Digital Film: Hiding the Raster

Thor Olson
Management Graphics Inc.
Minneapolis MN

Presented at the
133rd SMPTE Technical Conference
October 26-29, 1991
Los Angeles Convention Center, Los Angeles, CA
Preprint no 133-102

Abstract

The characteristics of image recorders having gaussian-shaped spots is investigated, particularly with respect to producing images which are free from raster artifacts. Human visual perception requires that intensity variations be less than 0.5%. A scanline spacing of 1.2 times the gaussian radius achieves this. Tighter spacing than this is required in order to mask the effects of digital roundoff noise and other noise sources in the deflection system. Sharpness, as measured by limiting resolution, is related to the spot radius, and no additional amount of pixel density will improve on it. Examples are given which illustrate this. Finally, the effects of delaying the film exposure are described. When the image data cannot be supplied at a rate to match the exposure, dark “data-late” streaks appear in the picture.

Introduction

It has become feasible to produce motion pictures digitally, merging special effects and inserting entirely synthetic scenes. Having worked hard to create realistic-looking imagery, it is important to not give it all away when you print each painstakingly assembled frame to film. There are a hundred and one ways that the output of a film recorder can give away its digital nature. This paper will discuss some of the more blatant ones, in particular those related to raster lines. By understanding them we can design the imaging system so that these digital artifacts are completely disguised; the movie audience will never know of the film's computer origins.

Underlying all digital images today is the raster. Scanline by scanline, a picture is assembled and a piece of film exposed. If the pixels were rectangular and fit exactly together, there would be little hint of the row and column organization of the image. Horizontal and vertical edges would be razor sharp but the other edges would show aliasing jaggies. Most imaging devices however, have fuzzy round pixels that don't pack perfectly. If we tightly place the pixels we obtain a nice uniform field free from visible raster lines but which suffers from lack of edge sharpness. If the pixels are not properly placed, the whole picture becomes contaminated by the texture of individually discriminable scanlines.

So where is the right place to be in order to compute a minimum number of pixels but not detect the raster? To answer this we need to closely examine an individual pixel.

Spot profile

Different film recording technologies will have different pixel characteristics, but in general, a small spot of light is used to expose pixels onto film. A number of technologies (for example CRTs and lasers) result in the spot being "gaussian". This means that the light intensity profile taken across the spot matches the mathematical formula for a gaussian distribution:

$$I(r) = e^{-\left(r/R_g\right)^2}$$

The light intensity falls off exponentially by the square of the radial distance r. The amplitude at the center is unity; at the “gaussian radius”, R_g, it is at 1/e. A plot of this intensity profile is shown in figure 1. Figure 2 is a simulated enlargement of a gaussian spot. The dark circles mark the the 50% and the 1/e intensity levels. The visual extent of the spot is much larger than either of these. In fact, the apparent diameter of the spot is roughly twice the gaussian diameter.

