554730 p10

N90-16209

CONTEXT DEPENDENT ANTI-ALIASING IMAGE RECONSTRUCTION

P. R. Beaudet

Westinghouse Electric Corporation
Advanced Technology Division
P.O. Box 1521, MS 3D12
Baltimore, MD

A. Hunt and N. Arlia
Westinghouse Research and Development Center 601-1B44C
1310 Beulah Road
Pittsburgh, PA

ABSTRACT

Image Reconstruction has been mostly confined to context free linear processes; the traditional continuum interpretation of digital array data uses a linear interpolator with or without an enhancement filter. In this paper, anti-aliasing context dependent interpretation techniques are investigated for image reconstruction. Pattern classification is applied to each neighborhood to assign it a context class; a different interpolation/filter is applied to neighborhoods of differing context.

It is shown how the context dependent interpolation is computed through ensemble average statistics using high resolution training imagery from which the lower resolution image array data is obtained (simulation). A quadratic least squares (LS) context-free image quality model is described from which the context dependent interpolation coefficients are derived.

It is shown how ensembles of high resolution images can be used to capture the a priori spacial character of different context classes. As a consequence, a priori information such as the translational invariance of edges along the edge direction, edge discontinuity, and the character of corners is captured and can be used to interpret image array data with greater spatial resolution than would be expected by the Nyquist limit. A Gibb-like artifact associated with this super-resolution is discussed. More realistic context dependent image quality models are needed and a suggestion is made for using a quality model which now is finding application in data compression.

I. INTRODUCTION

The work presented in this paper builds upon theory[1] that was developed further at the Westinghouse Advanced Technology Laboratory and more recent work at the Westinghouse Research and Development Center. The goal of this work is to develop optimal adaptive methods of interpreting image data. By matching the interpretation function to the local characteristics of the scene, a context dependent interpreter is designed which offers superior performance over context independent interpolation functions such as bilinear and cubic convolution.

When applied to sampled image data, this context dependent interpolation function yields an image which is free of the aliasing artifacts caused by image frequency content too high for the sampling frequency. Because of its ability to recognize familiar patterns in the sampled data before its interpretation, anti-aliasing interpolation selects the interpretation which is most probable given the a priori knowledge of context class patterns. This interpretation process is sometimes referred to as super-resolution.[2] The benefits of super-resolution are

- o Provides a method of contextually and artificially increasing the sampling frequency from which the known system modulation transfer function can be better compensated.
- o Gives a better procedure for image zoom.
- o May lead to adaptive methods of image gathering such as is provided in nature through eye movement and neural pre-processing

In Section II, a distinction is made between data interpretation vs data interpolation. Here, rationale is given for pursuing this work and the basic theoretical approach is given. In Section III, an experiment is described designed to show what benefits might be expected from super-resolution. In Section IV, results are presented of context dependent interpretation and some of the resulting artifacts are discussed. The discussion continues in Section V where a basis for future work is provided and a discussion of a more realistic quality model is presented.

II. INTERPRETATION VS INTERPOLATION

Images are often defined by their fourier content.[3]

$$I(x) = \iint_{-\infty}^{\infty} \frac{d^2\mathbf{k}}{(2\pi)^2} I_{\mathbf{k}}\, e^{\iota\mathbf{k}.\mathbf{x}}$$

where $\mathbf{x}$ and $\mathbf{k}$ are 2-vectors, $I(\mathbf{x})$ is the image and $I_{\mathbf{k}}$ is its fourier transform. A sampled image data set can be written as the sampling of an image at integer (or periodic) valued of $\mathbf{x} = \mathbf{i}$. (We select $\Delta x = \Delta y = 1$ throughout this paper.)

$$I(\mathbf{i}) = \iint_{-\infty}^{\infty} \frac{d^2\mathbf{k}}{(2\pi)^2} I_{\mathbf{k}}\, e^{\iota(\mathbf{k}.\mathbf{i})} = \sum_{\mathbf{m}} \iint_{-\pi}^{\pi} \frac{d^2\mathbf{k}}{(2\pi)^2} I_{(\mathbf{k}+2\pi\mathbf{m})}\, e^{\iota(\mathbf{k}.\mathbf{i})}$$

The sampled "image" generated by frequency $\mathbf{k}$ of unit amplitude and that generated by $\mathbf{k} + 2\pi\mathbf{m}$ for any integer 2-vector $\mathbf{m}$ are the same; A fundamental frequency $\mathbf{k}$ is indistinquishable from any of its aliasing frequencies $\mathbf{k} + 2\pi\mathbf{m}$. As a consequence, it is really impossible to determine the true frequency content of an image without some a priori knowledge. The typical engineering assumption that is made is that the true image has no frequencies larger than the Nyquist frequency $|k_x| < \pi$ and $|k_y| < \pi$.

