Chapter 2

Converging Computer and Television Image Portrayal

John Watkinson
Author

Key words: Dynamic resolution, Optic Flow, resolution, interlaced scanning, MPEG, video, bandwidth, data rate, TV, psycho-optics

Abstract: This paper is about optimizing television picture quality for a given bandwidth/data rate.

1. INTRODUCTION

With the inevitable convergence of television and computer imaging formats, the traditionally separate approaches are now a source of incompatibility which threatens to hinder progress, to no one's benefit. Regrettably there are already signs of entrenched attitudes.

The design of a new television broadcasting format is an opportunity which occurs rarely. The decisions made have a long lasting effect and must therefore be well considered. If a sub-optimal system is chosen, the cost of implementing or running the system may be higher than necessary, damaging profits. The consumer take-up may be low if the perceived quality falls below the viewer's expectations.

In this context it is the author's view that the only way to proceed is to design a format which, without incurring excessive complexity, gives the best subjective results for a given bandwidth/data rate. Anything else will simply cost more to run.

D. Gerbarg (ed.), *The Economics, Technology and Content of Digital TV*

Within this criterion of efficiency, the viewer can be offered any balance of quality and bit rate. The efficiency can be used to minimize bit rate in cost conscious applications, or to maximize quality in prestige applications.

In order to implement this strategy, only two important steps are needed. These are as follows:

1) Obtain an accurate model of the human visual system so that the sensitivity of the viewer to all relevant quality parameters is known.

2) Use that model to make objective comparisons between what is theoretically possible and any proposals. Any proposal coming close to the ideal can be selected, but if none do, work remains to be done.

In this paper it will be shown that little work remains to be done. Sufficient knowledge of the human visual system exists, and all of the fundamental technical concepts exist. An efficient, convergent moving image portrayal system with complete interoperability between television broadcasts and computer graphics can be created today with no more than an intelligent combination of existing technologies.

Once a choice, based on psycho-optics and physics, has been made, theory will be able to predict the performance of the equipment perfectly and actual demonstrations will confirm the accuracy of the theory. All of the proposals in this paper come into that category. They can all be explained in theory and they can all be shown to work in practice, singly and in combination.

It is the author's opinion that to continue to propose a sub-optimal system which violates established physics is either a misinformed belief or represents a vested interest. Unfortunately both conditions have entered the consumer TV verses computer debate, serving only to delay a rational outcome. To some extent genuine misconception is understandable and is much easier to deal with, requiring no more than an education process.

Some manufacturers of traditional broadcast equipment and consumer TVs, with their analog background, understand analog very well, but lack a wide and deep understanding of digital technology. Many aspects of today's television standards were established empirically before the relevant theory was understood. The computer industry naturally knows digital techniques backwards but tends to lack knowledge of psycho-optics and psycho-acoustics. Certainly manufacturers of traditional television equipment would rather deny the world a significant improvement in television quality so that they can cling to tradition (and their traditional profits). If this paper serves to expose only one such instance, it will have served its purpose well.

2. WHAT CAN WE SEE

Brevity requires this paper to concentrate on resolution or definition. As a high definition television (HDTV) system is to be created, this is reasonable. Considerations such as gamma and colorimetry, whilst important and interesting, cannot be treated here as they are not unique unto high definition.

The resolution of the eye is primarily a spatio-temporal compromise. The eye is a spatial sampling device; the spacing of the rods and cones on the retina represents a spatial sampling frequency. The measured acuity of the eye exceeds the value calculated from the sample site spacing because a form of oversampling is used. The eye is in a continuous state of unconscious vibration such that the sampling sites exist in more than one location. Effectively the spatial sampling rate is increased by this saccadic motion, but it can only be turned into resolution by a temporal filter which is able to integrate the information from the various different positions of the retina.

This temporal filtering is responsible for "persistence of vision" which is effectively the temporal frequency response of the eye's oversampling filter. The picture below shows the spatio-temporal response of the main viewing area in the eye. Note that between 50 and 60 Hz the temporal response starts to become negligible, hence the use of such frequencies for television picture rates.

