.

In order to represent an image on a video screen it has to be sampled. Sampling at intervals τ replicates the spectrum F(*k*) in the frequency domain at intervals τ^{-1}. The sampling process is completed by extracting a single copy of the replicated spectrum.

If the sampling is too coarse then some of the duplicated higher frequencies will overlap with the lower frequencies. The high frequencies will be indistinguishable from the lower frequencies: they appear under another alias. Familiar examples of aliasing are the jaggies which plague straight edges on a raster display. Preventing this from happening, or minimizing its effects, is known as anti-aliasing. An effective solution is to ensure that the excessively high frequencies of the source are first removed before sampling.

(7)

4. The ‘Atmosphere’-VDU-Eye transfer function.

To portray changing scenes we would like to adapt the information displayed to meet the minimum requirements of convincing the human eye. Computer-generated animation already depends on features of human visual perception to make the image synthesis problem tractable. The way the eye integrates changes in the visual field over time makes a succession of static samples of the optic array indistinguishable from the continuous original.

From a surface being viewed the light first travels through the atmosphere. Atmospheric effects cause high spatial frequencies to be lost. This kind of high spatial frequency loss is described in the following observation of Leonardo da Vinci [40]:

> In every figure placed at a great distance you lose first the knowledge of its most minute parts, and preserve to the last that of the larger parts, losing, however, the perception of all their extremities; and they become oval or spherical in shape, and their boundaries are indistinct.

Scattered light also causes loss of contrast (e.g., surfaces seem lighter, more blue).

The first step of perception is light entering the eye. Light rays are mapped onto the retina with a central projection, as in a camera. But the eye is nothing like a camera when it comes to capturing motion. A camera neutralises movement and produces a series of static frames. The eye has no shutter or scanning beam, it is constructed for continuous recording of optical change over time. Essentially the eye sees only optic flows. It deals with time-continuous perspective transformations [41]. Computer animation has to induce a perception of optic flow, not (necessarily) produce a sequence of complete static images.

4.1 Human Visual Perception.

Information is transmitted to the visual cortex. A fundamental insight of sensory physiology is that there are many parallel pathways within a sensory system. Each is specialised to carry information about part of the total stimulus space [42]. A classic example are the different structures which respond to different wavelength ranges of light. Colour video displays therefore need stimulate only these by using three different coloured phosphors.

Relationships in optical systems and higher level perception can be characterised by the modulation transfer function (MTF—Equation 9). The eye *and* the rest of the visual system act as a band-pass filter.

According to Gabor's theory of communication, if we represent a signal by a sequence of samples we have to compromise between uncertainty in time and in frequency [43]. The same uncertainty relation limits the joint resolution in all dimensions. In two-dimensions there is a trade-off between resolution for spatial frequency (e.g., detail) and orientation on the one hand and spatial resolution (e.g., position) on the other. The inescapable trade-offs are probably optimized for natural scenes by our visual system [44].

The major conclusion is that the higher visual functions can be analysed, at least to begin with, by Fourier theory. Thus it has a wider application than merely the optical stages of perception. The synthetic 'camera' could allow for the overall perceptual sampling processes so as to limit the sampling which it performs on a scene.

4.2 Motion Vision and Artificial Displays.

Real movement refers to the experience of motion when an object is continuously displaced [45]. Apparent motion occurs whenever the displacement of an object is discontinuous and motion is still perceived. A motion picture conveys apparent motion. Movement perception is not a failure to resolve space and time but rather an active search for an integration of the two.

It seems that apparent movement perception is mediated by two processes: (1) A *short-range process* operates with video and cinema displays and also handles real movement. This process precedes shape recognition. (2) Movement can also be seen when the short-range process does not operate but where shape recognition can take place, the *long-range* apparent motion effect [46].