This assumption is most often wrong and forcing it to be correct by placing a smoothing filter in the image gathering process may result in loss of information (interpretable spacial resolution). It does simplify the display process, however, which is then unambiguous.

To get a better appreciation of the information loss that is possible, consider the images of figures 1a and 1b. Here, a road, pipe, cable or other narrow-object illuminates a diagonal set of pixels which if interpreted according to the Nyquist assumption would lead to a display (interpretation) consisting of a string of blobs one for each diagonal pixel. It is a goal of this work to interpret this data more as a human might do as illustrated in Figure 1b. What other assumption than Nyquist could be used to better accomplish this human-like interpretation of the data? Surely it is that almost everywhere in a scene there is some direction of minimal spacial frequency, while its orthogonal direction may have a very high frequency content; frequencies even higher than the Nyquist frequency.

By data interpretation,[4] it is meant the generation of a continuous image function $I(\mathbf{x})$ from a sampled subset $I(\mathbf{i})$ <u>taking into consideration the directions in the image data over which there are minimal/maximal changes</u>. Such a process is context dependent because the interpolation process depends upon the scene patterns. In contrast, by data interpolation it is meant a context free interpolation of the data such as is provided by the sinc function interpolation based upon the Nyquist assumption.

$$I(x,y) = \sum_{i,j} I(i,j)\operatorname{sinc}(\pi|x-i|)\operatorname{sinc}(\pi|y-j|).$$

To implement a context dependent interpretation process, the neighborhood of each data sample must be classified into one of many context classes, K. This can be done by computing the local gradient and classifying based upon gradient magnitude and direction. More complex classes are also possible for images containing lines, line ends, corners, etc. For each context class, K, the interpolation formula

$$I(x,y) = \sum_{\mathbf{i}} I(i,j)\, g_K(x-i,y-j)$$ is used where g_K is the interpolation

coefficients (function) matched to the context class. It is anticipated that such a process would be capable of distinguishing the line-like objects of figure 1 and provide for a more human-like interpretation. The extent to which this is possible is the subject matter of this paper.

III. THE Experiment

An experiment was designed to determine the extent to which a priori knowledge of scene content could be used to improve sampled data interpretability. A real high resolution television picture was taken with a CCD camera, digitized to obtain a 512x384 digital image, and averaged over sixteen frames to reduce noise. It was an image of a white piece of rectangular cardboard tilted by about 30° relative to the camera axias (see figure 2). 15x12 blocks of pixels were averaged to obtain 180 coarse images of the scene, each 32x32 pixels. These 180 coarse images varied in the manner (phase) by which the data were averaged from the finer resolution data. The scenes were contextually classified as illustrated in Figure 3, and one of the coarse images was contextually interpreted to achieve the high resolution image shown in Figure 4. The philosophy for deriving the super-resolution interpretation functions is subsequently presented. This philosophy uses a least squares image quality model which is believed to be at the heart of the Gibbs-like[5] artifacts seen in Figure 4.

The interpreted function for each context class, K, is expressed as

$$I(x,y)=\sum_{i,j,\in N} I(i,j)\; g_K(x-i,y-i)$$

where N was selected as a 5x5 neighborhood centered at the pixel nearest the point x,y. The fine grid (512x384) was used for discrete points within each pixel.

The interpolated image was compared to the original (ground truth) data in a least squares manner yielding a cost function:

$$C = \langle (I(x,y) - \tilde{I}(x,y))^2 \rangle_K$$

Here $\langle\ \rangle_K$ is an ensemble average over all 180 images at all defined context class K centers. To obtain the "optimal" interpolation function, we simply take a functional variation of C with respect to g_K (x-i, y-j) giving a system of <u>normal equations</u> which decouple for each subpixel location x,y. The system of normal equations is

$$\sum_{i,j\in N} \langle I(i,j)I(i',j')\rangle_K\; g_K(x-i',y-j') = \langle I(x,y)I(i',j')\rangle_K$$

For each x,y and K, this is a system of twenty five equations for the contextual interpolation coefficients $g_K(x-i,y-j)$ to be applied at the 5x5 array in the neighborhood of each pixel classified as K to achieve the interpolation value I(x,y). The Matrix $\langle I(i,j)I(i',j')\rangle$ and vector $\langle I(x,y)I(i',j')\rangle_K$ are ensemble averages over the 180 processed coarse images and the original fine resolution image.