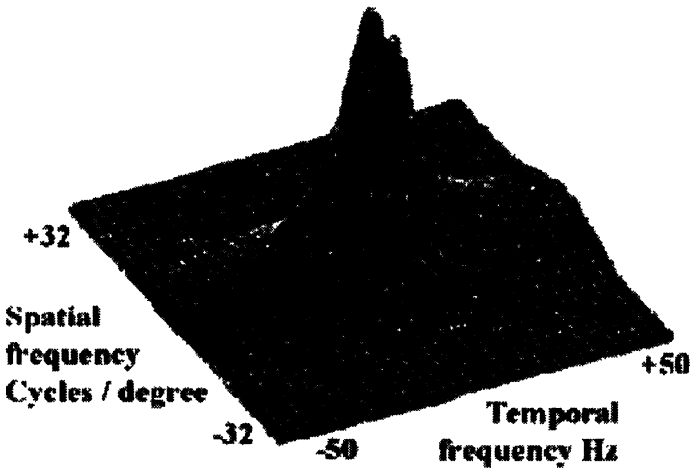

Figure 1)

The diagram below (Figure 2a) shows that when a detailed object moves past a fixed eye, the line of gaze effectively scans the object, and clearly high temporal frequencies will be created. These will be filtered by the temporal response of the eye (as shown in Figure 1), causing moving objects to blur.

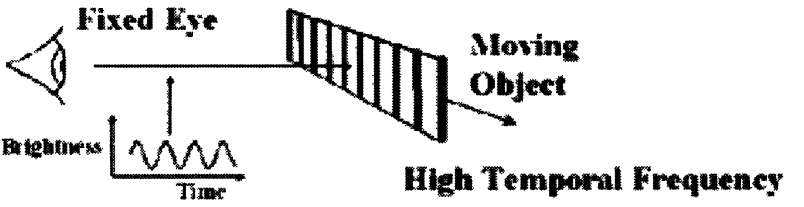

Figure 2a)

However, the situation shown in Figure 2a simply doesn't happen in real life. The human viewer has an interactive visual system which causes the eyes to track the movement of any object of interest. Figure 2b below shows that when eye tracking is considered, a moving object is rendered stationary with respect to the retina so that temporal frequencies fall to zero. In this case much the same acuity to detail is available despite motion. This is known as dynamic resolution and it's how humans judge the detail in real moving pictures. It astonishes the author that video engineers so often state that softening of moving objects is inevitable and acceptable, when it plainly isn't.

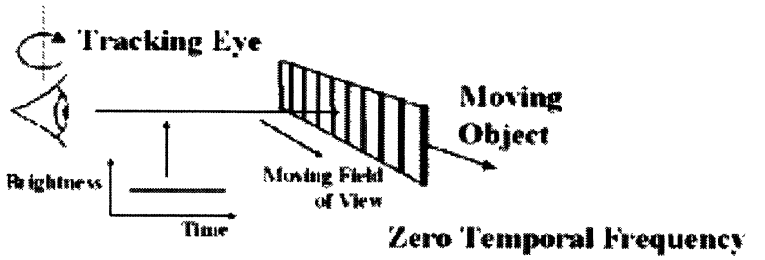

Figure 2b)

3. DYNAMIC RESOLUTION AND THE OPTIC FLOW AXIS

As the eye uses involuntary tracking at all times, the criterion for measuring the definition of moving image portrayal systems has to be dynamic resolution. Dynamic resolution is defined as the apparent resolution perceived by the viewer in an object moving within the limits of accurate eye tracking. The traditional use of static resolution in film and television has to be abandoned as not being representative of the viewing experience.

Figure 3a below shows that when the moving eye tracks an object on the screen, the viewer is watching with respect to the optic flow axis, not the time axis, and these are not parallel when there is motion. The optic flow axis is defined as an imaginary axis in the spatio-temporal volume which joins the same points on objects in successive frames. Clearly when many objects move independently there will be one optic flow axis for each.