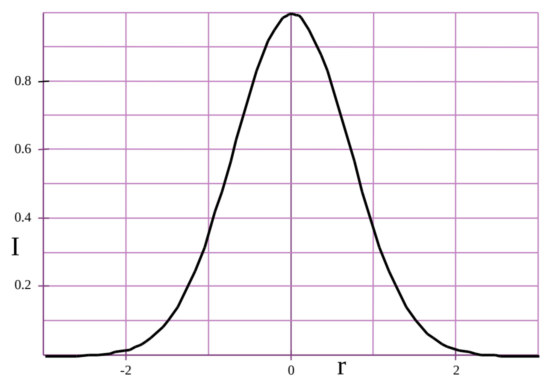

Figure 1
Gaussian spot intensity profile

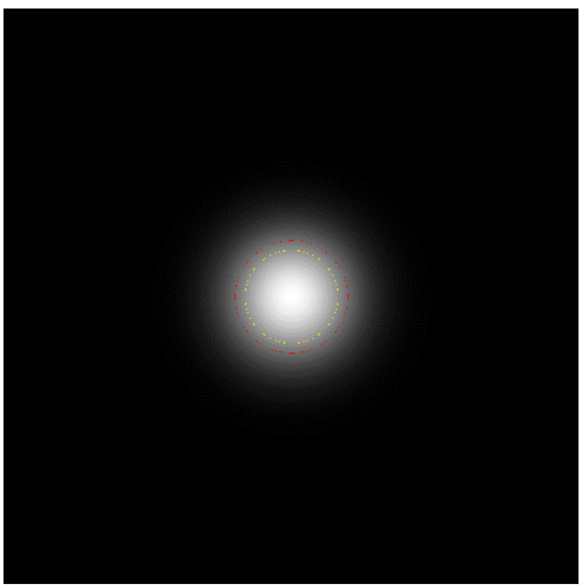

Figure 2
Gaussian spot with 50% (inner) and 1/e intensities marked

Now it is very unlikely that this is the actual exact intensity profile of the spot. Apart from the fact that no device is perfect, there are other effects which result in the spot not being perfectly symmetric and not being optimally focussed. Even so, the gaussian profile is a useful model for fuzzy pixels, and the gaussian radius, R_g, is used as a characterizing parameter. The actual physical distance units are not really needed for much of the discussion of scanlines and rasters. Instead, the normalized gaussian radius unit, R_g, will be used. The conclusions can then be scaled to any specific system whose spot characteristic is known.

Many imaging devices do not expose individual pixels. Instead, an entire scanline is swept and the intensity is modulated. Television is the best known example of this. Each scanline in a flat field image will be uniformly bright horizontally, but will have a vertical intensity profile. This profile turns out to be gaussian, with the same gaussian radius as the isolated spot. The intensity maximum of the scanline is $\sqrt{\pi}$ greater.

Contrast index, and threshold of discrimination

If one conducts an experiment to squeeze the scanlines in the raster closer and closer, eventually the individual scanlines will become indistinguishable. This is a measurement method known as the “shrinking raster”, and it is used to deduce the spot diameter of say, a CRT. We will use it in reverse, to determine the limit on how far apart the scanlines may be placed before the raster structure becomes visible.

To do this, we define a “contrast index” which attempts to quantify slight changes in intensity:

$$\mathbf{C}(\mathbf{y}) = \frac{\mathbf{I}(0) - \mathbf{I}(\mathbf{y})}{\mathbf{I}(0)}$$

This is the normalized difference between the intensity at location y and that at the center of the scanline (at y=0). It represents the relative change in the intensity from the maximum. The darkest point in the raster will be found halfway between two scanlines. Figure 3 shows the contrast index when plotted at this location as a function of scanline spacing in a raster. We see that it is close to zero until the scanlines start to become separated by more than one gaussian radius.

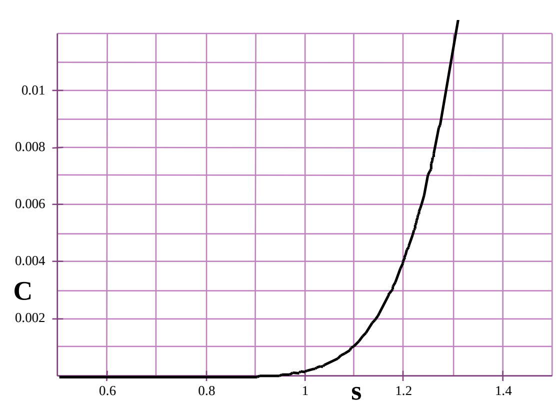

Figure 3
Contrast index as a function of line spacing

So how much contrast can we tolerate? Experiments in human visual perception have shown that people are not exceptional at identifying color or intensity differences over a large area. But they are extremely sensitive to slight shifts when they are right next to each other. In fact a good rule of thumb is that one can perceive about a one-half percent change between one visual sample and a nearly identical one. This applies to each of three coordinates, hue and saturation, as well as brightness. This has implications in how many colors are needed to make a smooth color ramp, but in the case of scanline density, it means that the contrast index as defined above, may not exceed 0.005 without the scanlines being noticeable.

The plot shows that staying within a separation of 1.2 R_g will assure us of a raster-free field. A plot of the full raster intensity is shown in figure 4. Here the raster lines diverge from one unit of separation. As the separation increases, the overall light intensity falls and at a separation of two gaussian radii, the individual scanlines are clearly discernible.

