LETTERS TO THE EDITOR

Dear Editor:

I was pleased at first to see the article on two-pass transforms in your January 1986 issue ("A Nonaliasing, Real-Time Spatial Transform Technique," Karl M. Fant, pp.71-80). However, I was unhappy to find that the author had apparently misunderstood the main gist of the argument Ed Catmull and I presented in our original 1980 article on the subject ("3-D Transformations of Images in Scanline Order," *Computer Graphics* [Proc. SIGGRAPH 80], Vol. 14, No. 3, July 1980, pp.279-286). Fant indicates that the filtering and sampling of a transformed image was of only "secondary concern" to us. On the contrary, the reason we found the two-pass technique so interesting was that it made solution of the aliasing problem easy—in hardware and software—reducing it from a 2D problem to a sequence of 1D problems. We stated this in the opening paragraphs of our paper, and illustrated it with figures of rotation, scaling, perspective, and bilinear patch mappings, which were all internally antialiased, completely and continuously.

The technique we used for antialiasing in our 1980 paper was the standard application of sampling theory on which the author expands for the main body of his paper, except we used second-order filtering (triangle) rather first-order (box) for better antialiasing. We used the same trick as the author of changing the filter width along the scan line to get perspective antialiasing.

I would also direct the author's attention to Ampex Corporation, which implemented the concepts in their ADO real-time video effects product which does rotation, scaling, and perspective in real time with the two-pass technique and appropriate antialiasing.

Sincerely,
Alvy Ray Smith, Vice President, Pixar

The author's reply:

The contents and results of the two articles are quite different. Each appeals to a fundamentally different theoretical view of the discrete image. From the evidence of the article, the letter, and the Ampex ADO system cited as an example, and with which I am familiar, the Catmull-Smith article assumes the standard sampling theory view of the image.

Within standard sampling theory the image is viewed as a field of uniformly spaced points. Each point represents a sampled intensity value (pixel). The intensity of each pixel is assumed to be concentrated in its corresponding point. These pixels are assumed to have been sampled from a continuous surface of intensity value and represent that surface. It is assumed that this continuous surface can be approximately regenerated from the sampled points and that an approximate intensity value for any point on this regenerated surface can be calculated from sample points neighboring the point of interest. It is assumed that including more neighboring sample points in this computation increases the accuracy of the approximate intensity value. Using only the single nearest sample point is least accurate and is called zero-order interpolation. Using the four nearest points raises the order of interpolation and presumably the accuracy of the intensity approximation for the point of interest. Sample points closest to the point of interest contribute more to the approximate intensity value than sample points farther away from the point of interest. This weighting of influence is represented by a set of coefficients that are applied to the intensity of each neighboring sample point to calculate the approximate intensity of the point of interest.

In performing a transform, the output pixel point location is mapped to a point in [the] domain of input pixels. Interpolation is then performed in the neighborhood about this point to determine the approximate intensity for that point on the surface of the input image. These interpolation neighborhoods are generally of constant size for all transforms. Indeed, their rationale is independent of the transform. The larger the neighborhood or order of interpolation, the more accurate the approximation, regardless of the specific transform. For the typical transform there will be considerable overlap between these interpolation neighborhoods. Most of the input pixels will contribute in varying degrees to several output pixels, and some input pixels will contribute more than others to the output image. In the case of a shrink there is a point at which these regions separate and there are input pixels that do not contribute at all to the output image. This internal texture aliasing phenomenon, in which input image features disappear because of gaps between the interpolation neighborhoods, is clearly apparent on the Ampex ADO system.

In mapping output pixel locations to the input domain, each output pixel either falls on the surface of the input image, in which case it is entitled to a full intensity value, or it falls outside the surface of the input image, in which case the output pixel intensity is zero. There is no concept of partial intensity output pixels based on their proximity to the edge of the surface of the input image. This results in a stairstep effect where the edge of the input image appears in the domain of the output image. This edge aliasing is generally dealt with by a separate function, called antialiasing, which creates partial intensity output pixels along these edges in the output image to smooth out the stairstep effect along the edge.

The image transform process within this theoretical view first maps each output pixel to the input image domain. Then the intensity of the output pixel is approximated by interpolating the input pixels neighboring the mapped output pixel location. Then antialiasing is applied to the output image to smooth the edges of the input image in the out-

Continued on p.8

LETTERS

Continued from p. 3

put image. The per-pixel computation can be very expensive, depending on the complexity of the spatial transform, the order of the interpolation filter, and the antialiasing technique used.