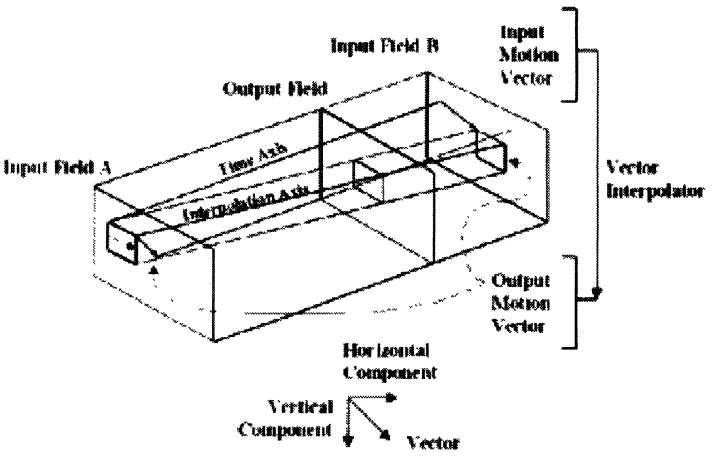

Figure 3a)

The optic flow axis is identified by motion vector steered frame rate converters to eliminate judder and also by MPEG compressors because the

greatest similarity from one picture to the next is along that axis. The success of these devices is testimony to the importance of the theory.

Figure 3b below shows that when the eye is tracking, successive pictures appear in different places with respect to the retina. In other words, if an object is moving down the screen and followed by the eye, the raster is actually moving up with respect to the retina. This has some interesting consequences. Although the object is stationary with respect to the retina and temporal frequencies are zero, the object is moving with respect to the sensor and the display and in those units, high temporal frequencies will exist. If the motion of the object on the sensor is not correctly displayed, or if these high temporal frequencies are not handled correctly, dynamic resolution will suffer.

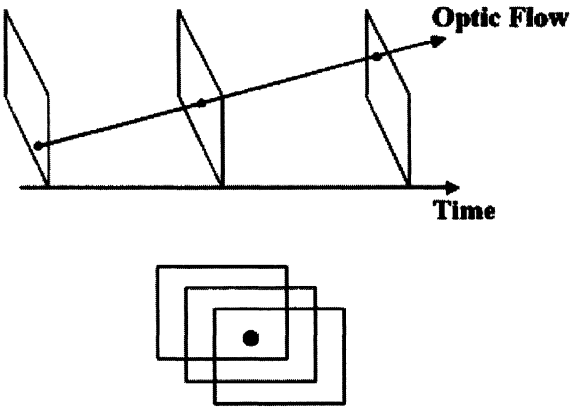

Figure 3b)

Display appears in different places with respect to a tracking eye, hence background strobing

When the eye is tracking a moving object in an image portrayal system, the background will be moving with respect to the retina. In real life this motion will be smooth, but in an image portrayal system based on periodic presentation of frames, the background will be presented to the retina in a different position in each frame. The retina separately perceives each impression of the background, leading to an effect called background strobing.

In practice the criterion for the selection of a display frame rate in an imaging system is sufficient reduction of background strobing. It is a complete myth that the display rate simply needs to exceed the critical flicker frequency. Manufacturers of graphics displays which use frame rates

well in excess of those used in film and television are doing so for a valid reason—it gives better results! Note that the display rate and the transmission rate need not be the same in an advanced system.

Perhaps non-intuitively, the dynamic resolution or perceived sharpness of a picture depends critically on the ability of the imaging system to portray motion. When the concept of dynamic resolution is used to examine competing image portrayal systems, it correctly predicts observed phenomena.

Dynamic resolution analysis confirms that both interlaced television and conventional projected cinema film are both seriously sub-optimal. In contrast, progressively scanned television systems have no such defects.

4. THE RESOLUTION OF INTERLACED SCANNING

Interlaced scanning is a crude analog bandwidth reduction technique which was developed empirically in the early days of television. Instead of transmitting entire frames, the lines of the frame are sorted into odd lines and even lines. Odd lines are transmitted in one field, even lines in the next. A pair of fields will interlace to produce a frame. Vertical detail such as an edge may only be present in one field of the pair and this results in frame rate flicker called "interlace twitter".

Figure 4, Case A, shows a dynamic resolution analysis of interlaced scanning. When there is no motion, the optic flow axis and the time axis are parallel and the apparent vertical sampling rate is the number of lines in a frame. However, when there is vertical motion, (see Figure 4, Case B), the optic flow axis turns. In the case shown, the sampling structure due to interlace results in the apparent vertical spatial sampling falling to one half of its stationary value.

