# EDGE-ONLY MATCHING TECHNIQUES IN ROBOT VISION

Shyamala Nagendran and Terry M. Caelli

Department of Psychology, The University of Alberta
Edmonton, Alberta T6G 2E9

**Abstract**

A computation procedure for detecting arbitrary two-dimensional signals embedded in scenes and independent of orientation/size is developed using pipe-line pixel processor procedures. Signals are encoded as edge-only features and cross-correlated with edge-only versions of the input scenes--both in cartesian and log-polar coordinates. These processes are incorporated into a robot visual system capable of locating, moving towards, and pointing to a target signal, again, independent of its size and orientation.

KEY WORDS: robot vision, pattern recognition, edge extraction, invariance coding

## 1. Introduction.

Many claims have been made as to the importance of edge information in coding images (Marr, 1982), pattern recognition (Rosenfeld & Kak, 1982) and fast computational vision in general. Secondly, convergent results from human psychophysics (Watt & Morgan, 1983), vertebrate physiology (Pollen, Andrews & Felden, 1978) and computional edge processing (Marr & Hildreth, 1980; Leclerc & Zucker, 1983) point to the "optimal" edge extractor as the logical intersection of band-pass Gaussian filters or pseudo-gamma function band-pass filters (Marr & Hildreth, 1980) approximated by $\nabla^2 G_\alpha$ operators (Watt & Morgan, 1983; Leclerc & Zucker, 1983): a Laplacian operator following low-pass Gaussian filters of specified bandwidths. In this paper we are concerned with using edge information for pattern recognition or pattern matching, as restricted to the two-dimensional environment--though including the problem of matching independent of signal orientation and size--required in our robot pattern recognition system (Figure 1).

Edge-based matching techniques are particularly useful in the context of pipe-line pixel processors where the execution time for convolution operations is critically dependent on kernel size, and the frame memories are restricted to 8- or 16-bit pixel sizes. Though our general model involves the correlation of edge-only pattern features at various levels of resolution according to a strict Laplacian pyramid format, in the present simulations we have restricted our analyses to the highest level of resolution: the original 512x512 input format.

A typical pipe-line pixel processor (Arithmetic Logic Unit: ALU-512) maps frame memories into frame memories with respect to the logical operations of (OR, exclusive OR, AND, 2's complement) and usual addition and subtraction operations. Each pass takes 33 msecs and operates on the full image, except when pixel protects are active. Information passes through a 16-bit register and the device also has an 8x8 bit multiplier--which enables convolution

**Figure 1**

*Robot and restricted visual environment used in simulations. Various image montages were placed about the walls and the robot's task was to detect where a specified signal was, move towards and point to it, based on visual information.*

operations on 512x512 images with $n$x$m$-sized kernels, and so taking $n$x$m$x33 msecs per convolution.

The robot was controlled by interrupt-driven subroutine calls capable of moving it in any horizontal direction and operating a three-joint arm according to the processed image information, as shown in Figure 1. Both image processing (Imaging Technology) and robot (RB Robots, Colorado) systems were controlled by a PDP 11/23 computer operating in RT11.

**2. Edge-extraction, invariance codes, and matching techniques.**

As implied above, the $\nabla^2 G(\alpha)$ operator (on an image $I$) is an adequate representative of band-pass filtering used in recent edge-extraction techniques. Defined by

$$\nabla^2 G(\alpha)(I) = \frac{\nabla^2 G(\alpha:x,y)(I)}{\partial X^2} + \frac{\nabla^2 G(\alpha:x,y)(I)}{\partial y^2} \quad (1)$$

where $G(\alpha:x,y) = e^{-\alpha^2((x-x_0)^2+(y-y_0)^2)}$. (2)

Here $(x_0,y_0)$ defines all convolution centers over the image. This filter is simple to approximate in a pipe-line pixel process by two convolution kernels. First, the Gaussian low-pass filter (2) can be approximated by the recursive use of an averaging kernels of the form:

$$A = 1/4 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad (3)$$

After $n$ recursions, it produces the bivariate binomial coefficients on an $(n+1)$x$(n+1)$ kernel as:

$$G \simeq B(n;x,y) = \frac{(n!)^2}{x!y!(n-x)!(n-y)!}, \; x,y = 0,..n \quad (4)$$

This takes $n$x132 msecs to complete and in the simulations to be reported here $n$ was set at 10 (a total of 33x4x10=1.32 sec).

The $\nabla^2$, or Laplacian kernel is defined (in finite difference form[2]) by

$$\nabla^2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (5)$$

taking 9x33=297 msec to complete.

