MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I.MEMO No. 668 1 June 1982

CARTOON: A BIOLOGICALLY MOTIVATED EDGE DETECTION ALGORITHM

W. Richards, H. K. Nishihara and B. Dawson

Abstract. Caricatures demonstrate that only a few significant "edges" need to be captured to convey the meaning of a complex pattern of image intensities. The most important of these "edges" are image intensity changes arising from surface discontinuities or occluding boundaries. The CARTOON algorithm is an attempt to locate these special intensity changes using a modification of the zero-crossing coincidence scheme suggested by Marr and Hildreth (1980).

Acknowledgement. This report describes research done at the Department of Psychology and the Artificial Intelligence Laboratory of Massachusetts Institute of Technology. Support for this work is provided in part by NSF and AFOSR under a combined grant for studies in Natural Computation, grant 79-23110-MCS. The helpful comments of many members of the Vision Group, especially Ellen Hildreth, were greatly appreciated. A portion of this work was begun with Elmar Noeth in the summer of 1980, particularly the effects of imbalanced masks.

1. Introduction

One of the first tasks faced by any vision processor — whether it be biological or artificial — is to encode the image on its retina into a more economical and meaningful form. Because most of the information in the image is carried by the intensity changes (Attneave, 1954; Barlow, 1961), a large effort over the years has been devoted to creating and describing these changes (see reviews by Rosenfeld and Kak, 1976; Pratt, 1978). Although many researchers have recognized the need for a more symbolic description of the gray levels in an image, Marr (1976) was the first to state clearly the goals of these early stages and to address the problems raised by these goals. Following leads from neurophysiology, Marr (1976, 1982) argues for the construction of a "Primal Sketch" — a primitive but rich description of the image intensity changes which are given labels such as "sharp edge," "shaded edge," "line," "termination" or "blob". Although these symbolic descriptions hint at some external physical cause, they are in fact only image tokens from which the physical features of the scene are deduced (Marr and Nishihara, 1978). Such features include shadows or specularities, surface markings or texture, shapes, contours and the like. One such physical event of particular interest here is the location of surface discontinuities, or where one material abuts another, such as at occluding boundaries, or where grass meets pavement (Rubin and Richards, 1981). The intent of the algorithm is to make strong assertions about the locations of these events in particular.

2.0 The Problem

In an artificial world made of smooth, matte surfaces, whenever two objects made of different materials overlap, there will usually be a step change in intensity in the image. In the natural world, this underlying step change is grossly corrupted by specularities, texture, and shading. Our task is to find the step changes in the presence of these confounding factors.

Figure 1 illustrates the problem in greater detail. The upper graph is an intensity profile taken through a vertical slice of an image (*PYRAMID*). In this profile, there are only seven changes in materials, as noted at the top of the figure. Nowhere is there an ideal "step edge." Furthermore, some intensity changes not associated with material changes are much greater than those arising from the edges of interest. (Compare the sky-to-pyramid transition with some of the texture profiles produced by hieroglyphic markings.)

The complexity of the problem appears to be further increased by the fact that, for biological systems at least, the raw image intensities are not readily available for analysis. Instead, the first stage of processing in neuron-based visual systems is bandpass filtering by so-called "center surround" operators (Kuffler, 1953). Our available input representation is thus not the image profile itself, but rather several filtered versions of the image. Two such filtered versions of the pyramid image profile are shown in the lower panels of Figure 1. In spite of initial appearances of further complexity, this bandpass filtering step will be shown critical to finding the locations of material changes in an image.

Figure 2 gives a clearer picture of the data base from which we start. The lower two panels show the image WINNIE filtered using one of two bandpass filters, in this case the difference of two Gaussians, which closely matches the first stage neural filter used in biological systems (Schade, 1956; Rodieck, 1965; Enroth-Cugell and Robson, 1966; Wilson and Bergen, 1979; Richter and Ullman, 1982). This type of spatial operator or "mask" has the desirable property of preserving both the location and waveform of an intensity change, as seen at the scale of the filter (Marr and Poggio, 1979; Marr and Hildreth, 1980; Sakitt and

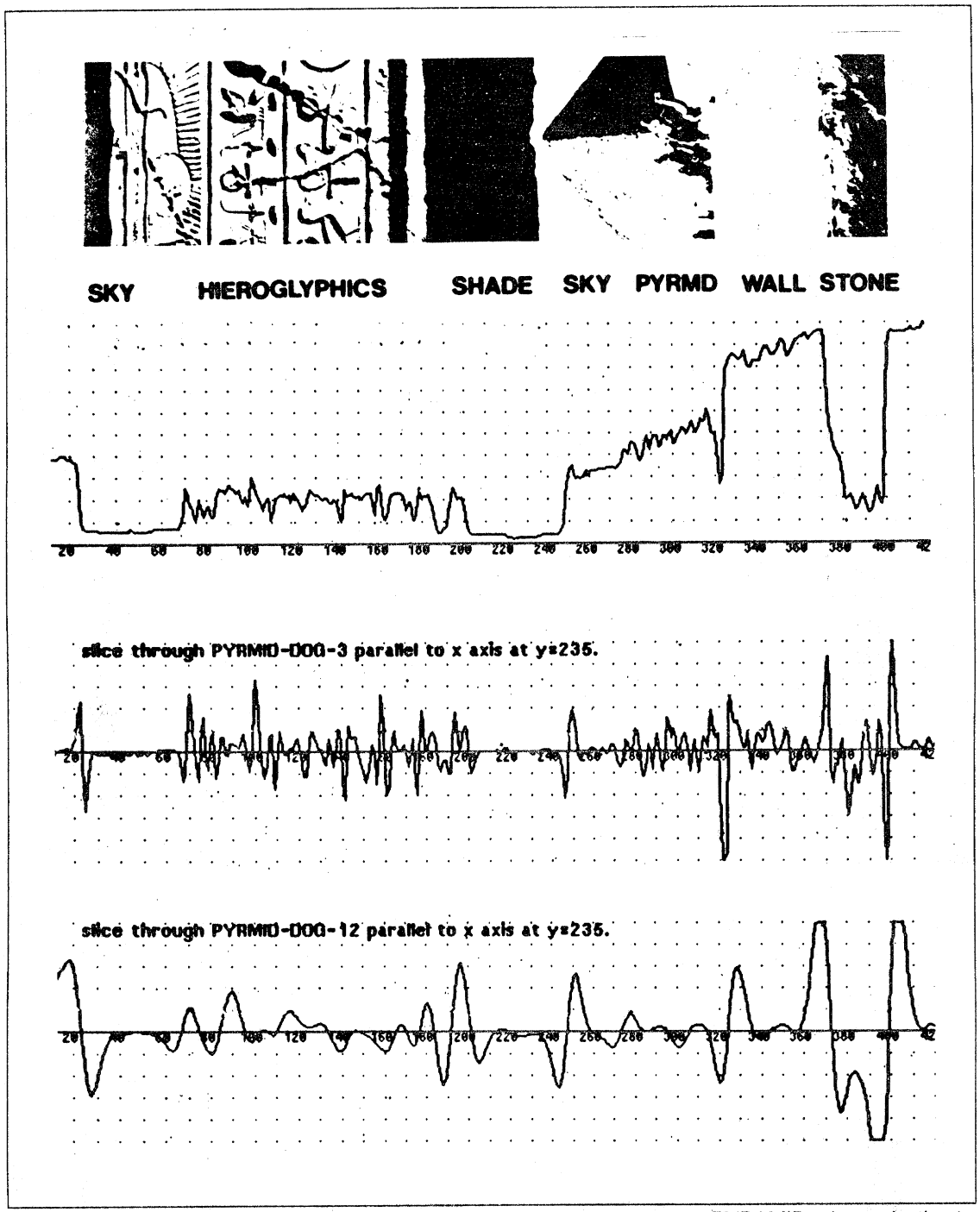

Figure 1 The first graph is the intensity profile of a slice of the image PYRAMID, shown in the top panel. The third and fourth panels show slices of the convolutions of the image, with each slice taken at the same positions as the intensity profile. The two image convolutions were made with a difference-of-Gaussians mask depicted in Figure 2, with mask widths of 3 and 12 pixels.

Barlow, 1982). It has the further property of being ideally suited to detect intensity steps corrupted by noise (Shanmugan et al., 1979; Jernigan and Wardell, 1981).

Because a difference-of-Gaussians filter approximates a second derivative operator, the zeros in the filter's output (i.e., the convolution) correspond to peaks in the intensity changes, which are located at the putative "edges" in the original image. In the above figure, these "edges" or their second-derivative correlates occur at each black to white