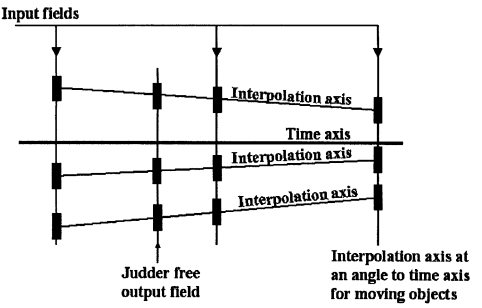

Figure 4)

Case A) With no motion, interlaced system has resolution based on number of lines in a frame.
Case B) In the presence of motion, interlaced system has vertical resolution halved to the number of lines in a field.

Consequently interlace does exactly what would be expected from a half-bandwidth filter. It halves the vertical resolution when any motion with a vertical component occurs. In a practical television system, there is no anti-aliasing filter in the vertical axis and so when the vertical sampling rate of an interlaced system is halved by motion, high spatial frequencies will alias or heterodyne causing annoying artifacts in the picture. This is easily demonstrated with a grating test card or a moving zone plate. Figure 4c below shows how a vertical spatial frequency well within the static resolution of the system aliases when motion occurs. In a progressive scan system this effect is absent and the dynamic resolution due to scanning can be the same as the static case.

c) Only alternate samples are present on optic flow axis, and aliased waveform (dashed line) results

Figure 4c)

When stationary, original spatial waveform (solid line) is sampled by line structure (dots) and waveform is correctly reproduced. In the case of motion, vertical sampling rate falls to one half. Only alternate samples are present on optic flow axis, and aliased waveform (dashed line) results.

This analysis also illustrates why interlaced television systems need to have horizontal raster lines. This is because in real life, horizontal motion is more common than vertical. It is easy to calculate the vertical image motion

velocity needed to obtain the half-bandwidth speed of interlace, because it amounts to one raster line per field. In 525/60 (NTSC) there are about 480 active lines so motion as slow as one picture height in about 8 seconds will halve the dynamic resolution. In 625/50 (PAL) there are about 600 lines, so the half-bandwidth speed falls to one picture height in 12 seconds. This is why NTSC, with fewer lines and lower bandwidth, doesn't look as soft as it should compared to PAL, because its dynamic resolution at low speeds can be higher.

The situation deteriorates rapidly if an attempt is made to use interlaced scanning in systems with a lot of lines. In 1250/50, the resolution is halved at a vertical speed of just one picture height in 24 seconds. In other words, on real moving video a 1250/50 interlaced system has the same dynamic resolution as a 625/50 progressive system. By the same argument, a 1080i system has the same performance as a 480p system. In high line number systems, interlace softening just kicks in at a lower speed and it's clear to the naked eye when this happens.

Whilst horizontal raster lines palliate the drawbacks of interlace they do nothing to help the CRT designer because this arrangement combines the highest scanning frequency with the greatest scanning deflection. With the move to 16:9 aspect ratio, the difficulty becomes even greater. With such a wide tube, it becomes logical to have vertical raster lines so that the deflection of the high frequency scan (and the current required) is nearly halved. The wide angle deflection is now only required at the frame rate. The use of interlace prevents this technique.

Interlaced signals are also harder for MPEG to compress. The confusion of temporal and spatial information makes accurate motion estimation more difficult and this reflects in a higher bit rate being required for a given quality.

Following this analysis, this author concludes that interlaced scanning has too many drawbacks to be considered in an advanced imaging system. Theoretical and subjective efficiency is low and interlace represents poor value for money. Wide-screen displays cost more than necessary, consume more power and dissipate more heat. Digital compression systems have to use a higher bit rate.

Interlace was the best that could be managed with thermionic valve technology sixty years ago, and we should respect the achievement of its developers at a time when things were so much harder. However, we must also recognize that the context in which interlace made sense no longer exists.

5. THE RESOLUTION OF FILM

Good dynamic resolution is essential for realism and will only be achieved if the motion portrayal is accurate. Accurate motion portrayal requires that the optic flow axis is reproduced without distortion.

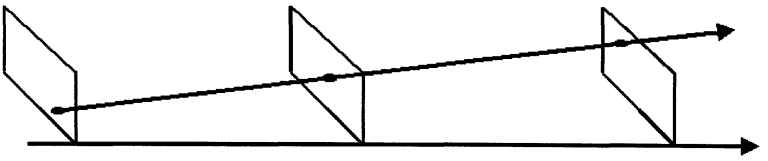

Figure 5a)

Figure 5a above shows how movie film is shot. For historical and economy reasons, the film is only exposed at 24 or 25 frames per second. The optic flow axis is correctly preserved on the film for moderate motion frequencies. However, 24 frames per second is below the critical flicker frequency of human vision and is unwatchable. The traditional palliative is to present each frame twice. The projector has a two bladed shutter which produces two flashes of light for each frame pulldown.