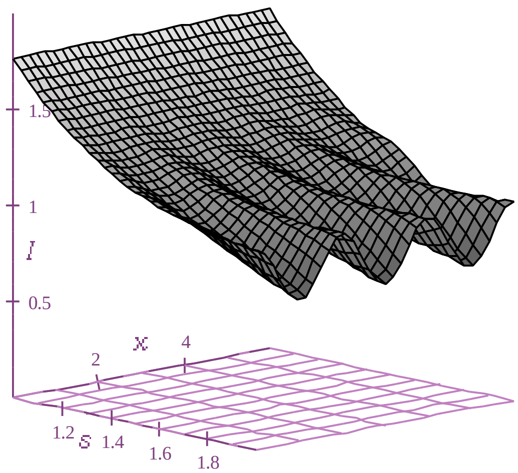

Figure 4
Scanline intensity over position as a function of line spacing

Figure 4 is a plot of an intensity profile. What does this really look like? To gain a feel for this, as well as to confirm our contrast threshold of 0.5%, we created a simulation of a raster and reproduced several samples for varying spot separations. The patterns (figures 5, 6, 7) show an area of a uniform field where the pixels stop (this could be the lower right corner of a stop sign). They also graphically depict the nature of fuzzy pixels, and the futility of obtaining razor sharp edges.

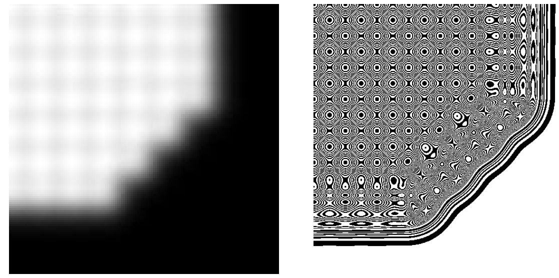

Figure 5
Scanline separation of 1.5 R_g

At a spot separation of 1.5 R_g, the pixels are easily seen. (Although the reproduction onto printed paper may not show this, film and CRT displays will). Keeping the spot size constant, their density is increased to where the pixels are separated by 1.2 R_g. At this point, most observers will not be able to detect the locations of the pixels. The contrast index is less than 0.5% . The uniform field appears flat, and the horizontal and vertical edges seem straight. A contour plot more clearly shows the actual intensity levels between black (0) and white (255). The data is identical but the contour image has painted all even intensity levels black, and odd ones white.

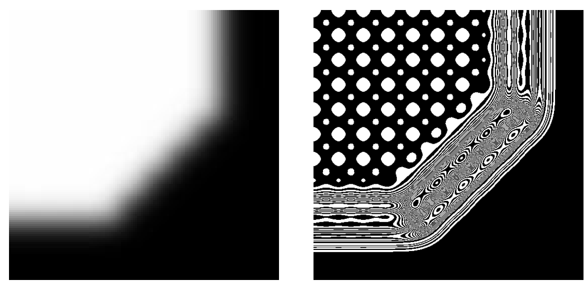

Figure 6
Scanline separation of 1.2 R_g

Even though the scanlines have visually merged, some observers will be able to detect a slight "lumpiness" along the diagonal section. One might expect that the pixel separation would have to be reduced further by the factor of $\sqrt{2}$ to account for the distance between pixels along the diagonal. This does not prove to be the case however. Empirically we find that a separation of 1.0 R_g is adequate to smooth the edge.

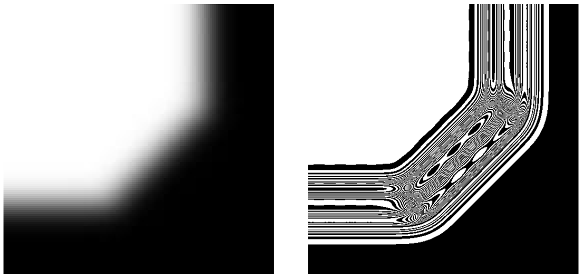

Figure 7
Scanline separation of 1.0 R_g

Some confusion about sharpness and resolution

We use the term *sharpness* as a qualitative indication of the visual acuity of the image. An image which is very sharp has very crisp, clean looking edges whereas an image which is not sharp will have soft blurry features. Often, the sharpness of an optical system is characterized by its *limiting resolution*, measured according to how closely one can place alternating black and white lines and still be able to detect that they are present in the image. The units "line pairs per millimeter" or "cycles per millimeter" usually quantify it. The more cycles per millimeter that can be detected, the higher its limiting resolution, and the sharper will be the image perceived.

Obviously if one can detect the raster of an image, the raster density is within the limiting resolution of the system. A test image that paints even scanlines white and odd ones black would be easily visible when the scanline spacing was 1.0. It would disappear when the spacing was reduced to 0.6 R_g. So the limiting resolution when expressed in the dimensions appropriate to scanlines is approximately 0.8 line pair per gaussian radius (one cycle in 1.2 R_g). Notice that sharpness is independent of scanline spacing; we only considered varying the spacing in order to measure where the limiting resolution occurs. The sharpness is a function of the spot radius only, not the scanline density.