There are many kinds of artificial displays, with one common feature: limited bandwidth that affects both spatial and temporal detail. Aliasing (§3.2.2) effects occur if the source contains higher frequencies than the display can handle. If the update rate of a display is too low then we cannot induce apparent motion effects (no 'real-time' display)

The limits to human spatial- and temporal-frequency sensitivity have been called a *window of visibility* [47]. Watson *et al.* provide a useful synthesis of results from vision research specifically for time-sampled displays and computer imagery. The limits to human spatial- and temporal-frequency sensitivity are relatively independent of each other. We can therefore place a limiting box over the frequencies (ξ, η, ν) of the image $f(x, y, t)$. Frequencies outside this range will be filtered out.

This 'window' predicts the critical sample rates required in space and time to render motion accurately on artificial displays. For computer-generated imagery it is suggested that various spatial frequency bands in such synthesized images be treated separately and that we display only those whose velocity does not produce aliasing. Since motion can be detected independently of the recognition of shape when images are presented rapidly, we may hope to sacrifice spatial fidelity in order to get movement at the correct speed.

4.3 Information Channels: from computer model to viewer's mind.

The theory being developed here is desired for the animation of scenes taken from nature. The human visual system is spatio-temporal and the theory should also combine spatial and temporal aspects. Information is conveyed to the viewer via several processing steps and each of the steps has an appropriate theoretical description. These steps involve changes in knowledge representation levels. The animation process starts with high-level knowledge within the machine: a dynamic three-dimensional simulation of a natural environment and its interrelations. The knowledge representation level proceeds downwards in steps. It ends up at a very low level representation: changing two-dimensional light patterns on a display. Human visual perception then extracts the information from the display to re-create a high-level representation. From this viewers are expected to reconstruct their own approximation to the author's intentions. Computer animation thus means solving the inverse of the problem addressed by computer vision [48].

5. Conclusion.

We have been exploring the benefits of regarding animation problems from the standpoint of the viewer. I have argued that realistic pictures do not necessarily depend on accurate physics. First, the natural world as we perceive it cannot usefully be described by the laws of physics alone. Second, the various imaging systems that lie between such a natural world and its imitation on a computer display have their own limitations and possibilities which have little to do with physics.

Developing a full account of the *viewer-centred approach* should be the aim of a complete scientific programme. In this paper I have developed an example based on the complete Renaissance notion of perspective. Simply analysing all aspects of perspective already requires mastery of a number of disciplines. This example itself leads to a number of fruitful results, briefly summarized in §5.1.

Our perception of a display is an active process that strives to create order and sense: real motion and full colour are seen where there are only static images and three-coloured phosphors. These examples are familiar but there are other effects that might also be exploited, for example, the limits in sensitivity to spatial and temporal frequencies, or the perception of continuously flowing images where parts of the image deform, rather than sequences of complete static images.

The need for a theoretical basis for synthesizing realistic pictures remains even if physics by itself proves inadequate. Such a theory can draw upon the topics discussed here as its starting point.

5.1 Formalizing "Faking".

We can develop a *spatial detail metric* [49,50]. The spatial metric is primarily dependent on distance. It measures perspective effects in the broad sense (the way the appearances of things change with distance). Detail is lost because of the relative diminution of areas with distance and because of atmospheric perspective and numerous other effects.

In an analogous way to the spatial metric a *dynamic* or *temporal detail metric* can be defined. The dynamic metric can measure the speed with which the projected images move. This speed can govern update rates. As pointed out previously, the observer in nature is confronted by a changing optic array—the optic flow field. The temporal metric can be extended to measure the various levels of optic flow effects. The levels of frame similarity indicated by the temporal metric corresponds to a hierarchy of different orders of frame-to-frame coherence.

Hofmann [51] examined the use of scene-shifting in traditional film making and presented suggestions for its use in computer animation. Scene-shifting is a mechanical way of approximating the three-dimensional effects which result from observer motion by two-dimensional approximations. His analysis is rather intricate. Optic flow analysis seems to be a more powerful tool which enables one to handle more complex effects.

When objects move their spatial frequencies get altered (§3.2). However, the passband of the *window of visibility*, (§4.2) remains unchanged. There is a trade-off between spatial and temporal detail. It is important to note that the use of these trade-offs depends on spatial anti-aliasing of the lower spatial resolution images. This can be computationally expensive.