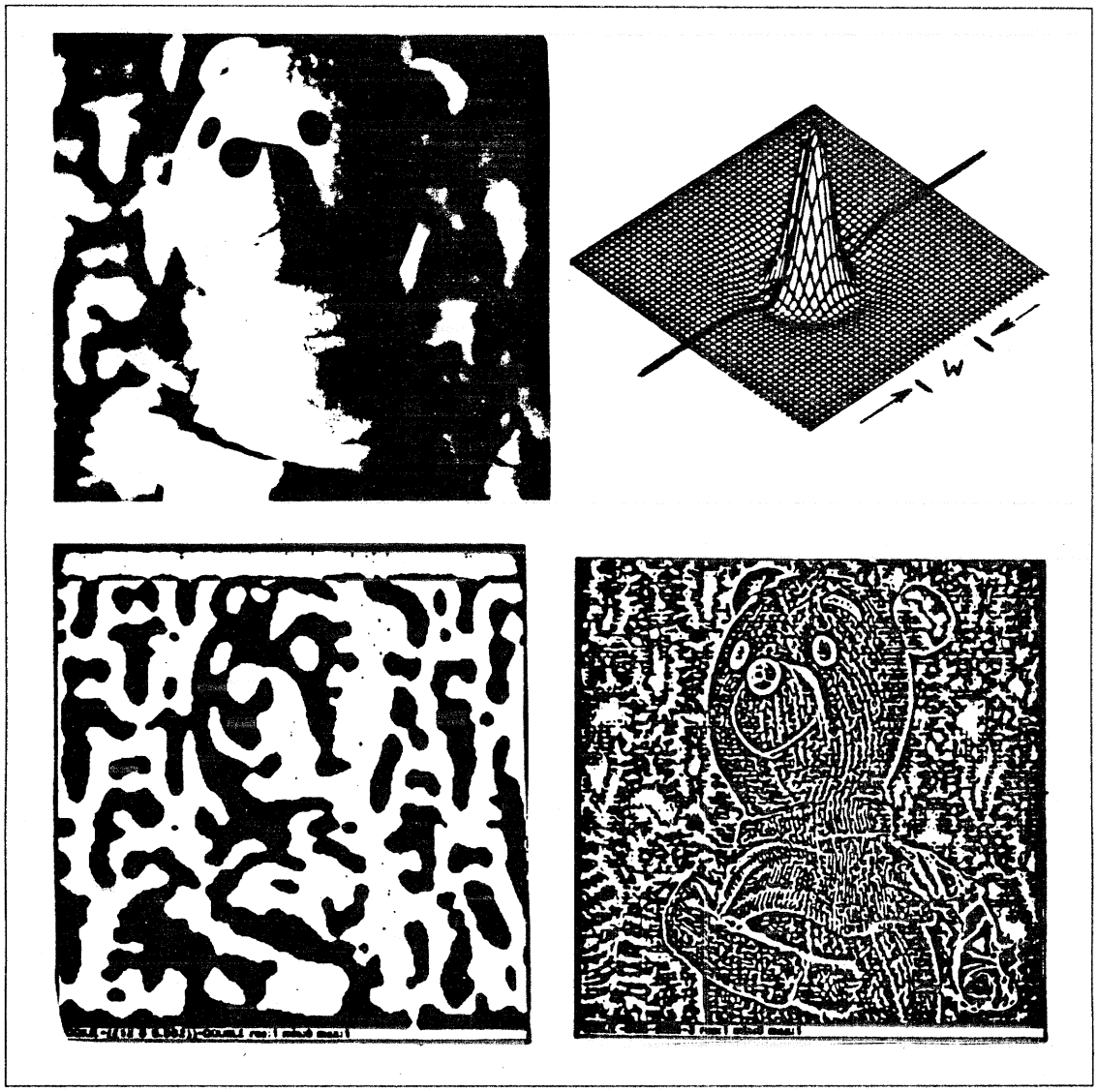

Figure 2 The original 512 x 512 image of *WINNIE* has been bandpass filtered using a difference-of-Gaussians mask of size W = 16 (lower left) and W = 3 (lower right). The mask profile is shown in the upper right panel, where W is defined as the width of the positive part of the mask. Only the sign bits of the convolutions have been displayed.

transition, which is where the sign of the convolution output changes from positive (white) to negative (black). Surprisingly, this representation which shows only the locations of the zero-crossings in the convolution at each scale is remarkably rich (Marr and Poggio, 1979; Marr, Poggio and Ullman, 1979; Nishihara, 1980), and does indeed constitute a simple data base from which the locations of material changes can be deduced with reasonable certainty.

If all material changes produced a sharp step in intensity, then finding these edges from the outputs of the bandpass filters would be straightforward. Because the circular masks are bandpass filters, their convolution with a step intensity profile will be zero when the mask lies exactly centered on the ideal step. This will be true for all mask sizes, as long as the edge intensity profile is straight and longer than the entire extent of the mask (Marr, 1976). Thus, for an ideal step edge, the convolution profiles for all mask sizes will cross the zero axis at exactly the same position, just as it does at the right edge of Figure 1. Certainly such a coincidence in the locations of the zeros in the mask outputs is a

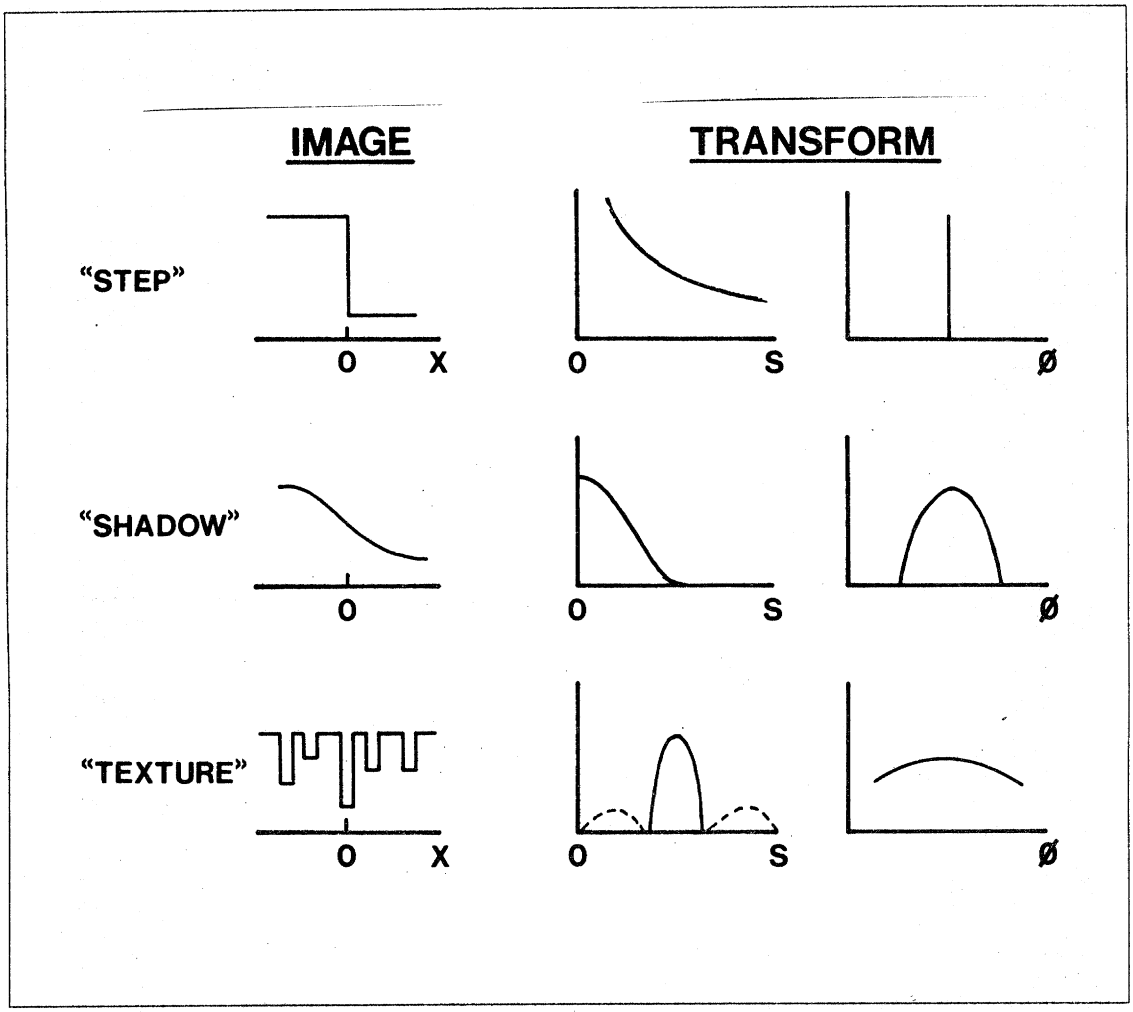

Figure 3 Image profiles of several types of "edges" and their one-dimensional Fourier transforms.

rare event and should be noted as having special significance (Marr and Hildreth, 1980). However, for natural images where ideal step profiles are perturbed and corrupted, such exact coincidences are rare. How then can this coincidence idea be made more robust?

4.0 A Solution

To identify those image intensity contours that arise from material changes we adopt a "reject-accept" strategy suggested by Rubin and Richards (1981). Quite simply, we wish to "reject" clear instances of image profiles that cannot arise from a material change, leaving a much smaller number of candidate profiles for closer examination. Two image characteristics that arise from material changes will be used: the broad spectral power of the resulting step edge and the phase coherence of its Fourier components.

Figure 3 illustrates the approach in more detail. As previously noted, the basic waveform underlying all material changes is the step edge. Regardless of the corruption of this ideal step, such as by shading or texture, the transform must contain power over a wide range of frequencies, S. The addition of texture or shading cannot eliminate this broad-based power of the underlying step, but only alter (broaden) the phase relationships ϕ among the different

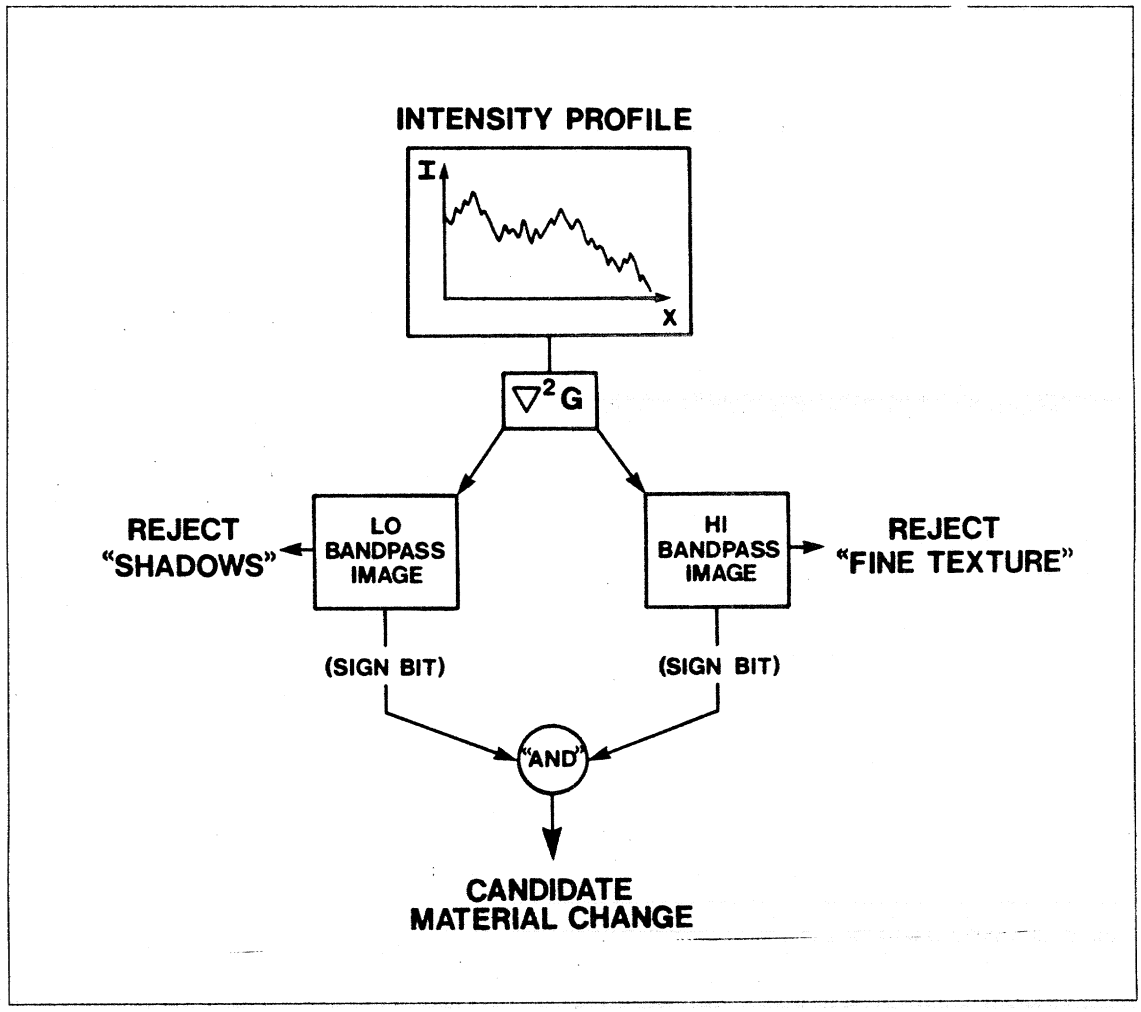

Figure 4 Schematic Flow Diagram of first stages of CARTOON algorithm.

frequency components.