It is easy to confuse this use of the term *resolution* with how we generally use it with digital pictures. It is common to refer to the number of pixels across or down in an image as its resolution. This is a measure of its *digital resolution* since these dimensions are how one locates any point in the image. The digital resolution relates to how accurately an image is represented by the set of pixel samples it comprises. The more samples, the less the aliasing artifacts (jaggies) will be. This is a separate issue from sharpness as defined above, though the visual effects of undersampled (aliased) images will often be described as "reduced sharpness." A full discussion of digital image sampling and reconstruction is beyond the scope of this paper, but is discussed in [MIT88].

Implications for number of bits required in a deflection system

Modern film recorders usually have a digitally based deflection system. The scanlines are positioned according to the contents of a digital to analog converter (DAC). This has many nice benefits for control and repeatability, but a major drawback is that scanlines can only be placed at discrete positions, according to the digital number contained in the DAC.

Depending on factors such as number of scanlines used, optical magnification, correction for pincushioning etc., it is possible that the prescribed location of a given scanline cannot be achieved. The desired position could be as much as one-half bit from the nearest possible position that the DAC can accommodate. It can also be true that the neighboring scanline could be one-half bit from its proper location *in the other direction*. This adds up to a possible separation error of a full "DAC count". The difference between the exact desired position and the possible discrete positions accessible by the DAC is generally referred to as *digital roundoff error* and is the subject of much study in numerical analysis. We will examine its visual effects on the raster.

Having established that the contrast visibility threshold is 0.5%, how small must the DAC count be in order to not see its effect? An extra DAC count inserted between a scanline and its neighbor will certainly be visible if the nominal scanline spacing is already right at the threshold. There is no tolerance for roundoff error in this case.

By placing a gap into the middle of an otherwise uniform field, we can evaluate the contrast index and find the regions that are below the visual threshold. Figure 8 is a plot of iso-lines of contrast index versus scanline separation s, and the inserted gap distance d. It is verified on this plot that no amount of roundoff error is tolerable when the scanline separation is 1.2; the 0.005 index contour intersects it.

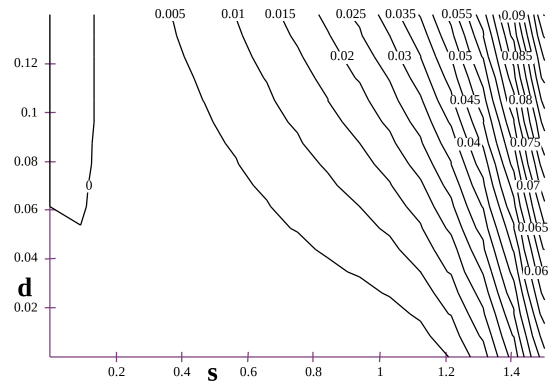

Figure 8
Iso-lines of contrast index as function of spacing s and gap distance d

At the tighter spacing of unit separation, a gap of up to 0.025 can be tolerated without exceeding the threshold. This means that 40 DAC counts of resolution are needed to represent one gaussian radius of distance. At this separation, using 40 counts between adjacent lines it is tolerable to have one count of digital roundoff noise without being able to detect it in the raster.

Let's say that we use a DAC that supplies 32 counts over a scanline separation distance of 0.9. This falls in the safe region of the chart. Now if there are 4096 scanlines, the minimum number of bits required will be 17. Good design practice will require the deflection system to position the beam off the image area, so add one more bit to obtain the extended range. It is seen that an 18-bit DAC is required to obtain a 4K image having no digital roundoff artifacts in the raster! Higher resolution images (8K and 16K) require even further precision.

The gaussian spot simulations show the sensitivity we have to this type of raster artifact. To illustrate the effect a scanline separation of 1.2 R_g is used. When a gap of one-tenth scanline is inserted, the result is a dark line through the center of the field (figure 9). The intensity contour plot shows another interesting effect. Even though the gap was a fraction of a scanline wide, the intensity drops over a region of four or five scanlines. This is a blessing and a curse. It means that the intensity falloff occurs over a wider region making the contrast difference more gradual and less visible. On the other hand, if it is a large enough drop to be noticed, it will *really* be noticed; it's four scanlines wide!