These spatial and temporal priority metrics are refinements of what is so inadequately called "faking it".

Acknowledgements.

My thanks for many discussions, particularly on optic flow, to Dr. Hilary Buxton of Queen Mary College, University of London. I am very grateful to Mr. David Hockney for allowing me to reproduce "Photographing Annie Leibowitz while she photographs me, Mojave Desert, February 1983".

References.

1. D.P. Greenberg, "Coons award lecture", *Communications of the ACM* **31**, 123–129,151 (1988). (quoted passage is on p. 125).
2 W.T. Reeves, *Statement for the panel:* "The physical simulation and visual representation of natural phenomena", *SIGGRAPH'87: Computer Graphics* **21**, No. 4, 335–336 (1987).
3 W.T. Reeves, *Quoted in:* Frenkel, K.A. "Physically based modelling *vs* 'Faking it'" *Communications of the ACM* **31**, 116–117 (1988).
4 G.W. Meyer, H.E. Rushmeier, M.F. Greenberg and K.E. Torrance, "An experimental evaluation of computer graphics imagery", *ACM Transactions on Computer Graphics*, **5**, 30–50 (1986).
5 J. Lasseter, "Principles of traditional animation applied to 3D computer animation", *SIGGRAPH'87: Computer Graphics* **21**, No. 4, 35–44 (1987).
6 This view of computer graphics is compatible with the phenomenology expounded, for the benefit of computer scientists, in T. Winograd and F. Flores, "Understanding Computers and Cognition: A New Foundation for Design" (Norwood, New Jersey: Ablex, 1986).
7 H. R. Maturana and F.J. Varela, *The Tree of Knowledge: The Biological Roots of Human Understanding* (Boston: Shambala, 1987) pp. 17,20–21.
8 R.N. Haber, "Stimulus information and processing mechanisms in visual space perception." **in** Beck, J., Hope, B. &

Rozenfeld, A. (**eds.**) *Human and Machine Vision* (New York: Academic Press, 1983) pp. 157–235. (See especially p. 183 and p. 209 ff. on stationpoint compensation.)

9 J.S. Ackerman, "Leonardo's eye", *Journal of the Warburg and Courtauld Institutes* **41** pp. 108–146 & plates 16–17, (1978).

10 Figure 1 has rather a history of causing confusion and controversy. This is mostly irrelevant to my discussion, although it might be helpful to point out that Sir Ernst Gombrich seems to miss its implications (E.H. Gombrich, *Art and Illusion*. (Oxford: Phaidon Press, 1959) 215–217). See also Ackerman, [9].

11 Plato, c. 428–348 B.C.—quoted in: L. Wright, *Perspective in perspective.* (London: Routledge & Kegan Paul, 1983), p. 35. See also: Plato, *The Republic.* (Hammondsworth, England: Penguin Books, Second Edition, Revised, 1987) Translated by Desmond Lee. The customary reference for the quoted passage is 598a–d.

12 P. Joyce, *Hockney on Photography*, (London: Jonathan Cape, 1988), p. 14.

13 See Joyce [12], p 152.

14 "It was mocking it ... with the camera that happened to be in my pocket" Joyce, [12], p 64.

15 E. Gombrich, "Mirror and Map: Theories of pictorial representation", *Philosophical Transactions of the Royal Society of London*, **B 270** (1975) 119–149 and plates 12–22. The quote is from p. 145. In this paper I have skirted the controversy surrounding curvilinear perspective. This survey by Gombrich (see p. 143 ff) is a good starting point for exploring it. The reader may also refer to: P.A. Heelan, *Space-perception and the philosophy of science*, (Berkeley: Univ. of California Press, 1983).

16 Fritz Stark, *Das Netzhautbild: Verfahren zur Herstellung des wahren Sehbildes nach dem Grundprinzip des menslichen Sehens angewandt auf die zeichnerische Konstruktion der Perspective,* (Neuss am Rhein: published by the author, 1928). The title can be translated as: "The Retinal Image: A method to restore the true perceptual image according to the basic principle of human vision applied to the construction of draughtsman's perspective."