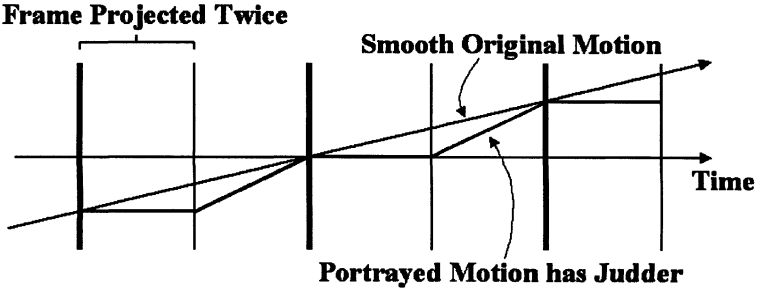

Figure 5b)

Figure 5b above shows that this corrupts the optic flow axis because there cannot be motion between the repeated frames. The eye tries to track the motion the best it can, but the optic flow axis of the film now oscillates

up and down with respect to the retina. Unlike interlace, which is worst on vertical movement, this effect is equally powerful in all directions. To a tracking eye, the two identical versions of a frame appear in different places on the retina. For slow movements, this results in an aperture effect which damages dynamic resolution. For rapid movements the result is visible as judder or multiple images.

Assuming the film has a thousand lines of static resolution, dynamic resolution will be halved by the aperture effect when a speed of one picture height in 40 seconds is reached. This is too slow to be useable, so the best dynamic resolution achieved by film hardly ever reaches half the resolution the film is capable of. The best that cinematographers can do is to mount cameras on very solid and smooth supports and move them slowly to avoid judder. Rapidly moving objects of interest must be panned. Quality films are shot like this because the filmmakers know the restrictions. Notice how good cinematographers use shallow depth of focus in order to blur the background and avoid background strobing.

The damaging effect of picture repeat in film means that although film manufacturers have dramatically improved the static resolution of film in recent years, the improvement cannot be seen by the moviegoer.

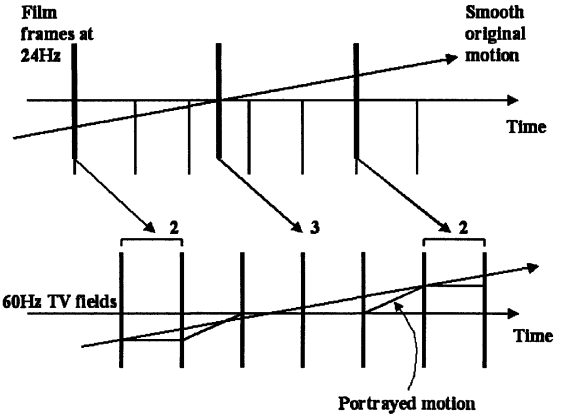

Figure 5c)

The picture repeating of film projection is carried over into telecine. To produce 50 Hz video in Europe, the 24 Hz film is run at 25 Hz and two fields are made from each frame. Production of 60 Hz video from 24 Hz film in the U.S. requires 3:2 pulldown, where one frame is made into three fields and the next is made into two fields. 3:2 pulldown has a devastating effect on the optic flow axis as shown in Figure 5c above.

Figure 6 shows that the action of the interlacing telecine is to display a frame sampled at one point in time as fields at two separate times. In the presence of motion the optic flow axis turns and these fields no longer superimpose. The shift of the fields with respect to one another causes an aperture effect which reduces the visibility of interlace aliasing. Consequently a motion artifact of film has the result of concealing an interlace artifact in video.

Bearing this in mind, using 24/25 Hz film material to test or demonstrate HDTV systems must be a very suspect practice indeed and the results are meaningless. The dynamic resolution of the TV system under test could be (and often is) quite poor yet the artifacts due to film judder could well conceal the fact.