Now consider first the case of texture or a shadow alone (lower two panels). The shadow edge has power only at low frequencies, whereas texture has principally a high frequency spectrum. In general, neither of these intensity profiles alone will cover a broad spectrum. Thus an intensity profile whose power is confined to *only* one portion of the spectrum can be rejected as being caused by a material change.

On the other hand, since a material change produces an underlying step edge with a broad spectrum, any intensity profile that contains both high and low frequency components should be considered as a candidate material change. We thus wish to retain profiles that will survive simultaneous low and high-bandpass filtering, and reject profiles that pass only one or the other filter alone. Our cartoon algorithm accomplishes this by taking the logical "AND" of the outputs of a low and high bandpass filter, as illustrated schematically in Figure 4. The two-dimensional intensity profile is first convolved with two separate difference-of-Gaussians masks, yielding a low and high-bandpassed image. The positive sign bit of this filtered image is then assigned the value "1," otherwise the value will be zero. These binary outputs of the two filters are then multiplied at each point in the array, in effect testing for "coincidences." The resulting product identifies the candidate step edges; isolated "shadow" and "texture" edges are rejected. Not rejected will be approximately 25% of the pixels, because 50% of the pixels in each of the two convolved images will be set to "1" anyway, at least in the case where the filtered images are independent. To reduce these

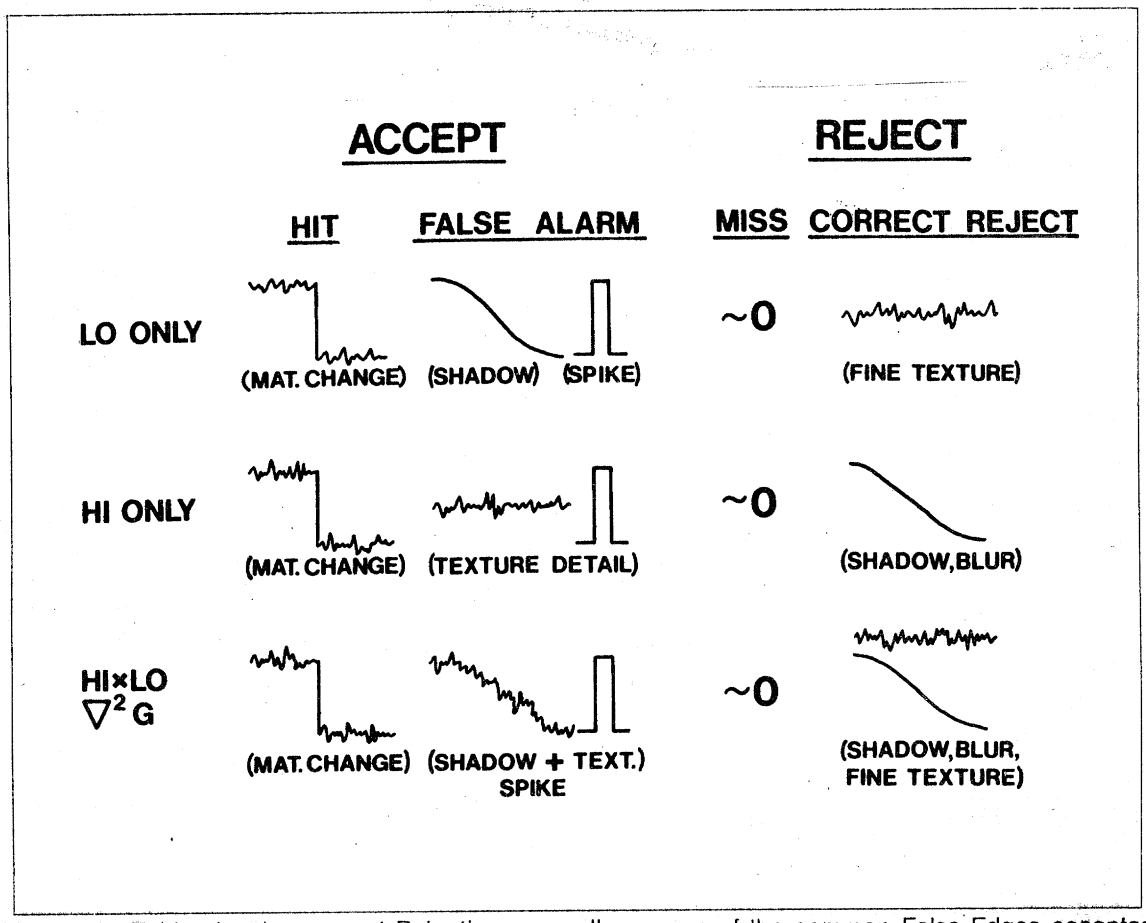

Figure 5 Table showing correct Rejections as well as some of the common False Edges accepted by the first stages of the CARTOON algorithm. The "ACCEPT" and "REJECT" columns correspond respectively to whether the incoming signal passes the CARTOON component shown in the left-most column. By logically combining the outputs from a lo- and hi-pass filter, the rejection rate improves (last column). To improve further the Hit Rates over False Alarms (third column), another stage must be included that incorporates the phase information provided by *POSitive* and *NEGative* masks.

false acceptances, additional independent masks might be used. There are more powerful techniques however, which are presented in the next sections.

4.1 Accepting "False" Edges

One example of a failure in the above procedure is that if a shadow edge is textured, then portions of this profile will pass the "AND"-filtering, yet it is not a material change. Such failures and successes of the basic algorithm can be seen more clearly by examining Figure 5. There are four conditions to be considered: a correct acceptance (*HIT*), an acceptance which is incorrect (*FALSE ALARM*), missing a material change entirely (*MISS*) and finally a correct rejection (*CORRECT REJECTION*). The first two rows of the table depict the profiles that are Accepted or Rejected (correctly or not) by either the lowpass filter or the highpass filter acting alone. For example, a lowpass filter will "accept" either a step edge (*HIT*) or a shadow or spike, which also has a broad spectrum (*FALSE ALARM*). It will miss no step edges and will correctly reject fine texture profiles. The highpass filter has a complementary result, correctly rejecting shadows, but giving false alarms to texture details, as well as to intensity spikes produced by specularities or "cracks." The result of

"ANDING" both filter outputs is shown in the third row. All material changes are correctly identified (as step edges); there are essentially no misses; the rejections are correct, but two serious cases of false acceptances occur: the textured shadow and the intensity "spike." How can these false acceptances be avoided? Alternately, do they occur sufficiently often to cause problems?

4.2 Utility of "Positive" and "Negative" Masks

One of the striking properties of all biological vision systems is that the early neural processing is performed by two complementary mask types (Hartline, 1938; Kuffler, 1953). The profile of one is simply the inverse of the other. Thus, the mask profile depicted in Figure 2 is designated as a "*POSitive*" mask, its complement, the "*NEGative*" mask will simply have the inverted stalactite profile. Of course, for neural systems whose digital signals must of necessity be positive numbers, it is obvious that both mask types are needed to represent the complete waveform of the convolution. However, there is a second, deeper reason for positive and negative masks: their complementarity provides a modicum of additional (phase) information about the intensity profile that reduces the false alarm rate by allowing texture within shading to be rejected or "spikes" to be identified if so desired. This use of phase coincidence is similar to that proposed by Marr and Hildreth (1980) although our scheme is quite different.

To see how a cartoon based on both Positive and Negative masks can reduce the False Alarm rate, consider the intensity profile in Figure 6. (This profile is an idealized version of the wall-stone image profile shown on the right side of Figure 1.) If a suitable large Positive mask is passed over such a textured plateau, the output will be all-positive within the plateau region of the image, whereas the Negative mask output will be all-negative over the same region (Figure 6, second row). Smaller positive and negative masks, on the other hand, will produce convolutions with both positive and negative values over the same region (Figure 6, third row). The CARTOON algorithm then takes these positive sign bits and multiplies them, to yield the pulses shown in the fourth row. Some high-frequency texture detail thus passes the "ANDING" of the two positive masks, leaving a false indication of several changes in material. On the other hand, "ANDING" the two negative masks (right column) rejects this fine texture detail, marking only the desired step edge (left) and the location of the shaded edge (right).

Unfortunately, because of the complementary nature of the positive and negative masks, the sign bits associated with step edges will lie on opposite sides of the zero-crossing as shown in Figure 6, row four (or Figure 7, bottom two rows). This seeming disadvantage is actually the additional phase information we seek, however. Since any true step edge (or material change) must produce adjacent sign bits for both mask types, this phase relation identifies the true step edge profile. Thus, the final stage of the CARTOON simply checks for *pairs* of adjacent sign bits obtained from using both *POSitive* and *NEGative* masks (Figure 6, last row). Our False Alarm rate has thus been significantly reduced for a textured shadow. (Either texture or color or both could be utilized to further eliminate this kind of false target if so desired — see Rubin and Richards, 1981, for a suitable operator.)