IV. Results

The results of this first experiment is shown in Figure 4. A context free bilinear interpolated image is shown in Figure 6. Clearly, the context dependent process retains the translational invariance along the edge and is much "sharper" than the context free bilinear interpolator which also shows the staircase aliasing artifact. But Figure 4 has a Gibb-like artifact which in itself is a distraction. Surely as humans, we wouldn't interpret the coarse data shown in Figure 5 with these Gibb-like oscillations! Where do they come from?

To understand the results of this experiment, all we really need to recognize is that the data are noisy.

A sharp edge discontinuity with a white noise background should have to be filtered a-la Weiner[6] if based upon a least squares fidelity criteria. The Weiner filter is a very sharp filter and will essentially truncate all spatial frequencies whose amplitude is below the noise level while preserving all those above the noise level. Figure 7 illustrates the response of an edge function to such a process. The high frequency truncation does a best-least squares job but results in the Gibb oscillations. It is believed that this Gibb's phenomena is the process at work here and results from the implicit assumption that errors near edges are just as important as those away from the edge; (least squares criteria).

V. Discussion and Future Work

The Gibbs artifact can be subdued if a different model to an image quality measure is taken. The sharp discontinuity of an edge, together with the least squares criteria, places a severe restriction upon the approximating function and admits short but large excursions from the edge function. Another image quality measure may prove more effective in yielding an edge approximation which is more in keeping with a human interpretive approach. It is a quality model which is finding application in DPCM data compression. The model considers human toleration to a slight "jitter" of the pixel (sometimes called rate distortion). Any error in the approximating function is compared not just to the expected noise variance, σ^2, but to the noise variance plus rate distortion. If $\Phi(x)$ is the approximating function to I(x), then $(I(x)-\Phi(x))^2$ must be compared to

$$\sigma^2 + h^2 \left| \frac{d\Phi}{dx} \right|$$ where h is the variance equivalent subpixel jitter.

The modified cost function is then $\left< \dfrac{(I(x) - \Phi(x))^2}{\sigma^2 + h^2 \left(\dfrac{d\Phi}{dx} \right)^2} \right>$.

This is a nonlinear functional which in the limit of very low contrast edges (noisy edges) leads to the previous least squares measure. But for large contrasts, the model compares the error to the slope of the approximating function. In this high contrast case, an edge gets approximated by the exponential function $\Phi(x) = 1-e^{-\alpha x}$ which distributes the representation error uniformly and provides the type of solution one might expect a human to choose. This edge function is shown in Figure 8. It trades off a more uniform transition in exchange for a sharper drop near the edge. The solution is forced to be a smooth transition because if

$\dfrac{d\Phi}{dx} = 0$ anywhere, then any error there is given a very large weight.

References

1. Beaudet, P. R., "Context Dependent Interpolation" Image Science Mathematics Symposium. November, (1976), Monterey, California. Editors Carrol O. Wilde, Eamon Barrett.

2. Marple, S. L., "Digital Spectral Analysis with Applications" Chapter 16, Prentice Hall, 1987.

3. Rosenfeld, A., Kak, A. C. "Digital Picture Processing" Academic Press, 1982.

4. Beaudet, P. R., "Context Sensitive Modeling", Proceedings of the ERIM International Symposium on the Remote Sensing of the Environment, Anne Arbor, Michigan, May 1981.

5. Wylie, C. Ray "Advanced Engineering Mathematics" McGraw Hill, 1975

6. Wiener, Norbert "Time Series" The M.I.T. Press, 1964

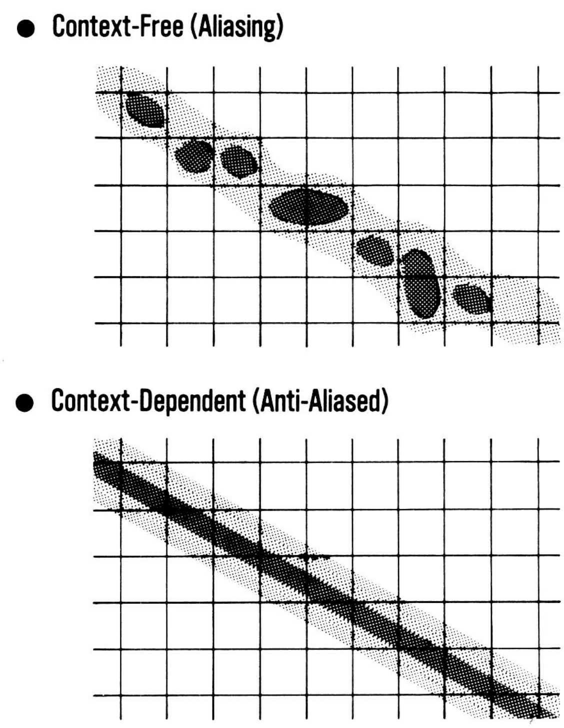

Figure 1a. Aliasing Artifacts (What is Aliasing)