6. FILM AND MPEG

In an MPEG environment the damaged optic flow axis from telecine causes compressors a lot of trouble. The field repeating means that motion vectors are zero between repeat fields but of doubled amplitude elsewhere. This alternating vector data means that the data available for picture differences fluctuates, causing quality loss. The current approach to MPEG compression of telecine video is to use a preprocessor which de-interlaces the fields back to progressively scanned frames. In 3:2 pulldown systems, the third field is entirely redundant and is discarded. The adoption of progressive scan at the same frame rate as the film material allows MPEG to work at its most efficient as the vector data is more stable from frame to frame.

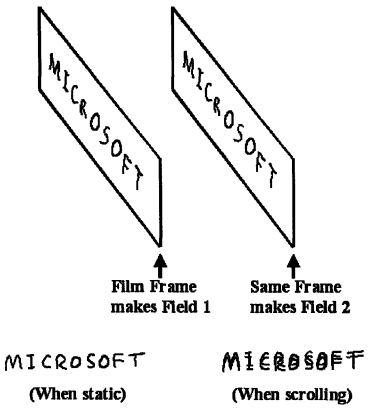

Figure 6)

Set-top boxes receiving MPEG film frames at 24/25 Hz have no trouble accurately decoding the frames, but display them by reading the output frame store at 50 Hz using interlace and at 60 Hz using interlace and 3:2 pulldown. This interlacing process recreates the damage to the optic flow axis which took place in the original telecine material.

Telecine machines are actually Standards Converters because the input and output picture rate is different. It is obvious that the only way to overcome the poor motion portrayal of the telecine machine is to use motion vector steering in the conversion process so that the optic flow axis is not distorted. A telecine which does not do this cannot be regarded as having high definition. The advantage of the motion vector steered telecine is that the output video has the same motion characteristics as video from cameras and so doesn't need to be handled differently by MPEG.

There is an enormous archive of 35 mm 24 Hz film material which will be heavily used to attract customers to new television services. The advantages of a high quality television system will be lost if primitive field repeating telecines are used.

7. OVERSAMPLING

People seem to think that high definition television needs lots of lines, but it's a myth. Cameras and displays need a lot of lines to overcome aperture effects and to render the raster invisible, but the transmission medium between doesn't. In the early days of television, the capture, transmission, and display formats had to be identical for simplicity, but that's no longer true or desirable.

A 480 line camera can't give 480 lines of resolution, but a 960 line camera with downconversion can. Effectively the camera is using oversampling. Although oversampling has totally dominated digital audio because of its obvious merits, it is harder to use it in conventional television because of interlace. Interlace puts half the picture data at another time and reduces the performance of spatial resamplers. Once interlace is dispensed with, oversampling becomes an obvious and attractive technology.

Oversampling overcomes practical limits in optical filters. In a CCD camera, the sensor elements sample the image spatially. The sensors are large for maximum light sensitivity and so a serious aperture effect is experienced. Ideally an optical anti-aliasing filter is needed between the lens and the sensor. Unfortunately it is difficult to make a filter that has a sharp cut-off and it is usually necessary to compromise between visible aliasing and picture softness.

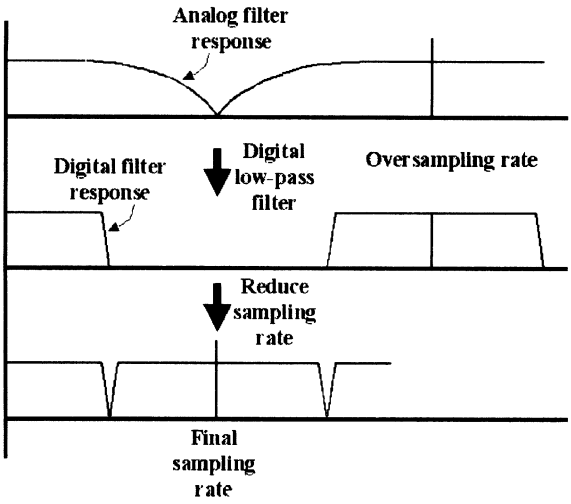

Figure 7)

Using oversampling, this compromise is unnecessary. Figure 7 above shows that in an oversampling camera, the spatial sampling rate must be increased by using a larger number of pixels in both dimensions (i.e., use a progressive HD camera). The optical anti-aliasing filter then only needs to prevent aliasing at the higher sampling rate. The output of the CCD element is spatially low-pass filtered and decimated to produce a TV signal with the target pixel count. It will contain no spatial aliasing, but will not suffer loss at the band edge.