The second type of false target is an intensity spike. The fourth column of Figure 7 shows the relation between the positive sign bits of the opposite sign masks. Clearly both the spike and the step edge produce adjacent sign bits from the two complementary masks. However, if the mask size is increased, each sign bit profile will be magnified along the horizontal axis as shown by comparing the crosshatched areas of the profiles shown in the two right-most columns of Figure 7. For the spike, this scaling will cause the locations where the two flanking sign bits abut the central mode to move outward as the mask size

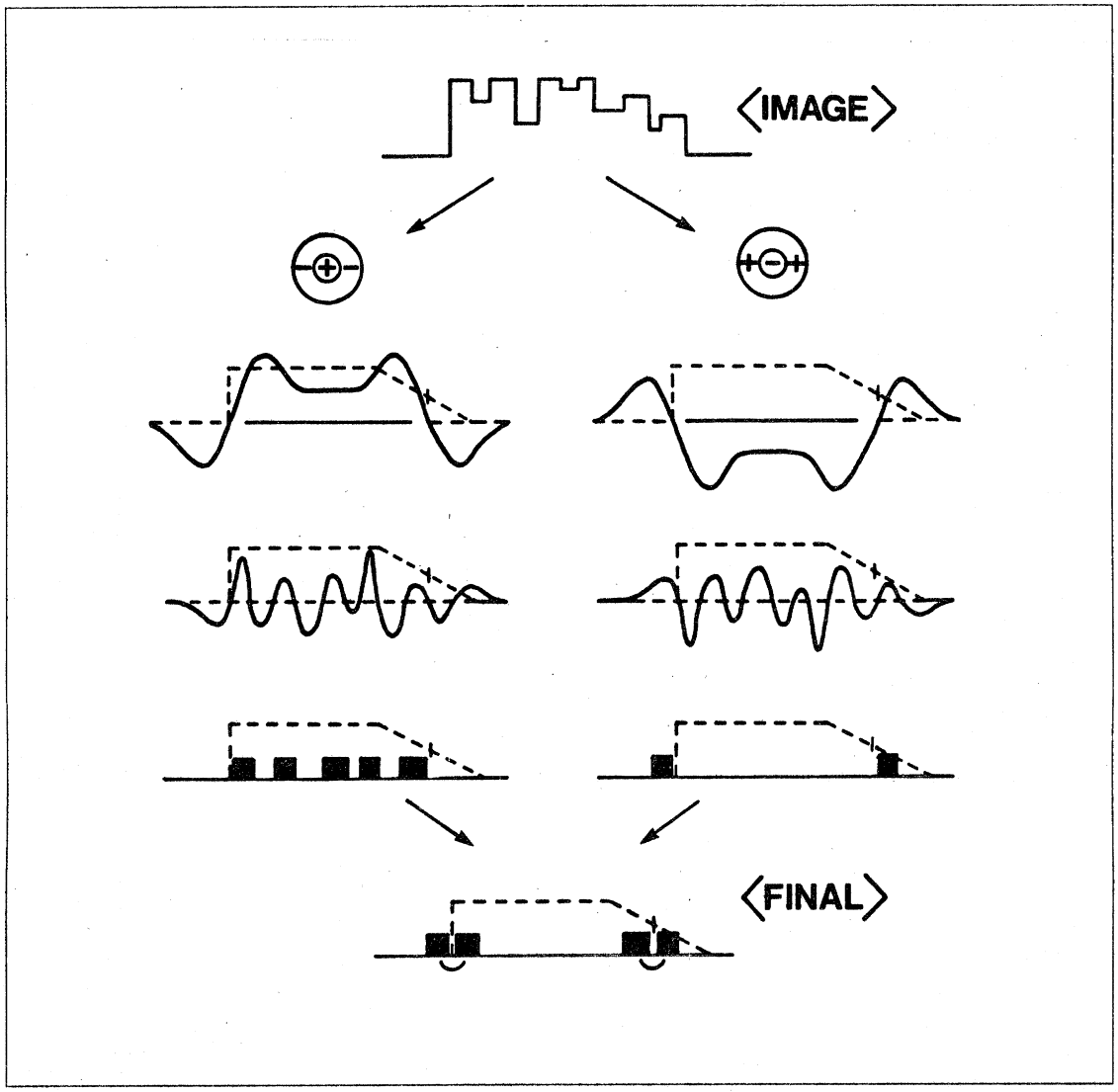

Figure 6 The hypothetical image intensity profile shown at the top center is crudely depicted in the subsequent figures as the dotted line. For appropriately tuned large masks the output of the convolution to this image profile will be either all positive (second row, left) or all negative (second row, right). The small mask convolutions, on the other hand, will have both positive and negative values because of the "texture." "ANDING" the outputs of the sign bits of the two masks will thus yield different patterns of coincidence (third row). By testing for neighboring coincidences in the *POSITIVE* and *NEGATIVE* CARTOON outputs, a truer separation of step edges is obtained (fourth row).

increases. This displacement of the abutting position will cause a gap between the resultant "positive" sign bits obtained after "ANDING" masks of two different sizes as shown by the filled portion of the cross-hatching. For the step edge, however, the position where the sign bits abut will remain fixed at the two scales, and this location is preserved after the "ANDING" operation. The final coincidence check illustrated in the bottom row of Figure 6 will thus preserve the step edge but reject the "spike."

To summarize, the complete CARTOON uses four mask types — a coarse and a fine plus the positive and negative profiles for each. A "positive" cartoon is then created by "ANDING" the outputs of the two positive masks. Similarly, a "negative" cartoon is produced from the two negative masks. The sign bits of the positive and negative cartoons are then circularly smeared at each point by half of the fine mask width, and then multiplied using the logical "AND." The smearing allows the sign bits which are adjacent in both the positive and

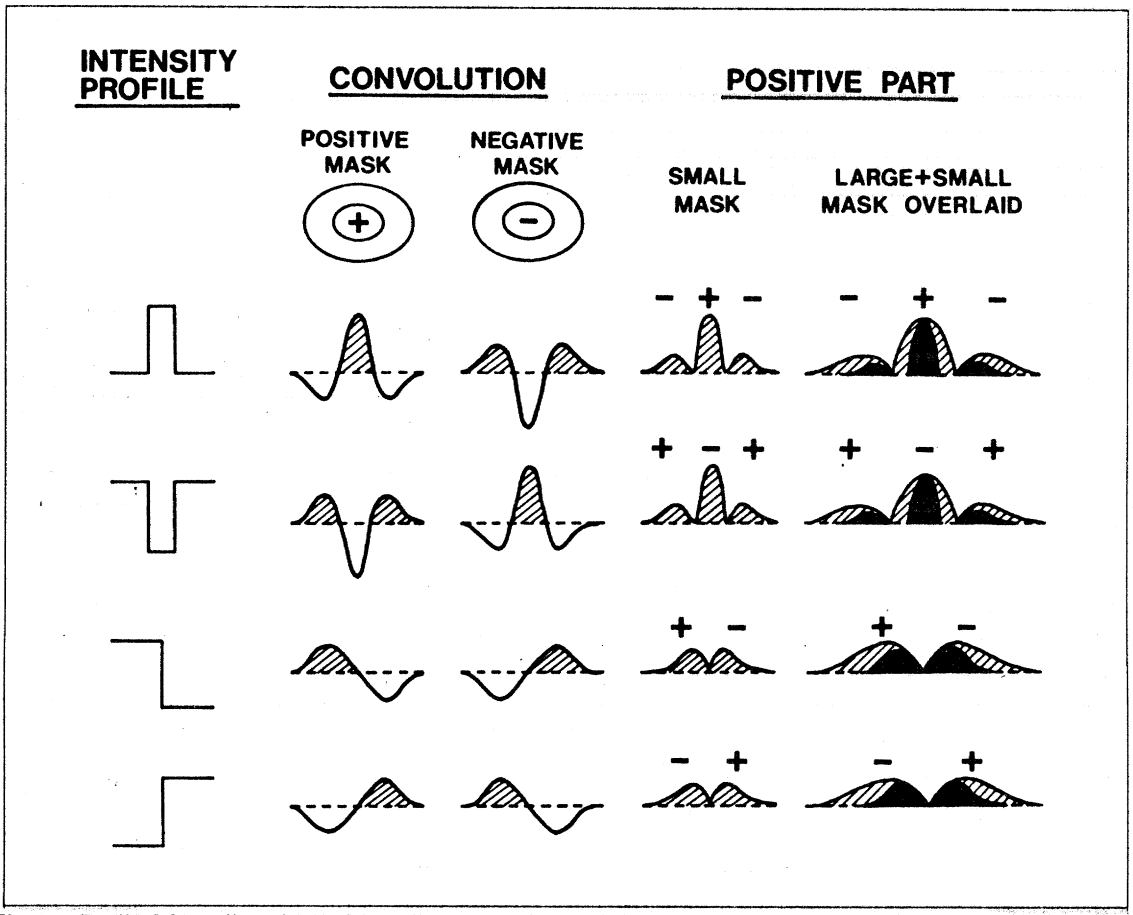

Figure 7 "*POSitive*" and "*NEGative*" masks provide some information about the phase of the Fourier components of an intensity profile. A negative "spike" or "step" can be discriminated from its positive inverse by the complemented position of the sign bits for each mask as shown in columns two and three. The "spike" and "step" waveforms can also be discriminated, in turn, by the union of the outputs of the POSitive (+) and NEGative (−) masks (column four). The effect of mask size on the step-spike discrimination is shown in the final column, which shows the result of "anding" the large and small mask outputs. Note that the combined (+) and (−) mask outputs are not immediately adjacent for spikes, whereas they abut for step edges.

negative images to be retained. Texture detail and clutter will be rejected as illustrated in Figure 6. (A slightly greater rejection rate is possible if the "smearing" operation is applied only to an oriented segment and not to isolated points, but this requires more computation.) The result is the final CARTOON that delivers the location of true step edges (or material changes) with high reliability. Figure 8 shows the results of these operations on the image "Winnie".

4.3 Noise Reduction

Both biological and artificial vision systems suffer from two types of noise. One noise source is at the receptor level, and is due to quantum fluctuations in the input, or simply, to differences in the transfer functions between receptors. Because these noise sources are associated primarily with the transduction stage of processing, they will be grouped together as transducer or external noise.