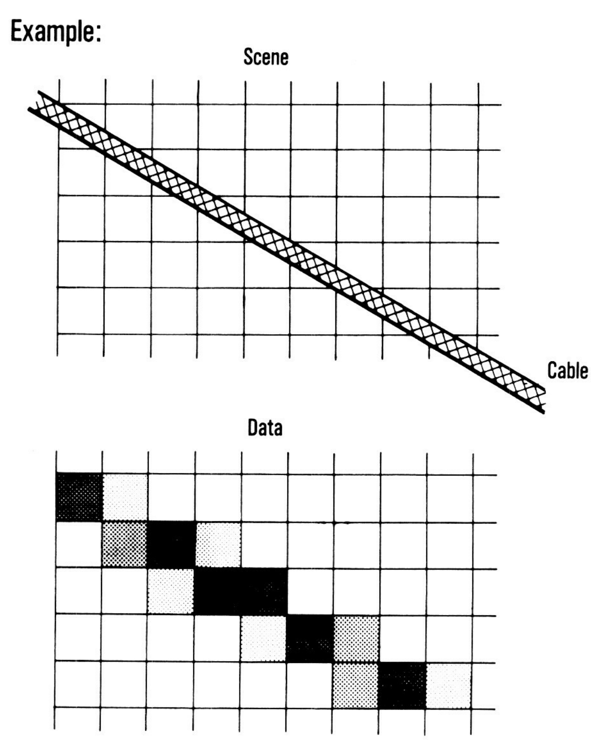

Figure 1b. Data Interpretation (Display)