As a CRT display is a sampled device, breaking the picture up into lines, it should ideally be followed by an optical filter. As before, this is not done because in order to eliminate the raster it would intrude into the passband. Oversampling can also be used to render the raster invisible. Once more a form of Standards Converter is required, but this now increases the number of input lines using interpolation. The aperture effect of the display filters out the raster, leaving the passband unaffected.

The adoption of progressive scan allows spatial oversampling to be easily implemented in both camera and display. The number of lines needed in the transmission channel between is then quite moderate.

Progressive scanned sensors and displays having 800 to 1000 lines connected by a 480p transmission channel are all that is required to deliver a truly high definition television service. The upconverter in the display is optional and lower cost receivers could omit it.

8. HOW TO IMPLEMENT A PROPER VIDEO SCALER (INTERPOLATOR)

Interpolation is the process of computing the values of output samples which lie between the input samples (i.e., the samples in the original video signal). It is thus a form of sampling rate conversion. One way of changing the sampling rate is to return to the analog domain using a Digital to Analog Converter and then to sample at the new rate. In practice this is not necessary because the process can be simulated in the digital domain. When returning to the analog domain a suitable low pass filter must be used which cuts off at a frequency of one half the sampling rate.

The impulse response of an ideal low-pass filter is a sinx/x curve which passes through zero at the site of all other samples except the center one. Thus the reconstructed waveform passes through the top of every sample, as shown in Figure 8a below. Between samples, the waveform is the sum of many impulses. In an interpolator a digital filter can replace the analog filter.

A digital filter can be made with a linear phase low pass impulse response in this way. As a unit input sample shifts across the filter, it is multiplied by various coefficients which produce the impulse response. Figure 8b below shows how this could be implemented. A 'windowed' sinx/x impulse response can be described using a set of coefficients stored in a look-up table (LUT).

The interpolation method usually employed involves taking the contribution of each input sample at the corresponding distance from the required output sample. All the contributions are summed to obtain the interpolated value. Figure 8c below shows the process needed to interpolate to an arbitrary position between samples. The location of the output sample is established relative to the input samples (this is known as the phase of the interpolation), and the value of the impulse response of all nearby samples at that location is added.

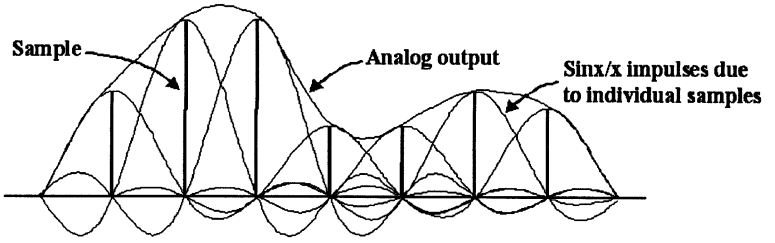

Figure 8a)

The coefficients can be found by shifting the impulse response by the interpolation phase and sampling it at new locations. The impulse will be sampled in various phases and the coefficients will be held in a look-up table. A different phase can then be obtained by selecting a different LUT page.

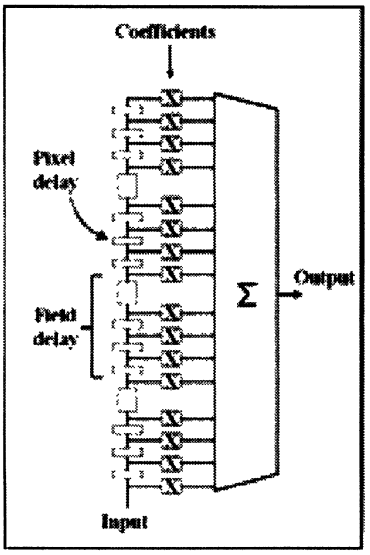

Figure 8b)