17 "Any attempt to use the retinal image as a picture to be perceived is doomed to lead all theorists and researchers into an acrimonious and futile debate and into silly experiments as well", Haber, [8] p. 159. I therefore refer instead directly to the stimulus at the eye, viz. optic array or visual solid angles.

18 "Das wissenschaftlich-exakte Netzhautbild", Stark, [16], p. 67 ff.

19 See Joyce, [12], p. 101.

20 Coleridge discusses how the power of poetry can arise either from a faithful adherence to nature or using imagination: "so as to transfer from our inward nature a human interest and a semblance of truth sufficient to procure for these shadows of imagination that willing suspension of disbelief for the moment, which constitutes poetic faith." see: S.T. Coleridge, *Biographia Literaria*, G. Watson (ed.), (London: Dent, 1975) pp. 168–169.

21 J.J. Gibson, *The Ecological Approach to Visual Perception*, (Boston: Houghton Mifflin, 1979).

22 See Gibson [21], p. 9

23 B.B. Mandelbrot, *The Fractal Geometry of Nature*, (New York: Freeman, 1982).

24 A.P. Pentland, "Fractal-based description of natural scenes", *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 661–674 (1984).

25 D. Marr and H.K. Nishihara, "Representation and recognition of the spatial organization of three-dimensional shapes", *Proc. R. Soc. Lond.* B **200** 269–294 (1978).

26 See Mandelbrot [23], p. 256

27 See Marr and Nishihara [25], p. 271

28 D.N. Lee, "The optic flow field: the foundation of vision", *Phil.Trans. R. Soc. Lond.* B **290** 169–179 (1980).

29 B.F. Buxton and H. Buxton, "Monocular depth perception from optic flow by space time signal processing", *Proc. R. Soc. Lond.* B **21**8 27–47 (1983).

30 J.J. Koenderink, "Optic flow", *Vision Research* **26**, No. 1, 161–180 (1986).

31 B. Neumann, "Optical Flow", *Computer Graphics* **18**, No. 1, 17–19 (1984).

32 W.A. van de Grind, "Vision and the graphical simulation of spatial structure", in *Proc. 1986 Workshop on Interactive 3D Graphics*, Chapel Hill, NC, Oct 1986, (New York: ACM, 1986) pp. 197–235.

33 Named after Joseph Fourier (1768–1830). His theory is often applied in the analysis and synthesis of sound.

34 E.H. Blake, *Complexity in Natural Scenes: A Viewer Centred Metric for Computing Adaptive Detail.* Ph.D. diss. Queen Mary College, University of London, 1989.

35 R.N. Bracewell, *The Fourier Transform and Its Applications.* 2nd ed (New York : McGraw-Hill, 1978).

36 D.E. Pearson, *Transmission and Display of Pictorial Information* (London: Pentech Press, 1975).

37 E.H. Adelson, and J.R. Bergen, "Spatiotemporal energy models for the perception of motion", *J. Opt. Soc. Am.* A **2** 284–299 (1985).

38 A.B. Watson and A.J. Ahumada, "Model of human visual-motion sensing", *J. Opt. Soc. Am.* A **2** 322–341 (1985).

39 P.M. Duffieux, *The Fourier Transform And Its Applications to Optics*, 2nd ed. (New York: Wiley, 1983) p. 71.. Originally published as: *L'inté grale de Fourier et ses applications à l'optique*, (Paris: Mason, 1970).

40 E. MacCurdy, *The Notebooks of Leonardo da Vinci*, (London: The Reprint Society, 1954, reprint of 1938 edition) Vol. **II** p. 351

41 G. Johansson, "Visual event perception", in *Handbook of Sensory Physiology.* Vol.**VIII** *Perception,* R. Held, H.W. Leibowitz and H. Tueber, eds. (Berlin: Springer, 1978) pp. 655–673.