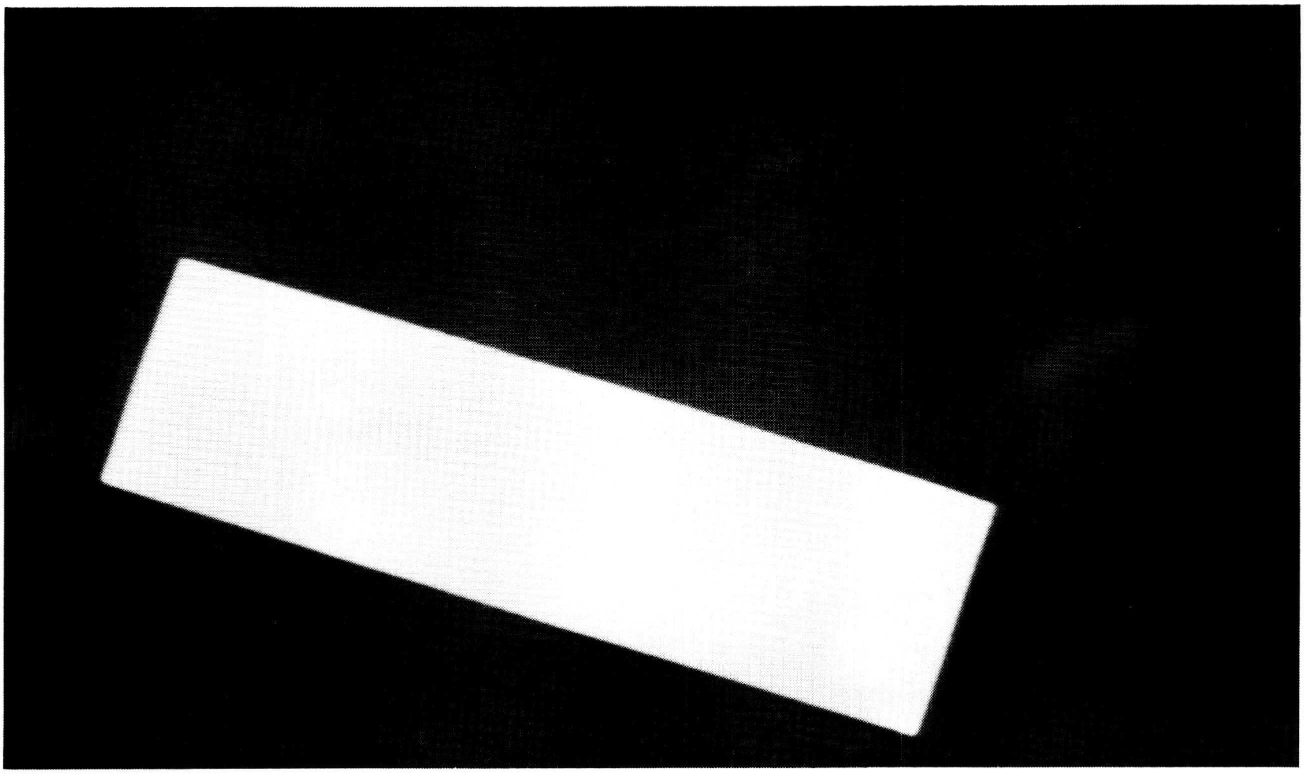

Figure 2. Original TV Image

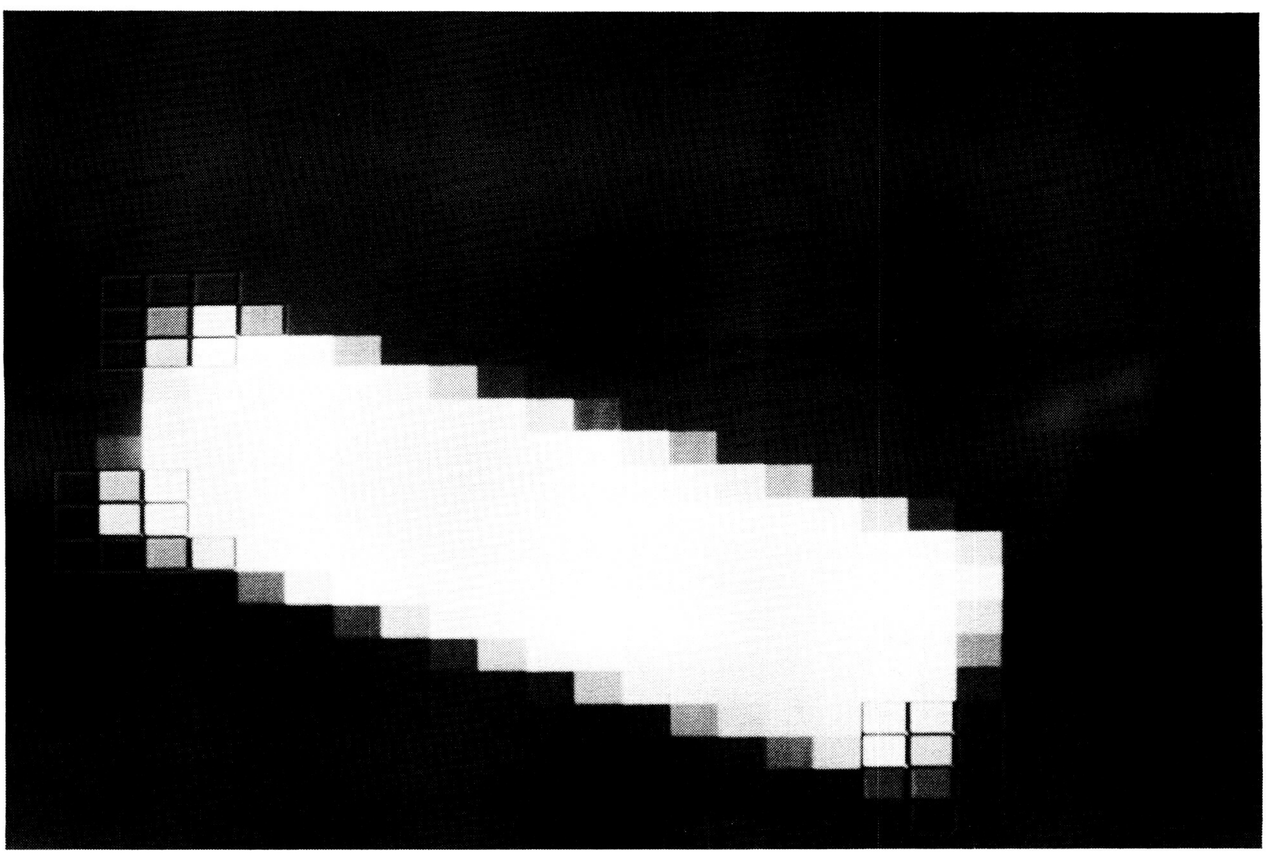

Figure 3. Some Pixels Classified as Corner-Like

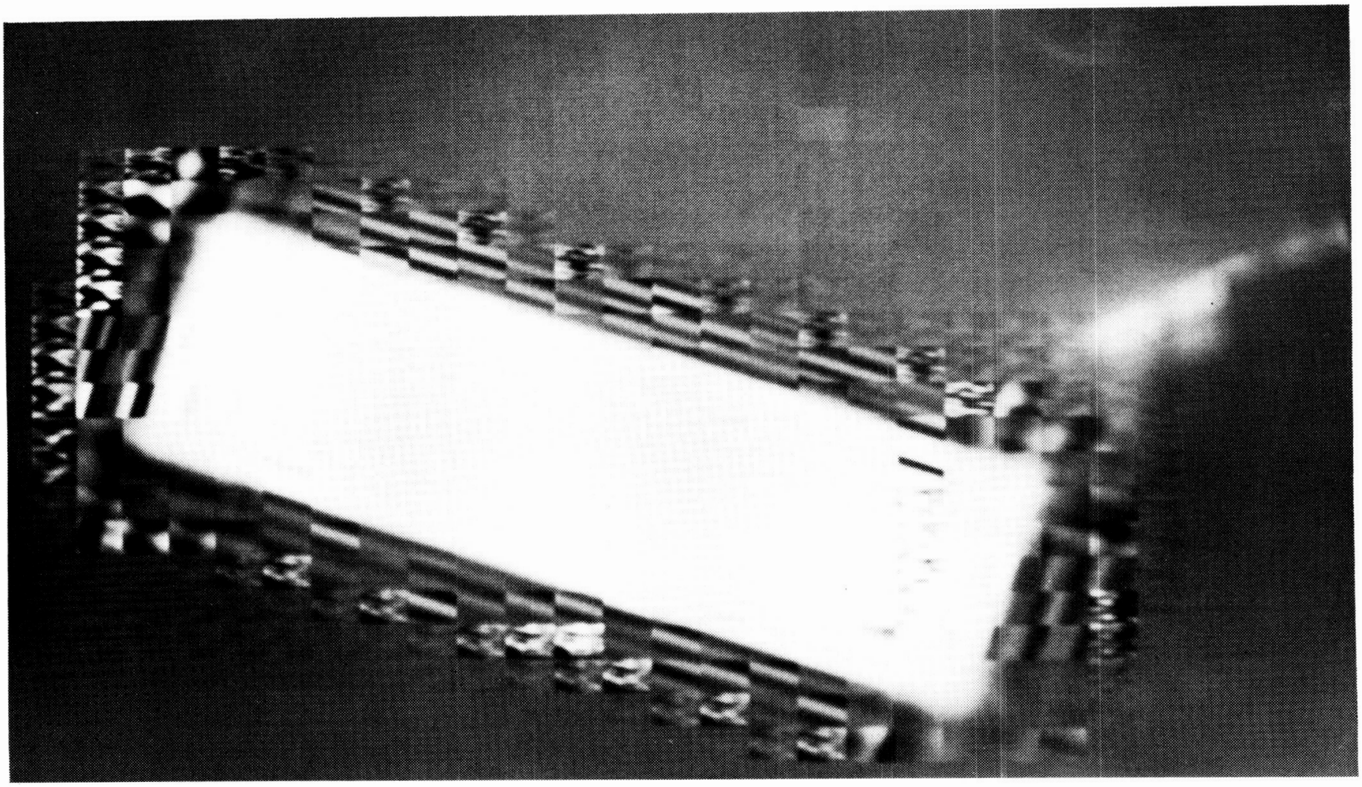