The algorithm presented in my article views the image as a rectangular array of identically sized squares or rectangles. Each rectangle represents a pixel. The intensity value of each pixel is represented as uniformly distributed over the surface of its rectangle. Each pixel is a perfectly uniform emitter over its surface. The surface of the image is filled with intensity. There is no between pixels. Every point on the surface is inside a pixel. For a transformation, the inverse mapping of each output pixel projects a footprint onto the input image. The output pixel is viewed as a detector that is uniformly sensitive over its surface.

Is there any reason to consider that any specific point or pixel under the footprint represents the value of the output pixel more correctly than the other points or pixels under the footprint? For instance, should pixels farther from the center of the footprint be given less influence in determining the intensity of the output pixel than pixels closer to the center? The answer is no. What is desired is a uniformly weighted average intensity value under the footprint of the output pixel. The intensity detected at the output pixel is the average intensity of the emittance of the portion of the input image bounded by the footprint of the output pixel. This is the intensity integral under the footprint, divided by the area of the footprint. This is precisely what the algorithm presented in my article computes. Duff and Pavlidis in their letter [see *CG&A*, April 1986, Vol. 6, No. 4, pp.66-67] understood this conceptual view of the image underlying the algorithm. I don't think this view can be called "standard" sampling theory.

In calculating the integral of the intensities under the footprint of the output pixel, does it make sense to approximate a continuous intensity surface or is it best to just use the discrete stepped surface of the raw image? We don't have the problem of standard sampling theory of points on the surface with no assigned intensity. Every point on the surface of the image has an intensity, so the only question is the quality of the intensity integral under the footprint of the output pixel.

Does approximating a smooth surface to calculate this integral provide a value for the output pixel that is better than the value obtained from the discrete stepped value image? Since what we are after is the average intensity and the intensity integral gets divided by the area of the footprint, the precision of the integral does not seem to be critical to the fidelity of the transform. Furthermore, higher order surface fitting functions involve considerable attenuation of high-frequency components of the image, which is not necessarily desirable. If filtering of the image is desired, it should be performed explicitly and not as an implicit side effect of intensity resampling.

It appears most desirable, from the point of view of the fidelity of the transform, to perform image intensity resampling to consider high-order interpolation functions.

The one-dimensional version of this view is described in detail in the article. INSFAC is the footprint area for each consecutive output pixel. INSEG can be viewed as a fractional position pointer into the input pixel stream. Input pixels are summed into the accumulator from the current input stream location (INSEG) until the footprint area (INSFAC) has been accumulated. It is then multiplied by SIZFAC, which is the same as dividing by INSFAC to provide the average intensity under the footprint of the output pixel.

The interpolation neighborhoods, or footprints, of each output pixel are contiguous. There are no gaps between interpolation neighborhoods and there are no overlaps between neighborhoods. The transform is complete in that every input pixel under the footprint of the output image contributes equally to the output image. No input pixel contributes more or less than any other input pixel to the output image. If a 512×512 image is shrunk to one output pixel, all 262,144 input pixels are averaged into the single output pixel. As pointed out above, interpolation neighborhoods within the standard sampling theory view can overlap and separate. Input image pixels do not uniformly contribute to the output image. This leads to internal texture aliasing, which does not occur with my algorithm.

Output pixels whose footprints fall off the input image receive partial intensity values according to how much of their footprint covers the input image. This is accommodated by the simple expedient of setting the initial value of OUTSEG to a fractional portion of INSFAC. The algorithm inherently produces unaliased edges, so the antialiasing step is not required. As mentioned above, aliased edges are an inherent product of the standard sampling theory view.

During expansion, the footprints of several neighboring output pixels can map into a single input pixel, providing identical intensity values to the output pixels and creating a blocky image. The one-pixel-wide input-pixel averaging discussed is provided solely to smooth out the blockiness during expansion to achieve a smooth fuzziness in the expanding output image and avoid the blockiness. This simple averaging window works quite well. It has not been considered advantageous to pursue "higher order" fuzziness. The averaging window should be disabled during shrinking.

The algorithm is considered nonaliasing because it does not exhibit the internal texture aliasing and the edge aliasing that the standard sampling theory approach exhibits. The algorithm is an exact implementation of the theoretical view and provides a precise and direct mapping from one discrete domain to another discrete domain.

I hope these comments will clarify the differences between the two articles.

Karl M.Fant
Honeywell