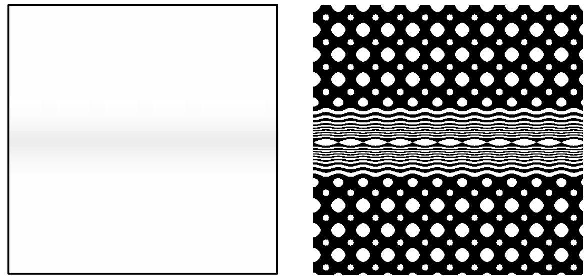

Figure 9
Inserting a 1/10 scanline gap when s = 1.2 R_g

This discussion about the visibility of digital roundoff noise also applies to other noise sources. Electronic signals always bear some amount of low level random noise. If this noise is large enough, it will appear in the raster, often as a "weaving" texture. How much noise is too much? Clearly, if it exceeds one DAC count, when the system has been set up with one DAC count of margin, it will be seen. One DAC count in an eighteen-bit DAC is 4 parts per million when related to its full scale range. One count out of 2^{18} can also be expressed as a signal to noise ratio of 108 decibels. This is a demanding level to the circuit designer responsible for the deflection system.

What does this mean to movies?

So what does all this mean to the moviemaker trying to determine the appropriate raster format for computer generated scenes? Clearly, he does not want to end up with a scene having television-like raster structure. Nor does he want to waste compute time on pixel overkill to guarantee raster-free images.

By knowing the characteristics of the imaging device, in particular the spot dimensions as recorded on the film, he can reliably predict how many scanlines are needed to assure a smooth raster-free picture. He will also know what sharpness to expect and will not waste production time by computing more pixels in the futile hope of gaining edge sharpness. (One can argue for more pixels to help with antialiasing, but there are other more efficient methods for this).

Here is an example to illustrate this. Lets say that the film recorder has a characteristic gaussian spot which has a 0.5 mil gaussian radius (2 mil diameter visible spot) at the phosphor of a CRT which has an image height of four inches. Lets further pretend that the lens is perfect and the film has infinite resolution (we’ll worry about them later). With this size spot, and assuming the 4 inch CRT dimension ends up as 0.63 inches on a 35mm academy frame, we know the limiting resolution (sharpness) will be 400 line pairs per millimeter on the film. Using the guide that the scanlines be placed 1 gaussian radius apart, fully 8,000 scanlines are required to fill the image area and prevent the visual detection of the raster!

Most motion picture images are not computed at 8K resolution. What saves them from this requirement is the fact that lenses aren't perfect, and neither is film. These elements effectively enlarge the spot, so that they may be placed further apart and fewer are needed to fill the image area. Some film recorders help out by actually scanning more lines in the raster than the data contains, pixel replicating in order to fill up the area.

The data-late streak

Another important and annoying artifact is known in our lab as the dreaded "data-late" line, also known as "streaking" or "dark lines". It shows up as an isolated darker scanline or group of scanlines among an otherwise fine looking picture. It occurs whenever the stream of data to the film recorder is interrupted momentarily and the exposure of the film must pause. When the data for the next line of the picture finally arrives, the exposure resumes, but by then it is too late. When the film is developed, the delayed scanline will be slightly darker than its neighbors.

We have investigated this, initially thinking that some cooling effect was occurring in the phosphor of the CRT we use to expose the film. Instead, having failed to measure the light falloff with photometers, we now feel it is an effect in the film, similar to reciprocity failure, called "intermittency effect". It is known by professional photographers who use multiply timed strobe flashes to light their subject.

An ideal film would be a perfect light integrator. Flashing a strobe twice to obtain double the light energy should result in a photograph that is twice as bright. This actually does hold true when the flashes are simultaneous. But if one is delayed, the effective speed of the film is reduced. The longer the delay, the less bright the resulting photograph. The photographer knows this and compensates by opening the f/stop slightly.

It seems that film is not a perfect integrator. It has a loss factor which depends on time. The film recorder, which today at least must expose the film scanline by scanline (instead of flashing the entire image instantaneously onto the film), is subject to the same laws of photochemistry as the photographer. We have seen how the raster exposure must blend adjacent scanlines together in order to form a uniform field. Each scanline occurs on a schedule separated by a few milliseconds.