Figure 4. Contexturally Interpreted Coarse Data

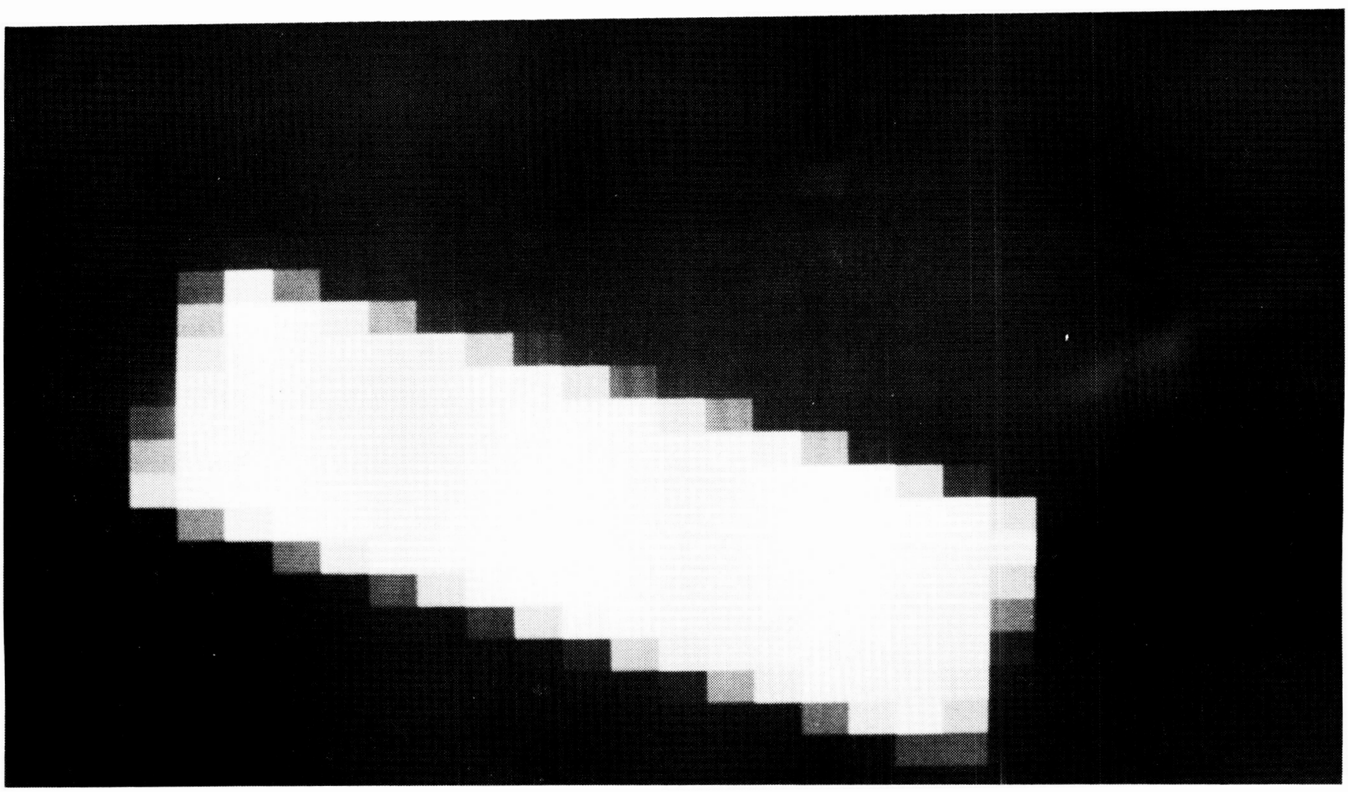

Figure 5. 32 × 32 Coarse Scene of the Rectangle

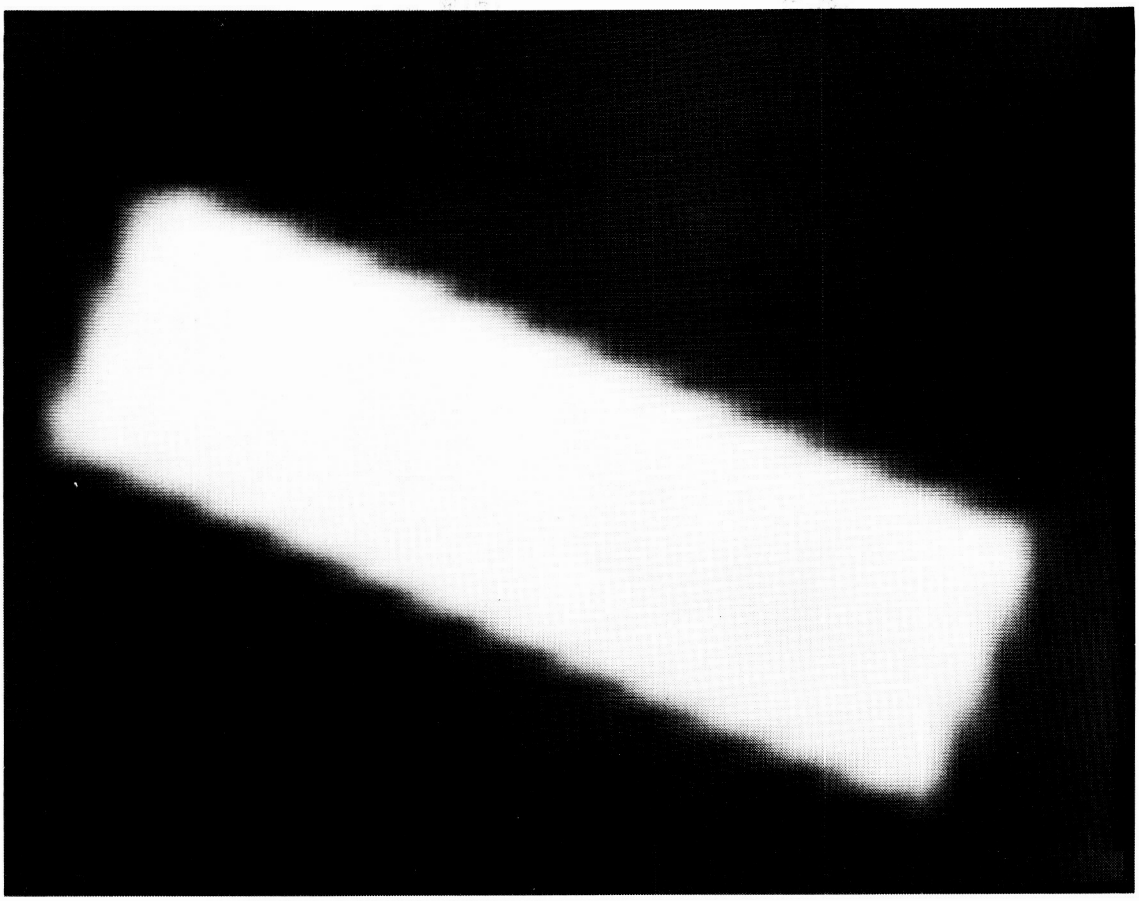

