A locally connected network for real time target detection

Gregory Budzban
Southern Illinois University, Department of Mathematics
Carbondale, IL 62901-4408

Arthur Forman and Richard Skoblick
Martin Marietta Electronic Systems, MP 1304
Orlando, FL 32862-8007

Abstract

The detection of human-made objects with low false alarm rates in IR imagery remains a technically challenging problem. In addition, many currently proposed systems for autonomous air vehicles require the algorithms to process images at the rate of 30 frames/second (real-time). Parallel distributed processes, such as neural networks, offer potential solutions to problems of this complexity.

The current algorithm takes advantage of the presence of both long straight lines and curvature points in human-made objects. These features are among those recognized pre-attentively by the human visual system. It is a generalization of work done by Sha'Ashua and Ullman at MIT on the extraction of so-called salient features. The addition of curvature detection, however, is what allows the algorithm to achieve acceptable false alarm rates. On simulated FLIR imagery taken from the U.S. Army C^2NVEO terrain board, low false alarm rates have been achieved while maintaining 100% target detection.

1. Introduction

The detection of human-made objects in imagery of natural scenes is an extremely useful, but challenging problem. Detection as a vision module can be seen as the initial stage in a system that culminates in object identification. What is required from a detection algorithm is a kind of screening mechanism. In a sense, target detection acts as a system of attentional focus for the later stages of processing in a complete vision system.

Historically, there has been a great deal of work in the area of target detection. The majority of the effort falls into the general category of image segmentation. Segmentation attempts to determine homogeneous regions in the image of some predetermined size. Segmentation algorithms are usually used in conjunction with region labeling routines which declare a given region to be target versus non-target (for example). In Marr's vision hierarchy [1] segmentation and region labeling fall somewhere between low-level and mid-level vision depending on the complexity of the algorithm used. The process is often extremely unreliable, with infinitesmal changes in the image producing large changes in the output of the segmentation algorithm. This lack of continuity can be traced to the difficulty of defining exactly what one means by the term "region".

The process that controls changes in attentional focus in the human visual system, however, can find targets of interest reliably and almost instantaneously. In fact, that which controls attention must itself be pre-attentive, that is, pre-cognitive. Segmentation is too complex an operation to be the correct model for an activity that occurs this rapidly. The proper computational technique should be a local, parallel procedure based on properties of edges and not regions.

0-8194-0765-8/92/$4.00

This paper describes such a procedure. It generalizes the seminal work of Sha'ashua and Ullman [2]. It is based on the computation of a measure that is maximized for those structures which immediately attract attention in an image independent of the size or shape of the structure. Structures of this type will be refered to as "salient". Structural saliency can be thought of as a local measure of the probability that a physical boundary passes through a given pixel. The main contribution of this paper is the addition of the computation of curvature as the local measure of the probability of the existence of a zero dimensional boundary (i.e. a corner) at a given pixel.

The organization of the paper will be as follows. After this introduction, a discussion of structural saliency and our extension will follow. Section three will contain a detailed description of the geometry of the network and the algorithm. A discussion of results follow in section four and the paper will close with some final remarks.

2. Structural saliency and generalizations.

Sha'ashua and Ullman speak of two types of saliency: local and structural. Local saliency occurs if a an area of the image draws attention by virtue of some purely local qualitative property, such as color or intensity contrast. An area of the image possesses structural saliency due to a global geometric property, such as length or a particular structural pattern.

Most human-made objects bear a structural resemblance to one another. The success of many computer graphics modeling packages is based on the notion that most everyday objects are geometrical simple. In fact, polygonal solids can be used as rough low resolution models in many cases. The geometric patterns of these objects are structurally salient.

Structural saliency can be thought of as geometric information theory. Shannon's measure of information content in a one-dimensional signal is based on the probability distribution of the possible states of the signal. If there is only one state possible, then the information transmitted at each time interval is zero. Information exists only in the presence of change. For analytic surfaces, geometric change is measured by derivatives: a constant surface contains no geometric informtion. The geometric changes that encapsulate the most important information are the physical boundaries of an object.

Objects have boundaries of three types: surfaces, edges, and corner points. Surfaces are two dimensional boundaries and are the hardest to determine since they are so region-like. In fact as indicated previously, their nature may make the pre-attentive detection of them impossible. Edges are the one dimensional boundaries of surfaces and corner points are the zero dimensional boundaries of edges. The structural saliency algorithm we will descibe can be considered the determination of the potential that an edge or corner point exists at a given pixel.