42 O. Bradick, F.W. Campbell and J. Atkinson, "Channels in vision: basic aspects", in [41] pp. 3–38.

43 D.M. MacKay, "Strife over visual cortical function", *Nature* **289** 117–118 (1981).

44 D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells", J.Opt.Soc.Am. A **4**, No. 12, 2379–2394 (1987) .

45 S.M. Anstis, "Apparent movement", in [41] pp. 655–673.

46 O. Bradick, "A short-range process in apparent motion", *Vision Research* **14** 519–528 (1974).

47 A.B. Watson, A.J. Ahumada, Jr. and J.E. Farrell, "Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays", *J. Opt. Soc. Am.* A **3** 300–307 (1986).

48 T. Pun & E.H. Blake, "Relationships between Image Synthesis and Analysis: Towards Unification?" *Computer Graphics Forum* **9** 149–163 (1990).

49 E.H. Blake, "A metric for computing adaptive detail in animated scenes using object-oriented programming" in *Proceedings of Eurographics 87,* G. Maréchal, ed. (Amsterdam, August 24–28, 1987) (Amsterdam: North-Holland, 1987) pp 295–307.

50 See Blake [34]

51 G.R. Hofmann, "The Calculus of the Non-Exact Perspective Projection: Scene-Shifting for Computer Animation." in *Proceedings of Eurographics'88,* D.A. Duce and P. Jancene, eds. *(Nice, France, 12–16 Sept. 1988)* (Amsterdam: North-Holland, 1988) pp. 429–442.

Glossary

AI: artificial intelligence, a broad research field in computer science which includes knowledge representation, computer vision and robotics (any closer definition of this field invariably leads to controversy).

bandwidth: the range of frequencies which has to be stored or transmitted.

curvilinear: a system of coordinates, where the axes are not the normal Cartesian orthogonal, linear, set. In curvilinear perspective straight lines in an object are no longer mapped to straight lines in the image.

declarative: a method of programming computers where relations are specified which have to hold continuously between the data values.

detail: detectable features at a particular scale of measurement.

differential geometry: the local geometry of deforming surfaces.

Fourier Analysis: the analysis of objects and changes into an infinite sum of sinusoidal waves which is their spectrum. Any shape can be described in this way. Smooth shapes or motion have relatively larger low frequency components than rough shapes or sharp changes.

filter: a system which alters the spectrum of its input according to its transfer function. For example, attenuating higher frequencies would cause at a shape to be smoother.

frame-to-frame coherence: the way in which one frame of an animation is normally very much like the preceding and succeeding frame.

frequency: the rate at which a sinusoidal wave repeats.

frequency domain: The space in which all objects are described by their Fourier components rather than by their geometry.

image analysis: the general term for computational techniques for automatically interpreting and recognizing pictures.

image synthesis: the general term for computational techniques for producing (moving) pictures. In some sense the opposite of image analysis.

integral: the (often infinite) sum of infinitesimal parts, the sum exists when a finite sum tends to a fixed value as the parts over which it is taken are made smaller and smaller.

jaggies: the jagged edges of straight lines on raster displays.

modulus: the absolute value of a number, negative signs are ignored.

optimize: any method of computing some best value subject to constraints.

passband: the frequencies which a filter will allow through.

phosphors: the coloured dots which are arranged in a grid on a video display, the physical equivalent of a raster.

procedural: a method of computation where explicit procedures are specified for arriving at a result.

raster: the two-dimensional grid or array of numbers which represent images in a computer. It corresponds to the coloured dots which are displayed on a video screen.

ray tracing: a method of synthesizing images by tracing a light ray for each point on the image raster.

real-time: the response of, or a simulation by, a computer system where events inside the machine occur in the same time-frame as corresponding events in the real world. In computer animation it actually means generating pictures quickly enough to produce convincing apparent motion.

solid angle: the three-dimensional analogue of a (two-dimensional) angle.

spatial domain: the space in which objects are described by their geometry. The complement of the frequency domain.

spectrum: the collection of frequencies which correspond to a particular object.

translation: straight line movement.

vector: a geometrical object which has both magnitude and direction, represented by an array of numbers in some coordinate system.

visual cortex: the area of the brain which receives information from the eye.

VLSI: Very Large Scale Integration: the placing of very many, very small, electronic components on a small piece of silicon to make the chips so essential to modern electronics.