Consider a single fixed point on the surface of the film being exposed to a uniform flat picture. As the raster approaches, it starts to see small pulses of light separated by the amount of time between scanlines. As the raster passes over, the pulses are significant for three or four scanlines. When all of the scanlines arrive in a timely manner, each point of the film has seen pretty much the same exposure schedule as any other point on the film, and the result is a uniform field. If however, one scanline is delayed with respect to its predecessors, the points on the film along that scanline image have seen a different exposure schedule than their neighbors. The result is a reduction in film speed similar to the photographer's, but localized to the immediate neighborhood of that scanline. It shows up as a darker line easily identified in the picture. Figure 10 shows some experimental data relating the density increase to the delay in exposing the next scanline.

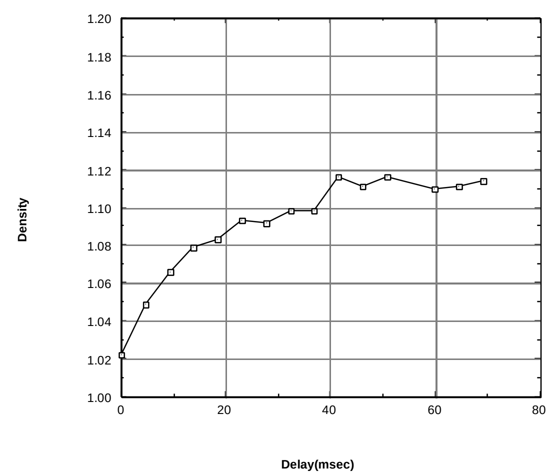

Figure 10
Density increase due to data-late conditions

What is the cure? Don't allow your scanlines to be late! It is tempting to try and compensate for delayed data, like the photographer opening the f/stop. With CRTs, this is fraught with other problems, such as finding a beam current increment that will exactly compensate for the delay. Even if one could be found, it would only be good for that delay on a specific film type and one brightness level (due to the square-law characteristic of CRTs). All the other colors would still shift.

No, the practical solution is to never get off schedule. Laser printers have a similar constraint but for a different reason. Once the paper starts moving, it must stay moving or alignment and uniformity problems result. The solution was to build a full-frame buffer into the printer. This is the only bullet-proof solution for film recorders as well, but because of the size of full-color images, the buffers sizes become enormous! The expense of such a buffer is rapidly falling, and it will not be long before they will be found in film recorders. But in the meantime, the burden is on the host computer to keep supplying data at the rate it is being exposed.

There are a few things that can be done. An obvious one is to slow the film recorder down to the rate that can be reliably imaged without "data-late" occurrences. This is unsatisfying for two reasons. First, the worst-case image determines the rate of all of the frames. Second, there are zillions of frames to shoot for a movie and never enough time to do it. Slowing the frame exposure rate slows the production schedule.

One of the most effective solutions is to provide a dedicated I/O channel to the film recorder, one which has no other traffic on it. Careful software driver design will allow a reliable deterministic frame rate to be maintained. When full frame (or at least full color pass) buffers become available on film recorders, the need for this effort at the system level will go away and the data-late line will fade into history.

Conclusions

We have covered a select few topics in the use of film recorders for making digital pictures. In particular for making movies without raster line artifacts we can state the following:

1. People can detect intensity differences of 0.5%.

2. Gaussian scanlines must be separated by no more than 1.2 R_g to not detect the raster.

3. Image sharpness is a function of spot dimension only. When measured by limiting resolution, it is 0.8 cycles per R_g.

4. Aliasing jaggies, which are a function of digital resolution may be reduced by increasing the pixel density (higher digital resolution, or smaller image size).

5. Digital roundoff error or circuit noise in the deflection system may cause scanline offsets which are highly visible because the intensity reduction extends over the neighboring scanlines. To counteract these, the scanline spacing must be reduced to below 1.2 R_g to provide margin for error.

6. Darker scanlines occur when the digital data arrives too late and the film exposure process experiences a delay.

Hiding the raster is only one of a number of topics that are important to making digital movies. We will continue to investigate in other areas such as smooth color shading and edge transitions. Eventually, we would like to watch a movie and be unable to identify the synthetic scenes from the real ones.

Acknowledgements

I would like to thank Loren Schoenzeit for his assistance with the LanSlide system on which the hardcopy versions of the gaussian spot simulations were made. Thanks also to the MGI staff members who submitted to the visual perception experiments on which this paper is based.

References

[MIT88] Mitchell, Don P., and Netravali, Arun N, "Reconstruction Filters in Computer Graphics", *Computer Graphics*, Vol 22, No 4, August 1988 (SIGGRAPH 88 Proceedings), p221.

[HUN87] Hunt, R. W. G., *Measuring Color*, John Wiley and Sons, New York, 1987.