Due to the presence of clutter and noise in an image, getting a strong return locally from a given edge operator can not be taken as undeniable evidence of the existence of a line. In the same way, high return from a curvature operator does not always imply the existence of a corner point. The definiteness of a boundary element can be measured in many ways. For one dimensional boundaries, if the edge has the properties of length, connectedness and lack of curvature one can be most sure that the edge is the true physical boundary of some target. Corner points, in addition to high curvature, must exist coincident with physical boundaries.

3. Network and Algorithm

As indicated in the introduction, the pre-attentive nature of detection is among its most important features. Thus the algorithm that computes the measure of structural saliency must be a local and parallelizable one. The measure derived by Sha'ashua and Ullman is indeed parallizable and is computed on a network of processing elememts whose geometry we will now describe.

At the heart of the saliency network are processing elements refered to as orientation elements. The orientation elements can best be thought of as directed elements in a graph. Each cell in the network has eight orientation elements, one directed at each of the eight nearest neighbors (see Fig. 1a). A given fixed orientation element has weighted connections with each of the eight orientation elements of the cell at which it is directed (though some of the weights may be zero). The orientations of the elements in each cell can be thought of as the eight possible local orientations of a line passing through the cell.

A simple edge operator is used to initiate the salience algorithm. The edge operator is run over the image and a threshold is applied. In practice, one wishes the procedure to be as independent of the threshold as possible, thus the smallest possible one is applied. For our results approximately five percent of the maximum value was used successfully. The algorithm is also not sensitive to the type of edge operator chosen. Both a Sobel and the difference of Gaussians (DOG) were used with good results, though the DOG did seem marginally better. The conjecture, though, is that any zero crossing operator will perform marginally better due to the connectivity of the resulting level curves.

At the start of the salience algorithm, the edge operator output is binarized and input into the network. Recall that the measure is optimized for edges that are long, connected, and straight. The formula, then, for the structural salience at the j th iteration for the i th orientation element is the following.

$$E_j^i = \sigma + \rho \max_{k=1,\ldots,8} \left[f_{ik} E_{j-1}^k \right] \quad \text{for } j = 1, \ldots, \text{ maxiterations}$$

Certain coefficients remain constant throughout the procedure. Sigma is either one or zero and is the input value from the edge operator. Rho is the penalty for gaps in the edges. It is set to one if sigma is one and set to 0.6 if sigma is zero, in our experiments. This value was simply decided upon empirically.

The value f_{ik} is the penalty for curvature. This value is defined by

$$f_{ik} = \exp\left(-2\alpha_{ik} \tan \frac{\alpha_{ik}}{2}\right)$$

where α_{ik} is the orientation difference between the element coming into the cell and element currently being updated.

This coefficient is best explained by considering figure 1b. Figure 1b shows the eight orientation elements that interact with a given center cell in the network. Information flows into the cell only from these particular elements. A given fixed one of these elements, say the one pointing from the north, forms eight different angles with the orientation elememts of the center cell. It is these angles or orientation differences to which the formula refers.

Thus, at each iteration the algorithm adds the weighted sum of the maximum salience contained in a neighboring cell. This maximum represents the neighboring cell most likely to be part of a line passing through the cell at that given orientation. Another implication of the formula is that at each iteration

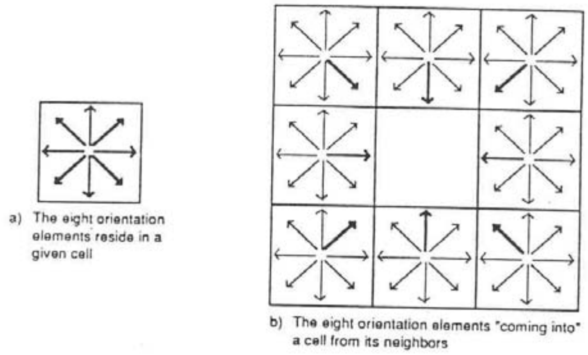

Figure 1: Directional Orientation of the Eight Elements for a Cell

the cell "remembers" from which direction it received the maximum energy. This can be used after completion of the salience algorithm to do boundary tracing.

As an example, consider a perfectly straight connected line of n pixels running from north to south in an image. After an edge operator is run, each pixel has a value of one. For this example, consider three pixels: the two endpoints and the midpoint. After n/2 iterations, the north pointing orientation element of the northern endpoint will have a value of n/2, as will the south orientation element of the south endpoint. At the midpoint, both the north and south pointing orientation elements will have a value of n/2. After n iterations, the north and south endpoints will have have values of n in their respective orientation elements while the values at the midpoint will not have changed. All subsequent iterations will produce no change in the values for the network.