Figure 6. Aliasing Artifacts Caused By Context-Free Bilinear Interpolation

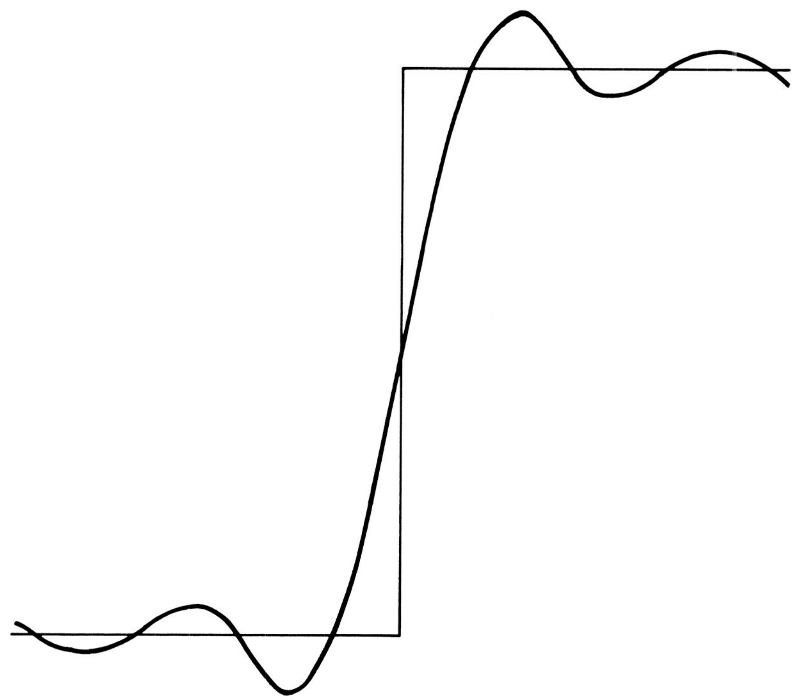

Figure 7. Gibb-Like Oscillations Caused By Least Square