A second source of noise is internal to the system and arises from receptor instabilities

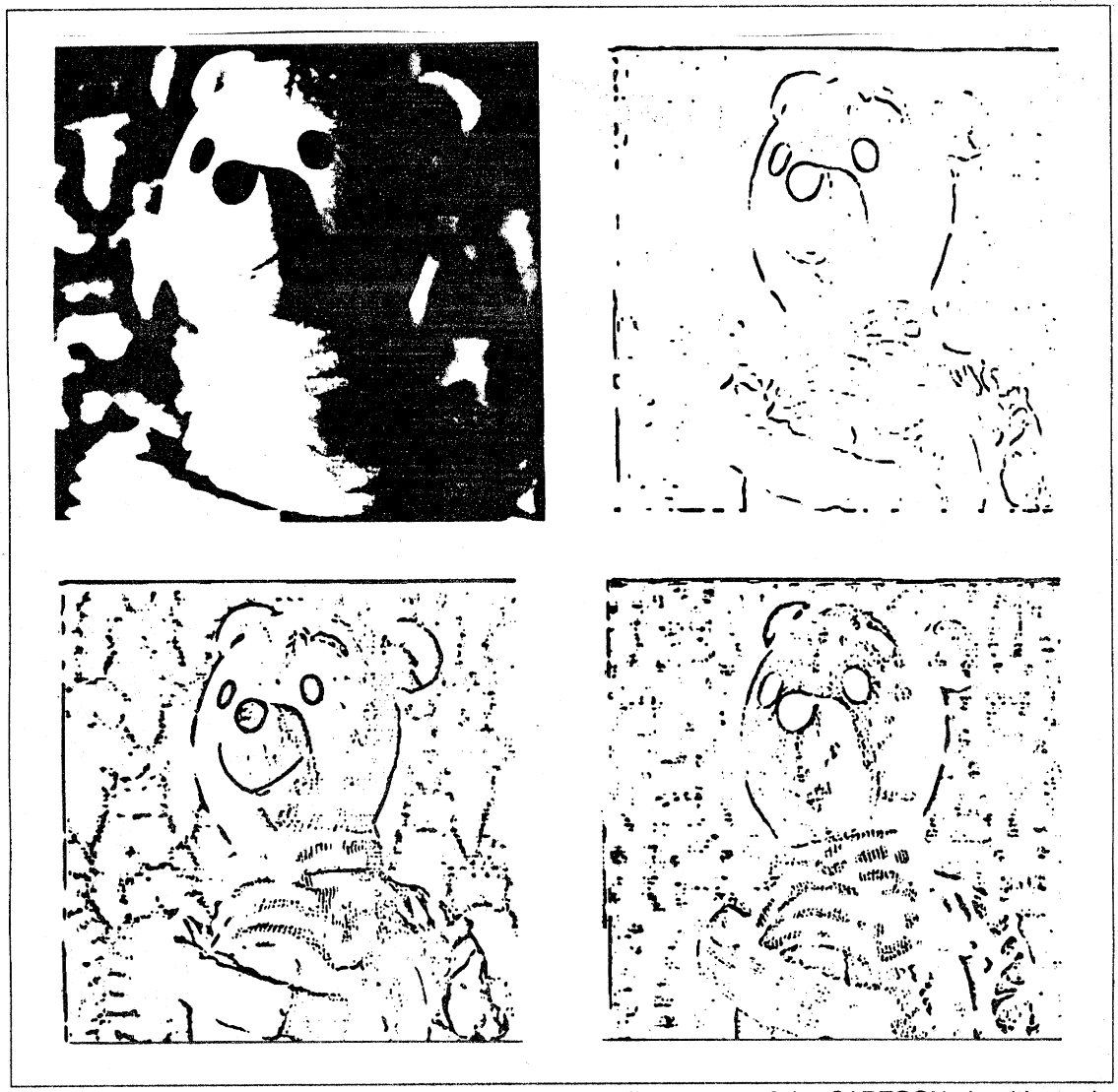

Figure 8 The lower two panels show the output of the first stages of the CARTOON algorithm using either positive (left) or negative (right) mask profiles applied to the original image of WINNIE (upper left). Note that different contours are passed by each operator. After smearing the positive and negative cartoons, the sign bits are multiplied to check for the coincidence of common contours, yielding the final cartoon shown in the upper right panel. (Mask sizes, $W = 3$, $W = 16$; 10% threshold by using an imbalanced mask.)

(usually thermal) or to noisy components (i.e., neurons or computer hardware). For biological systems based on neurons, internal noise can be substantial (see Figure 14). However, if such noise is common to two channels, then it can be eliminated by subtracting the outputs of the two channels, thereby eliminating the common noise component. The fact that biological filters for visual signals consist of two types, whereby one is merely the inverse of the other (i.e., "positive" and "negative" units), suggests that this scheme for reducing some internal noise is plausible. Referring to Figure 1, we see immediately that the positive and negative parts of the convolution of necessity will overlap nowhere. Thus, any signal common to both the positive and negative mask outputs must be internal noise, which can be eliminated by subtracting the positive outputs from each mask type from the other before assigning the sign bit.

Returning to the problem of eliminating the noise associated with transduction, several options are available. The two most obvious are temporal averaging and thresholding. Both

appear to be used by primate visual systems. Our choice is to threshold the output of the filters at 5 to 10%. The rationale for the minimum of 5% is as follows: let Q be the number of quanta received per receptor in a time frame (30 msec). Let N be the total number of receptor elements per filter. The mean signal is thus NQ and since the quantum distribution is Poisson, the fluctuations will be $(NQ)^{\frac{1}{2}}$ (de Vries, 1943; Rose, 1942). As a percentage of the mean, these fluctuations V will be

$$V = 100 \cdot (NQ)^{\frac{1}{2}}/NQ = 100/(NQ)^{\frac{1}{2}}$$

A 4 cycle/deg DOG filter (center width $W = 8'$) will collect about 5000 quanta per 30 msec in the center portion of the mask at a daylight level of $50cd/m^2$, assuming 10% quantum efficiency (Jones, 1957). An estimate for the fluctuations is thus about 2% of the input signal, but could easily reach 5% under twilight conditions where the quantum input is reduced 10-fold. Furthermore, in addition to the quantum noise, there will be additional errors in analog to digital conversion (both for artificial and biological systems), which are in the order of 1/2 to 1%. A threshold of 5% of the maximum signal is therefore a reasonable lower bound. For our particular camera and hardware, we have chosen a threshold of between 5 to 10%, depending upon the light level and contrast of the picture being inspected. (The effects of this threshold on the cartoon will be illustrated later.)

4.4 The Thresholding Method

For machine vision, the simplest thresholding technique is simply to set to zero all values of the filter that are less than 5% of the maximum output. This scheme is easy to implement, because the output range of the camera and the subsequent convolution is known. For biological systems, however, the voltage range of the receptors depends upon the light level, whereas the neural signal range is fixed. Thus, although 5% of the receptor output is a trivial analog signal at low light levels, when the incident intensity is raised 1000 fold or more, this small percent of the higher output would greatly exceed the input range for neural signals. In order that the thresholding be independent of the actual signal input, a possible scheme would be to bias the weights of the center and surround components of the filter such that a steady signal over the entire field actually caused a 5% inhibition. (Thus, the smaller Gaussian component would have only 95% of the volume of a larger Gaussian counterpart.) For our implementation, we mimicked this neural scheme because it has the further advantage of automatically producing a 5 to 10% threshold without the need to normalize for the intensity range.

5.0 Scale

Textural shading, "cracks," specularities, and shadows depend upon the scale of the viewer, as does what constitutes a material change. From the eye of an ant, the 5 mm "crack" is a haven of safety and each blade of grass constitutes a separate occluding object. But from the point of view of man, the grass is merely one carpet of the same material, and the ant's haven is just a crack. The scale of the chosen filters thus depends upon the viewer and the size of his "world."

5.1. The Low-Bandpass Scale

From the human point of view, wrinkles or cracks in a surface are an important aspect of texture. Intensity changes of such origin are to be rejected by the algorithm. "Cracks" are usually less than 5 mm wide, and are of sufficient depth to create a high contrast pulse in the intensity profile. As seen at a height of 60 cm, a "crack" in the floor, for example, will thus be less than 1/2 deg in width. Since more distant cracks will appear still smaller, negative high contrast ($> 50\%$) pulses of intensity that subtend less than 1/2 deg visual angle are likely candidates for "cracks."

Similarly, specularities at a sharp edge represent intensity spikes that also are to be rejected. Most smooth surfaces with sharp edges that generate specular spikes in intensity have a radius of curvature less than 5 mm. The visual angle of specularities reflected off such a surface will thus also be less than 1/2 deg.

To reject fine texture "cracks" and specularities that subtend 1/2 deg or less, the low-pass filter must be located below 2 cycles/deg. For a bandpass filter of 2-3 octave width similar to that used by the human visual system (Spitzberg and Richards, 1975), the center frequency should thus be slightly below 1 cycle/deg for the reject strategy to succeed. This choice is consistent with the locations of one of the four spatial-filtering channels based on psychophysical measurements (Richards and Polit, 1974; Wilson and Bergen, 1979; Richards, 1980).

5.2 The High-Bandpass Scale

The task of the high-pass filter is to reject the shallow intensity gradients due to shadows, highlights and changes in surface orientation. If, at the same time we desire this output to indicate "cracks" and specularities, then the location of this filter must be at about 4 to 5 cycle/deg if it is roughly 2-3 octave bandpass. Such a filter will also reject all intensity gradients below 2 cycle/deg, to varying extents depending upon the nature of the low frequency fall-off. (A difference-of-Gaussians filter is still at one-quarter its peak value at two-octaves below the peak frequency.)

For shadows, there is a wide range of possible edge profiles from very shallow to very sharp. As a crude estimate of the scale, consider a shadow cast from above, as from the branch of a tree. Then, since the lowest branch normally encountered will be at eye level, the penumbra cast by the 1/2° overhead sun onto the ground will be in the 1/2° range (if the occluding edge is sharp). Branches that are higher up, or objects that are farther away, or sun angles other than high noon, will produce shallower penumbrae. Most shadows seen on the nearby ground, therefore, will have penumbra that are larger than one degree. Shadows seen at greater distances, of course, will be proportionately "sharper," but also are less likely to be seen because of low contrast or because they are occluded by surface undulations. If the 1/2° penumbra is thus taken as a rough lower bound in the width of the intensity ramp, then this estimate can be combined with the maximum contrast step expected for a shadow. For overhead sun with either blue sky (complete clouds will, of course, produce little shadow), the contrast will be less than 20% (Richards, 1982). A 2 octave bandpass channel located at 5 cycle/deg will miss this intensity profile, and all others that are shallower still, such as shading on most hand-held objects.

5.3 Summary of Filter Choices

In summary, to reject fine texture, cracks and specularities, a low-bandpass filter should be located at less than 1 cycle/deg, which corresponds to a difference-of-Gaussian filter having a space constant of about .5 degrees. Similarly, to reject the shallow gradients due to highlights, shading and shadows, the high band-pass filter should be located near 5 cycle/deg, which corresponds to a difference-of-Gaussians filter with a space constant of about .1 degrees. These choices are remarkably consistent with the locations of the foveal spatial frequency channels found in man using psychophysical methods (Richards and Polit, 1974; Wilson and Bergen, 1979) and suggest that the proper size ratio between the large and small masks is about 5 to 1. (See also Sakitt and Barlow, 1982, for a theoretical discussion of these ratios.)