After iterating the appropriate number of times, (usually less than or equal to the longest expected edge of the target of interest) the curvature operator is run. Experimentally, the best results occur after the image has been regularized with a significant Gaussian blur with a standard deviation of six or greater. Then a standard curvature gradient operator is run.

Final computation proceeds as follows. Normalize the salience output by the number of iterations, and for each cell determine the maximum orientation element and declare that value to be the salience for that given cell. Next, multiply the curvature gradient at each pixel by the largest salience value in the 3 x 3 neighborhood surrounding the pixel. Finally, do a local integration of both the salience and curvature values and chose those areas of the image which maximize the product of these two sums. Clearly, the size of the local integration depends upon the size of the targets of interest.

The algorithm descibed above is clearly parallel and is easily adapted to any standard four-connected SIMD machine. The operations for each iteration are simple arithmetic ones and with nearly any of the well known SIMD machines real-time computation is easily within reach. Only the integration steps are more than linear complexity as a function of the size of the target.

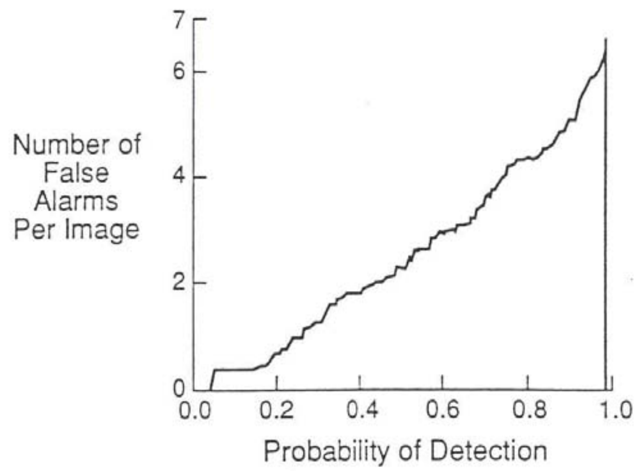

Figure 2: Salience Only Detection Results

4. Results

A series of exhaustive experiments were run on simulated FLIR imagery from a terrain board. The first series of experiments utilized only the edge boundaries and did not incorporate the curvature measurement. There were a number of features of this imagery that produced false alarms. Most notably, the horizon was a high contrast connected edge in some of the images and was easily the most salient structure using Sha'ashua and Ullman's definition. The receiver operating characteristic (ROC) curve for the series of experiments using only edge salience is shown in figure 2.

Next the same series of imagery was run using the curvature enhanced salience described in section three. In both experiments, the number of images was around one hundred and fifty. The addition of curvature as a pre-attentive detection of corner points produced a drastic improvement in the results as shown in the ROC curve of figure 3. With almost ninety-nine percent detection probability of target detection, there was well less than one false alarm per image.

5. Final comments

The attempt here describe how structural saliency can be used to measure local geometric imformation and thus act as a powerful target detection scheme. Other measures of geometric information are possible and are being explored. In particular, a natural way to use formal language theory would be to define a geometric pattern grammer and measure its information content.

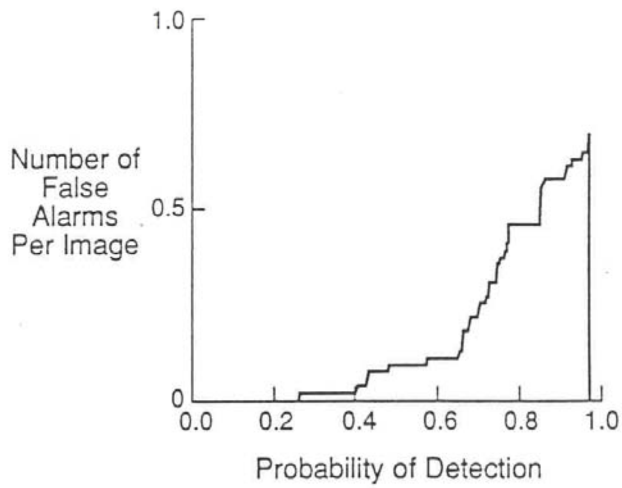

Figure 3: Curvature Enhanced Detection Results

References

[1] Marr, D., *Vision* San Francisco, CA: Freeman, 1982.

[2] Ullman S. and Sha'ashua, A., "Structural Saliency: The Detection of Globally Salient Structures using a Locally Connected Network," MIT AI Memo/061, 1988.