Though various options are available for extracting the image edges as $(x,y)$ coordinates from the $\nabla^2 G$ images, we have used the following technique, since it only involves two passes through the pipe-line pixel processor. Since the output of the $\nabla^2 G(\alpha)$ operator can be positive or negative, 127 was added to the output, followed by an .AND. with 128, giving the complete form:

$$\{E(x,y)\} = \{\text{all } z(x,y)>0, \text{ from } [G_2[\nabla^2 G_{10}(I)+127]]\cap 128\} \quad (6)$$

Here, also, a second smoothing operation $(G_2)$ was employed to suppress isolated non-zero points resultant from the $\nabla^2 G_{10}$ operation. An illustration of these processes is shown in Figure 2, taking a total time of 1.75 sec. In our application the edge set of points (6) was (spatially) uniformly sampled to less than 256 points (rectangular area shown in Figure 2a), as shown in Figure 4b.

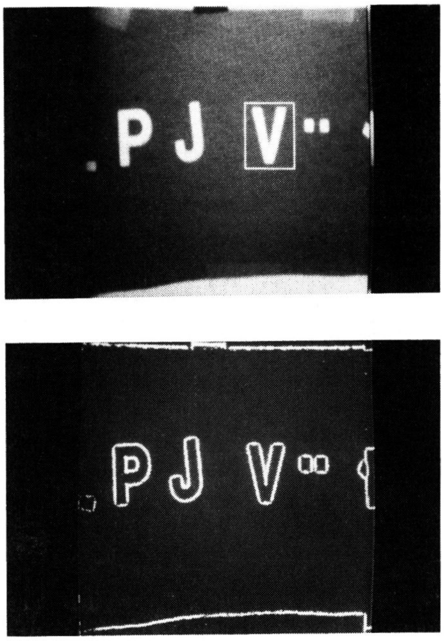



**Figure 2**

*(a) Input image ($I$ and (b) edge-only version created by: $G_2(\nabla^2 G_{10}(I)+127) \cap 128$ (rectangular area in (a) corresponds to signal used in matching process, Figure 3).*

Once this edge code has been established for a given signal, cross-correlation of the signal with any arbitrary scene can be reduced to a number of passes through the pipe-line processor equal to the number of signal points. That is, the cross-correlation function

$$C_{sg}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(\alpha,\beta)\,g(x+\alpha,y+\beta)\,d\alpha\,d\beta \qquad (7)$$

reduces to $C_{sg}(x,y) = \sum_{n=x_\ell}^{x_h} \sum_{m=y_\ell}^{y_h} s(n,m)g(x+n,y+m),$ (8)

$(x_l, y_l, x_h, y_h)$ corresponding to the sub-array containing the signal. For both $s$ and $g$ being binary valued (edge-only) images (8) reduces to

$$\hat{C}_{sg}(x,y) = \sum_{n,m\in E(x,y)} s(n,m)g(x+n,y+m). \qquad (9)$$

This correlation function can be executed by adding the image $g$ to itself shifted by the locus of points (x,y values) depicting the signal, or

$$\hat{C}_{sg}(x,y) = \sum_{n,m\in E(n,m)} g(x+n,y+m) \qquad (10)$$

since $s.g=1$ if, and only if, both signal and image edge components are present. (10) is readily executed via a series of passes through the pipe-line pixel procesor where the results are accumulated--as illustrated in Figure 3b.

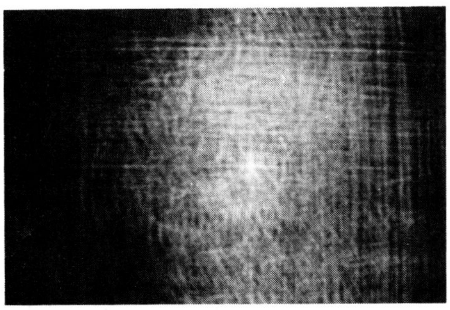


**Figure 3**

*Edge-only matching (cross-correlation) between sampled signal (a: see Figure 2b) and image (b). Luminance corresponds to the likelihood of signal presence.*

The problem with this matching technique, however, is that it is not invariant to size and orientation changes of the signal in the image. This is a particularly relevant problem in the case of robot vision studied here since movements of the robot towards the field wall introduces size and possibly small angular changes in the signal (Figure 1).

Techniques for matching which overcome rotations of signals have been recently developed by Hsu et al.[7] using circular harmonic functions whereby a Fourier decomposition of the polar transformed signal and rotated version is enacted along the orientation parameter. Matching occurs since both images have identical power spectra, with the phase giving the orientation differences between them. The problems associated with these techniques are (1) determining the initial estimate of the signal and image centers to enact the polar transforms, and (b) that the Fourier transform needs to be computed.