6.0 Summary of Algorithm

Figures 9 and 10 schematize our final "CARTOON" algorithm used to encode the locations of occluding edges or material changes. The output from a 1000×1000 CCD camera is fed into convolution hardware that performs a 2D difference-of-Gaussians filtering on the image at 3 and 16 pixel center width (Nishihara and Larson, 1981). (The exact choices are not too critical because the angular size of our pictures was arbitrary.) Both "*POSitive*" and "*NEGative*" masks were used at each of the two scales. To remove camera noise, these filters or "masks" were thresholded by $\theta = 5\%$ imbalance (in some cases a 0 or 10% imbalance was used as indicated). The positive sign bit of the four convolutions was then retained and the binary images obtained from the same signed mask were multiplied by "anding" to yield a "*POSITIVE*" and "*NEGATIVE*" cartoon. These two images were then smeared by ± 2 pixel elements (one half the center width of the highpass mask) and multiplied again to produce the final cartoon.

In some cases, it may be desirable to restore more of the original image, in which case a minimal amount of information about gray level can be superimposed on the final CARTOON (Figure 11). It is clear that for most biological systems some feeble D.C. signals survive bandpass filtering, as evidenced by after-images or low-frequency sensitivity for example (Brindley, 1970; Stromeyer et al., 1982) or the pupillary response (Alpern et al, 1963). When gray level information was desired, the image intensities were divided into one of 4 equally spaced intensity levels and then this 2 bit image was low-pass filtered with a 24 pixel wide mask. The cartoon contours were then superimposed upon this "blurred" 2 bit image to create the final picture. This scheme reduced the original 8×10^6 bits to an average of about 10^5 per picture.[1]

[1]Schreiber and Troxel's (1981) two-channel picture coding system uses a related procedure to hide visible noise encountered during bandwidth compression.

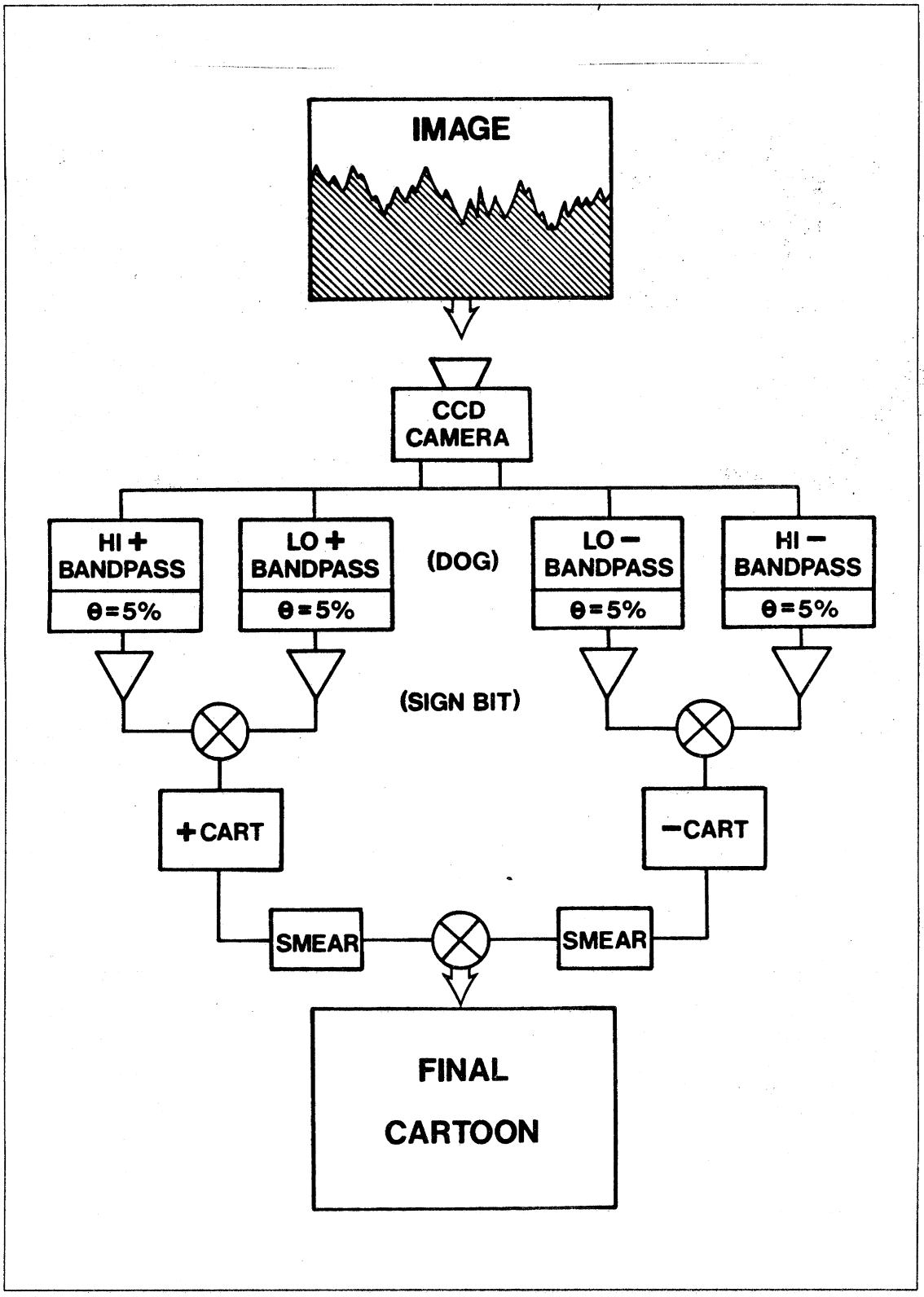

Figure 9 Final CARTOON Algorithm. (See text for details.)

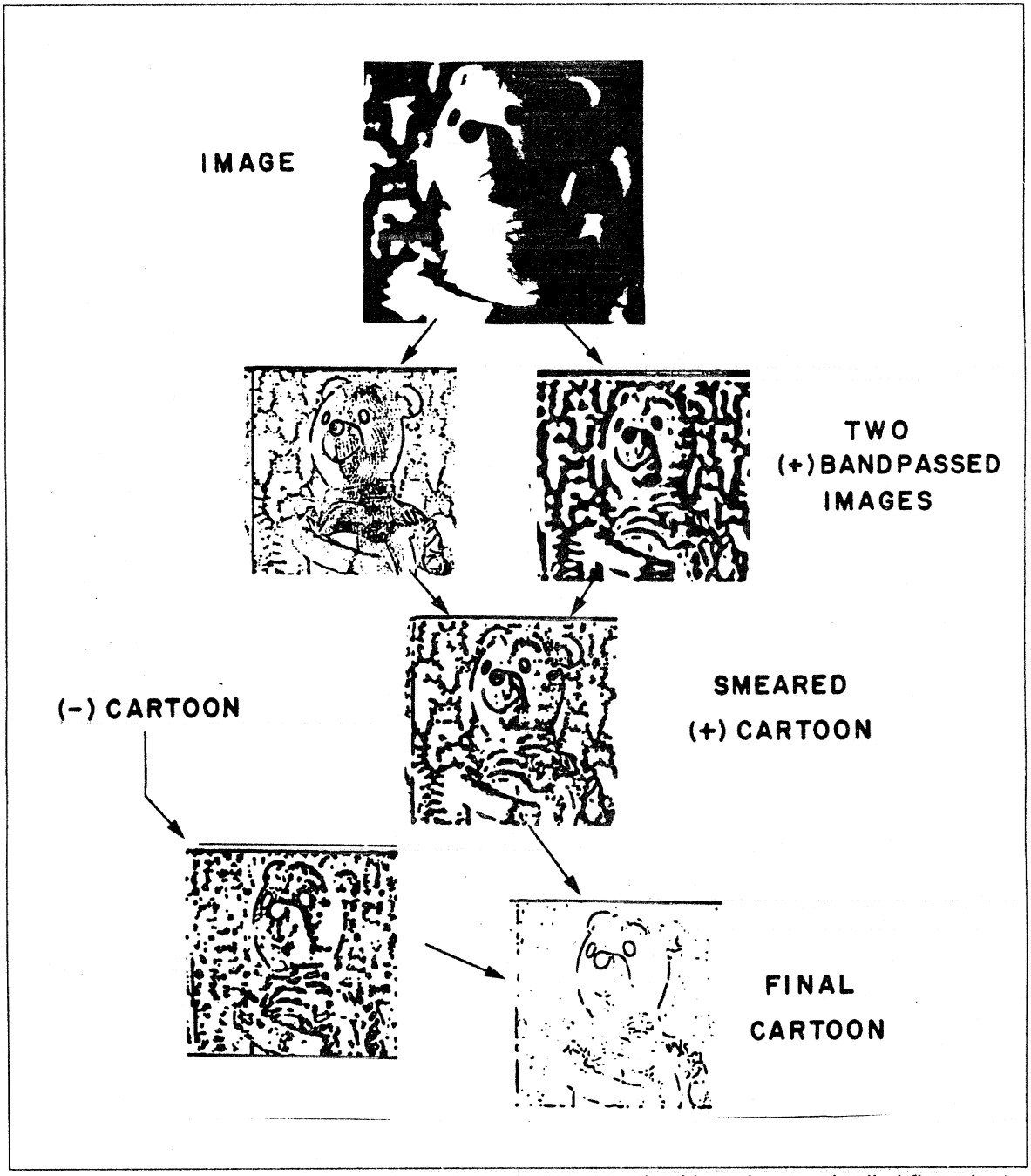

Figure 10 Pictorial representation of stages for the cartoon algorithm. A more detailed flow chart appears in the previous figure.