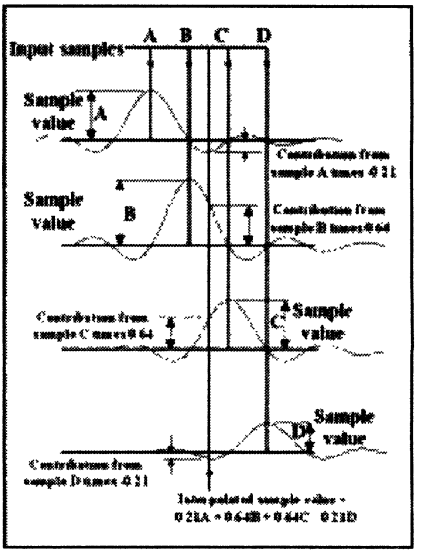

Figure 8c)

9. MOTION VECTOR STEERING

Oversampling can also be used in the time domain in order to reduce or eliminate display flicker and background strobing. A different type of Standards Converter is necessary which increases the input picture rate by interpolation. Such an oversampling converter should use motion vector steering, otherwise moving objects will not be correctly positioned in an interpolated picture and the result will be judder.

A conventional linear frame rate converter either just uses a frame store, or better, filters along the time axis by feeding the same pixel from several successive frames into an FIR filter. A temporal aperture of four frames is common although for some applications only two frames are used for economy. With such a short aperture, it is not possible to reach an acceptable compromise between roll-off and ripple and eliminating beating between the input and output frame rates is very difficult.

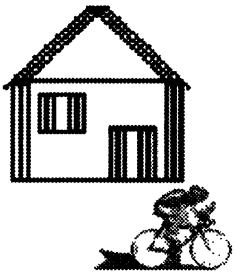

Figure 9a)

Linear filters (or no filtering at all in the case of just using a frame buffer) suffer from a major defect when used for frame rate changing. If an object is moving, it will be in different places in successive fields. Interpolating between several fields results in multiple images of the object. The position of the dominant image will not move smoothly, an effect which is perceived as judder. If, however, the camera is panning the moving object it will be in much the same place in successive fields and Figure 9a above shows that it will be the background that judders.

Motion vector steering is designed to overcome this judder by taking account of the human visual mechanism. It is a way of modifying the action of a frame rate converter so that it follows moving objects along the optic flow axis to eliminate judder in the same way that the eye does. The basic principle of motion vector steering is simple. In the case of a moving object, it appears in different places in successive source frames. Motion vector steering computes where the object will be in an intermediate target frame and then shifts the object to that position in each of the source frames prior to temporal interpolation.

An alternative way of looking at motion vector steering is to consider what happens in the spatio-temporal volume. A conventional Standards Converter interpolates only along the time axis, whereas a motion vector steered Standards Converter can swivel its interpolation axis off the time

axis onto the optic flow axis. Figure 9b below shows the input frames in which three objects are moving in different ways. It will be seen that the interpolation axis is aligned with the trajectory of each moving object in turn.

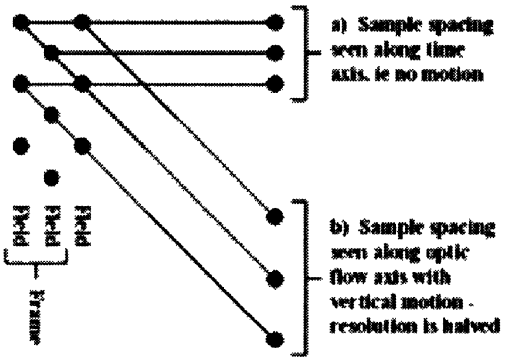

Figure 9b)

When this is done, each object is no longer moving with respect to its own interpolation axis, and so on that axis it no longer generates temporal frequencies due to motion and temporal aliasing cannot occur. Interpolation along the correct axes will then result in a sequence of output frames in which motion is properly portrayed. The process requires a Standards Converter that contains filters that are modified to allow the interpolation axis to move dynamically with each output field.

The signals that steer the interpolation axis are known as motion vectors and one of these must be available from the motion estimator for every pixel in the target frame. These are not just block based motion vectors. It is the job of the motion estimation system to provide these pixel accurate motion vectors. The overall performance of the converter is determined primarily by the accuracy of the motion vectors. An incorrect vector will result in unrelated pixels from several fields being superimposed and the result is unsatisfactory.

Motion vector steering should also be used when converting an interlaced scan video signal into progressive scan format. As the interlace process reduces the information in each field, the job of the motion estimator is somewhat harder. It is only economically feasible to use motion vector steered de-interlacers within TV studios for converting archive material for transmission in the new progressive transmission format.