Error Criteria and Sharp Truncation in the Fourier Plane

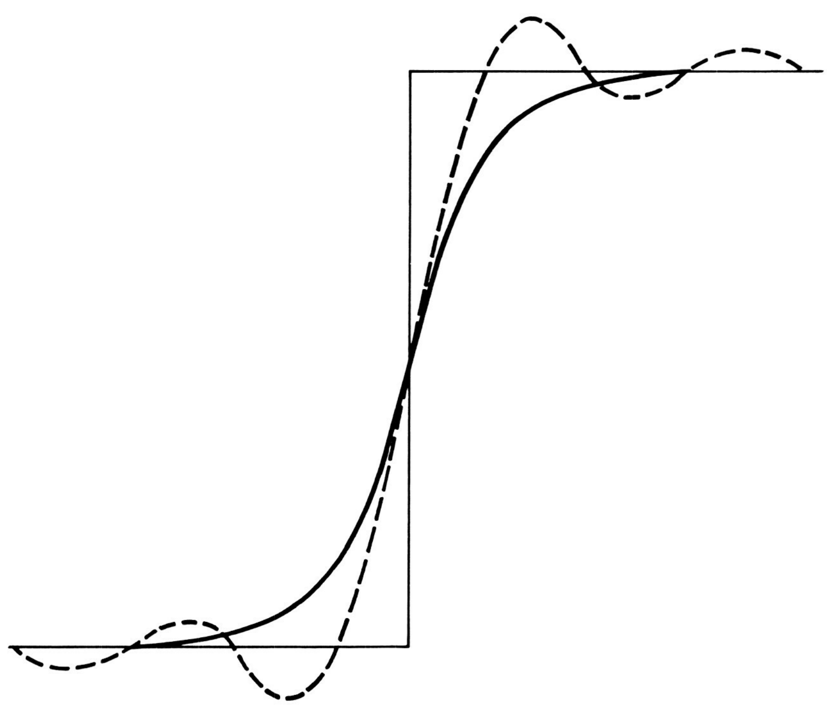

Figure 8. Exponential Edge Interpretation Based Upon a Modified Image Quality Model