7.0 Trying it out

7.1 "Winnie"

The result of applying the full CARTOON algorithm to the image "WINNIE" has already been shown in Figure 8. Of some interest is how this final image will change as the threshold imbalancing of the masks is altered. The left column of Figure 12 shows the positive mask output for a 0%, 5% and 15% threshold created by imbalancing the center and surround

Figure 11 Demonstration of the value of retaining some gray level information, and how contours mask the blur of low-pass filtering for achromatic signals, just as they do for chromatic signals. The lower right panel is a reconstruction of the upper left panel obtained by adding the positive cartoon shown on the lower left to the blurred 2-bit gray level image shown in the upper right. (Reprinted through the courtesy of Whole Life Times.)

portions of the mask. (The 10% threshold condition appears in Figure 8.) The middle column shows the resultant postive cartoon, whereas the right column is the final cartoon obtained by combining the positive and negative stages of the cartoon. By comparison with Figure 8, which is based upon a 10% threshold, we see that the major effect of thresholding is to eliminate the background clutter, with the most significant improvement coming from the 0 to 5% threshold change. Although increasing the threshold does further reduce the unwanted lines, the next major step in clutter removal comes from the combining of the positive and negative cartoons. This step is thus a significant noise reduction operation, which could be further improved if oriented smearing was used. Note that in all cases with the exception of 0% threshold, the major outlines of WINNIE are roughly the same, although some loss of detail occurs around the mouth for the higher thresholds. Based upon such

Figure 12 Effects of threshold at 0% (upper row), 5% (middle row) and 15% (lowest row) on the CARTOON product. The left column is the postive mask image; the middle column is the positive cartoon, the right column the final cartoon. The 10% threshold appears in Figure 8. Note that superficially either increasing the threshold imbalance (left column) or combining the positive and negative mask outputs (right column) causes a similar reduction in contour. (Compare lower left with middle panel of right column.) Upon closer inspection, however, the remaining contours are seen to be of different origin, as expected.

comparisons with several images, we conclude that the product of the cartoon algorithm is not very sensitive to the exact choice of mask imbalance provided it at least equals 5%.

Figure 13 The cartoon of a market scene is shown in the upper right quadrant. The lower two panels are the positive (left) and negative (right) cartoons. Mask widths are 3 and 16 with a 10% imbalance. (Reprinted through the courtesy of American Airlines and George Olson.)

7.2 Market

One of the most difficult images in our library posed to the CARTOON is the market scene shown in the upper left quadrant of Figure 13. The final cartoon (upper right) leaves only a complex splattering of sparse contours that, unlike the face of "WINNIE", are largely uninterpretable. Nevertheless, certain distinctive features of the original image are clearly highlighted in the cartoon, namely the melon in the lower center, the oblique edge of the crates, and the panels at the upper right. These seem to be the features that first attract the eye when the scene is first inspected.

The lower two panels show the earlier stage positive (left) and negative (right) cartoons. Note that the texture detail preserved in one image is often absent in its complement. (Compare the melons and the texture on the wall in both panels, or the lettering on the boxes.) In some cases, therefore, it may be advantageous to stop the cartoon at either the positive or negative stage — but which and how to determine when? Certainly scale factors play a significant role here, for if only the region of the melons, or only the shopper were inspected at increased resolution, the caricature would be clearer. CARTOON is not presented as a conclusive algorithm for isolating all material changes, since it is scale dependent and does not yet utilize spectral or textural information. Rather it is a biologically motivated procedure that is generally quite successful at noting the significant material changes present in the image at the scale of the masks used.

8.0 Relation to Neurophysiology

Since the work of Hubel and Wiesel in 1962, it has been known that biological vision systems compress image information presented on their retinae by first noting the location of intensity changes, and then grouping these locations into oriented segments. The first stage of processing occurs in the retina itself using circular masks (or units) of the type shown in Figure 2 (Kuffler, 1953). A subsequent (cortical) stage then combines the outputs of several such neural units aligned along various common orientations, presumably to make explicit information about the contours of the image. At this cortical level of processing, the neural response will be a rather abstracted version of the original image, as shown in Figure 14, lower left. (This figure shows a plot obtained by Creutzfeld and Nothdurft, 1978, of the activity of four cortical cells to various portions of the Bullfinch image shown in the upper left panel.)

For comparison, the negative (smeared) cartoon is shown at the lower right, while the upper right panel is the final cartoon. The cortical response and the cartoon images seem quite similar.

Presumably other models might be invoked to simulate such cortical data more directly than the CARTOON algorithm, which requires several intermediate steps. For example, the first stage of the CARTOON is the filtering of the image with bandpass masks of at least two different scales. How does the output of these masks then compare with neuronal units at a comparable stage of processing?

Fortunately, Creutzfeldt and Nothdurft (1978) also obtained geniculate responses to the image BULLFINCH, whose neuronal properties resemble the circularly symmetric masks used in the CARTOON. Figure 15 makes the comparison between the behavior of these neural "masks" (left) and the output of the cartoon filters (right). Once again, the initial resemblance between the biological and artificial vision systems responses is striking. Two further comparisons can be quickly made from this figure. First, the upper two rows correspond to masks of different sign (POSitive or NEGative), showing that both types of mask do not always capture identical aspects of the image. (This is due in part to the image intensity profiles and in part to the imbalancing of the masks for noise reduction.) The second comparison that can be readily made is the effect of scale. The panels in the lower two rows are both generated using NEGative masks, but at different scales (Middle, $W = 3$, Lower, $W = 16$).

For all the first stage panels shown in Figure 15, there is considerable noise and clutter which does not appear in the original image. Yet by the time these images have been brought together in the cortex, the debris and clutter has been eliminated (Figure 14). The CARTOON algorithm accomplishes this noise reduction without any arbitrary thresholding, but simply by corroboration of data points at the various locations in the filtered image. It

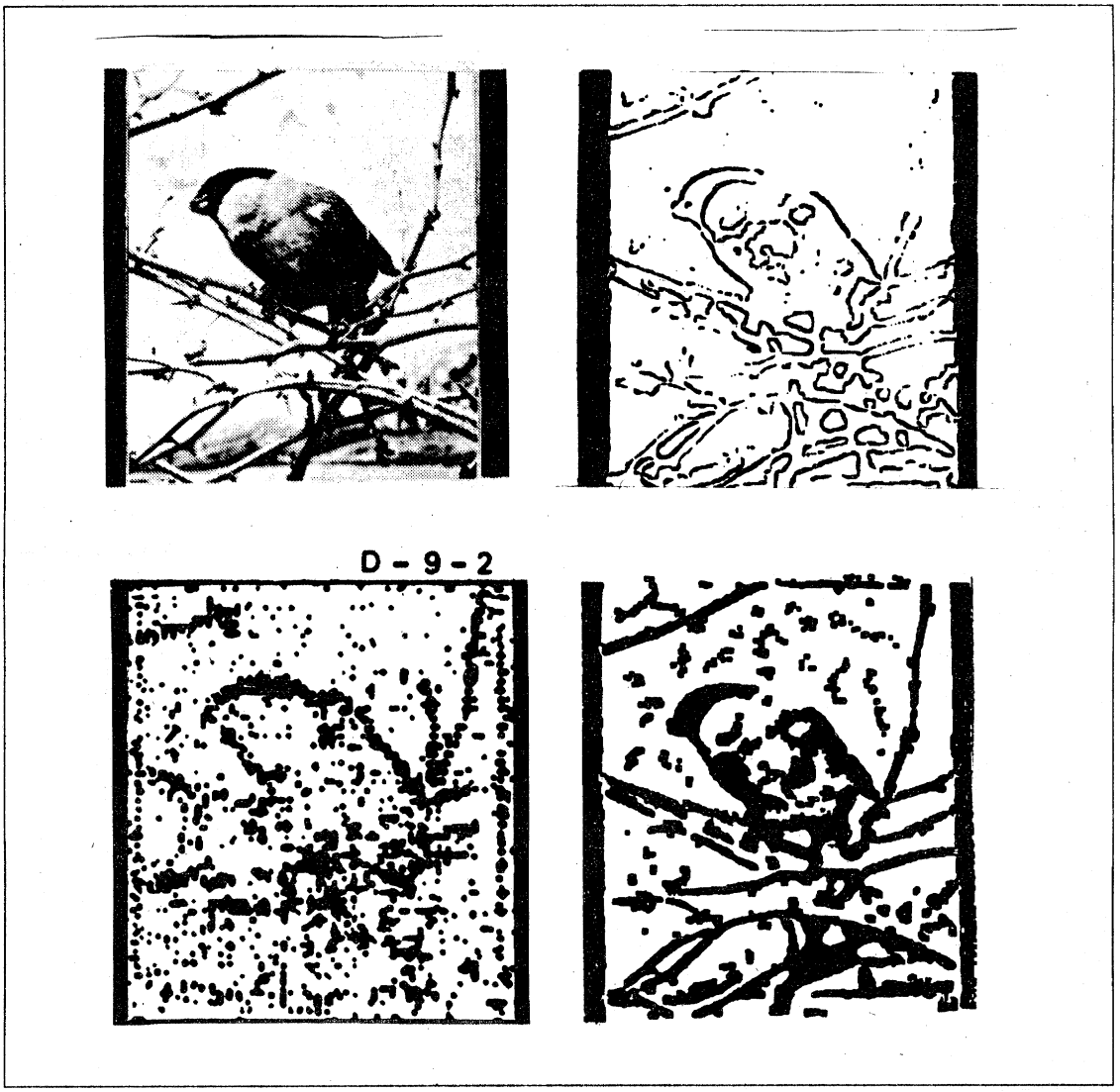

Figure 14 Bullfinch on Twigs Original image, upper left. The positive (smeared) cartoon is shown at the lower right, to be compared with the cortical cell's response depicted at the lower left. The upper right panel is the final cartoon.

is tempting to speculate that the neural processing proceeds along similar lines, probably adding the additional constraint that any near-coincidence of mask outputs must satisfy a smooth contour condition that would require orientation encoding (Marr and Hildreth, 1980). If so, then the orientation constraint should be imposed only by the higher-frequency masks. The cortical organization recently reported by Hubel and Livingstone (1981) is attractively compatible with such a modified CARTOON algorithm.

9.0 Beyond the CARTOON

The CARTOON algorithm is simply a procedure for making strong assertions about the most probable locations of points on occluding contours, where one material changes to another. It does not make these contours explicit, but is only a precursor to actually identifying these contours. Marr (1976, 1982) makes this distinction quite clear when

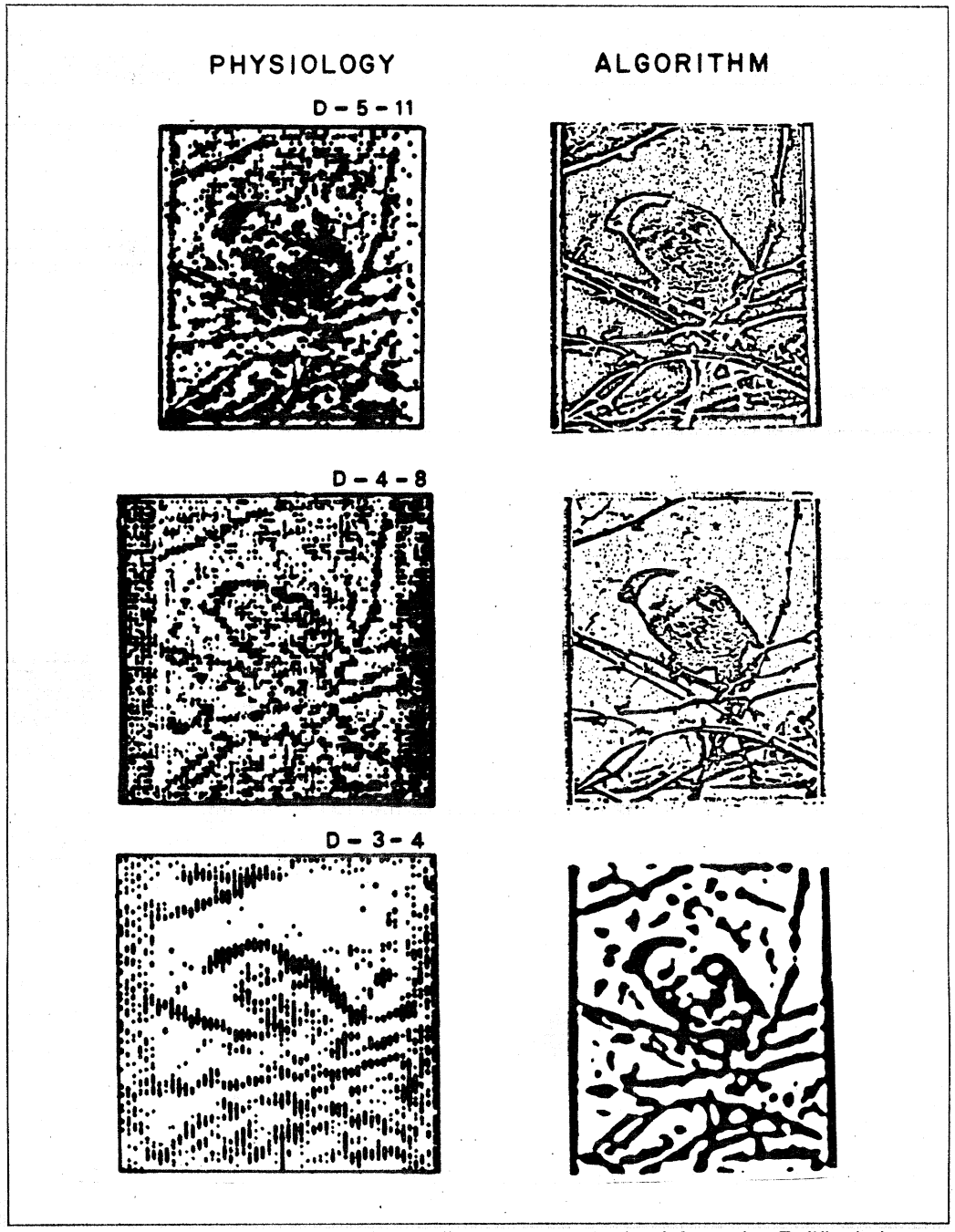

Figure 15 The left panels are geniculate cell responses obtained from the Bullfinch Image. The right panels are the mask outputs that generated the CARTOON of Figure 14. A comparison of the top two rows shows the effect of inverting the mask type. Scale effects are illustrated by comparing the bottom two rows. TOP: Negative mask (or cell), $W = 3$. MIDDLE: Positive mask (or cell), $W = 3$. BOTTOM: Positive mask (or cell), $W = 16$.

examining how zero crossing positions can be joined to form zero crossing segments — the precursors of the contour representation. Similar strategies must be invoked to take the CARTOON into a representation of occluding edges (or material changes).

Once these contour segments have been made explicit, then there is the further task of linking these isolated contours. One important lesson demonstrated by the CARTOON algorithm is that, although occluding contours of necessity must be closed, they will rarely appear closed in a filtered image. Certainly color, texture, and grouping strategies based

upon the 3D properties of objects and surfaces are needed to determine which image contours properly belong together. These are challenging issues beyond early vision.

REFERENCES

Alpern, M., McCready, D.W. & Barr, L., "The dependence of the photopupil response on flash duration and intensity," *J. Gen. Physiol.* **47**(1963)265-278.

Attneave, F.,"Informational aspects of visual perception," *Psychol. Rev.* **61**(1954)183-193.

Barlow, H.B., "Three points about lateral inhibition," *Sensory Communication*, W.A. Rosenblith, ed., MIT Press, 1961, 217-234; 782-786.

Brindley, G.S., *Physiology of the Retina and Visual Pathway*, Williams & Wilkins, Baltimore, Maryland, 1970.

Creutzfeldt, O.D. & Nothdurft, H.C., "Representation of complex visual stimuli in the brain," *Naturwissenschaften* **65**(1978)307-318.

DeVries, H.D., "The quantum character of light and its bearing upon the threshold of vision," *Physica* **10**(1943)553-564.

Enroth-Cugell, C. & Robson, J.G., "The contrast sensitivity of retinal ganglion cells in the cat," *J. Physiol.* **187**(1966)517-552.

Hartline, H.K., "The response characteristics of single optic nerve fibers of the vertebrate eye to illumination of the retina," *Amer. J. Physiol.* **121**(1938)400-415.

Horn, B.K.P., "The application of Fourier Transform methods to image processing," *M.I.T., M.S. thesis,* 1968.

Horn, B.K.P., "The Binford-Horn Line-Finder," *MIT AI Memo No. 285*(1973).

Hubel, D.H. & Wiesel, T.N., "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol. Lond.* **160**(1962)106-154.

Hubel, D.H. & Livingstone, M.S., "Regions of poor orientation tuning coincide with patches of cytochrome oxidase staining in monkey striate cortex," *Soc. of Neurosci. Abstr.* **7**(1981)357.

Hueckel, M.H., "An operator which localizes edges in digital pictures," *J. ACM* **18**(1971)113-125.

Jernigan, M.E. & Wardell, R.W., "Does the eye contain optimal edge detection mechanisms?" *IEEE Trans. Systems, Main & Cyber., SMC-11*(1981)441-444.

Jones, R.C., "On the quantum efficiency of scotopic and photopic vision," *Jrl. Wash. Acad. Sci.* **47**(1957) 100-108.

Kelly, D.H., "Spatial frequency selectivity in the retina," *Vis. Res.* **15**(1975) 665-672.

Kuffler, S.W., "Discharge patterns and functional organization of mammalian retina," *J. Neurophysiol.* **16**(1953)37-68.

McLeod, I.D.G., "Comments on techniques for edge detection," *Proc. IEEE* **60**(1972)344.

Marr, D., "Early processing of visual information," *Phil. Trans. R. Soc. Lond. B.* **275**(1976)483-524.

Marr, D., *VISION: a computational investigation into the human representation and processing of visual information*, Freeman, San Francisco, 1982.

Marr, D. & Hildreth, E., "A theory of edge detection" *Proc. R. Soc. Lond.* **207**(1980)187-217.

Marr, D. & Nishihara, H.K., "Visual information processing: Artificial intelligence and the sensorium of light," *Tech. Rev.* **81**(1978)1-23.

Marr, D. & Poggio, T., "A computational theory of human stereo vision," *Proc. Roy. Soc. Lond. B.* **204**(1979)301-328.

Marr, D., Poggio, T. & Ullman, S., "Bandpass channels, zero crossings, and early visual information processing," *J. Opt. Soc. Am.* **69**(1979)914-916.

Nishihara, H.K., "Reconstruction of DOG filtered images from gradients at zero-crossings," *Proceedings of an Image Understanding workshop; report by P.H. Winston*(1980).

Nishihara, H.K. & Larson N.G., "Towards a real time implementation of the Marr and Poggio stereo matcher," *DARPA Image Understanding Workshop, Washington, DC, April 23, (Report No. SAI-82-391-WA)* (1981).

Persoon, E., "A new edge detection algorithm and its applications," *Computer Graphics and Image Processing* **5**(1976)425-446.

Pratt, W., *Digital Image Processing,* J. Wiley & Sons, New York, 1978.

Richards, W., "Quantifying sensory channels," *Sensory Processes* **3**(1980)207-299.

Richards, W., "Ideal lightness scale," *Applied Optics* **21**(1982)2569-2582.

Richards, W. & Polit, A., "Texture matching," *Kybernetik* **16**(1974)155-162.

Richter, J. & Ullman, S., "A model for the temporal organization of X and Y-type ganglion cells in the primate retina," *Biol. Cybern.* **43**(1982)127-145.

Rodieck, R.W., "Quantitative analysis of cat retinal ganglion cell response to visual stimuli," *Vis. Res.* **5**(1965)583-601.

Rose, A., "The relative sensitivities of television picture tubes, photographic films and the human eye," *Proc. Inst. Radio Erys.* **30**(1942)295-300.

Rosenfeld, A., *Picture Processing by Computer,* Academic Press, New York, 1969.

Rosenfeld, A., "A non-linear edge detection technique," *Proc. IEEE* **58**(1970)814-816.

Rosenfeld, A. and Kak, A. *Digital Picture Processing*, Academic Press, New York, 1976.

Rubin, J.M. & Richards, W.A., "Color vision and image intensities: When are changes material?" *MIT AI Memo No. 631; Perception, in press*(1982).

Sakitt, B. & Barlow, H.B., "A model for the economical encoding of the visual image in the cerebral cortex," *Biol. Cybern.* **43**(1982)97-108.

Schade, O.H., "Optical and photoelectric analog of the eye," *J. Opt. Soc. Am.* **46**(1956)721-739.

Schreiber, W.F. & Buckley, R.-R., "A two-channel picture coding system: Adaptive companding and color coding," *IEEE Trans. Commun., COM 29-12*(1981)1849-1858.

Shanmugan, K.S., Dickey, F.M. & Green, J.A., "An optimal frequency domain filter for edge detections in digital pictures," *IEEE Trans. on Pattern Analysis and Machine Intelligence PAMI-1* (1979)37-49.

Spitzberg, R. & Ricards, W., "Broad band spatial filters in the human visual system," *Vis. Res.* **15**(1975)837-841.

Stromeyer, C.F. III, Klein, S., Dawson, B.M. & Spillmann, L., "Low spatial-frequency channels in human vision: adaptation and masking," *Vis. Res.* **22**(1982)225-233.

Troxel, D.E., Schreiber, W.F., et al., "A two-channel picture coding system: Real time implementation," *IEEE Trans. Commun. COM 29-12*(1981)1841-1848.

Wilson, H.R. & Bergen, J.R., "A four mechanism model for spatial vision," *Vis. Res.* **19**(1979)19-32.