An alternate approach to the problem is simply to transform edge-only images into log-polar coordinates and enact matching using pixel processor operations. Again, this process has the problem of estimating the image center for the log-polar transform. For both signal and image the best *a priori* initial estimate of the center is the peak of the cartesian cross-correlation functions; though we are at present investigating a pyramid search technique to improve the accuracy and the speed of estimating the log-polar centre.

The log-polar (conformal) mapping is:

$$r' = 77.8 \log_e(r_0+1) \qquad (11)$$

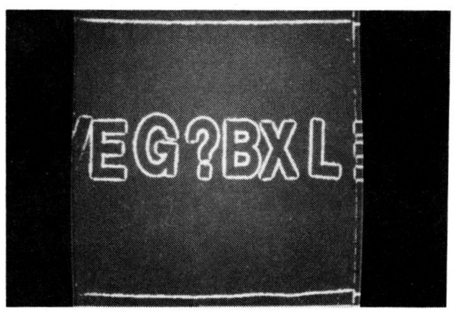
and $\theta = \tan^{-1}(y-y_0/x-x_0)$

where

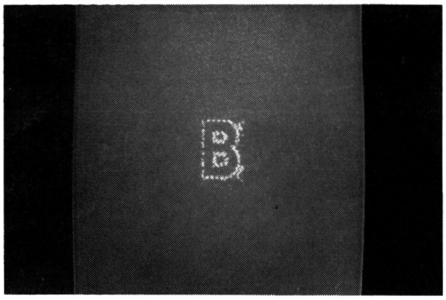
$$r_0 = \sqrt{(x-x_0)^2 + (y-y_0)^2} \qquad (12)$$

for $(x_0,y_0)$ corresponding to the peaks of signal autocorrelation and image cross-correlation images. Figure 4 shows an an example of signal and images transformed by the above procedure. Here, the peak of the log-polar cross-correlation function determines the particular signal size and orientation detected--with respect to the center chosen by the cartesian cross-correlation procedure.

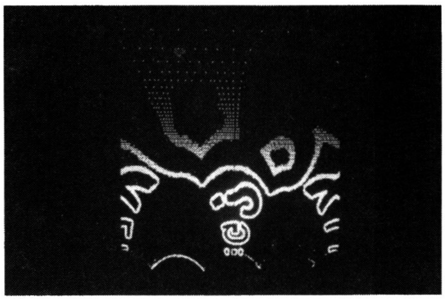
(a)



(b)



(c)



(d)

(e)

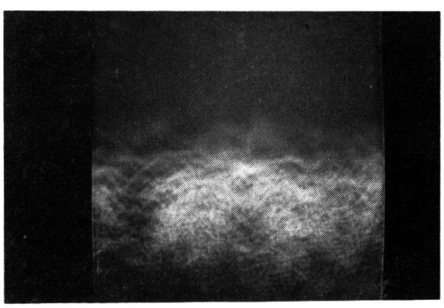


**Figure 4**

*(a), (b) show cartesian edge-only image and sampled signal respectively, while (c), (d) show corresponding log-polar versions of (a), (b). (e) Shows log-polar cross-correlation image where the peak defines the orientation and size of the signal embedded in the size which best matches the input signal.*

Though the basic computational vision procedures are defined and illustrated, a number of related problems to implementing these ideas into a pseudo-real time robot visual system must be resolved--particularly related to the analysis of error, setting of thresholds and adaptation procedures. Even in the restricted environment used in these simulations ambient luminance introduces photon fluxuations such that even redigitizing exactly the same image does not necessarily result in identical pixel values, albeit they are close. Of particular variability is the luminance projections around the field walls about which the robot moves.

For these reasons a signal matching threshold was set by digitizing the signal and then redigitizing the same area and determining the edge-only cross-correlation function. The peak value could then be used to estimate the signal match threshold--which in these simulations was set at 60% the number of signal points. Due to the rounding errors associated with the log-polar transform, the polar signal matching threshold was typically set at two-thirds that of the cartesian. Although these thresholds seem remarkably low, the checking procedure resulted in no false alarms in our restricted simulations--though more detailed signal detection analysis is being pursued.

**5. Conclusions.**

In these simulations we have demonstrated how pipe-line pixel processor technology may be effectively employed to enact edge-only matching of arbitrary signals to images. Though to this stage the choices of thresholds are arbitrary, our results using the log-polar mapping procedure for matching under rotations and size change along with the cartesian procedure resulted in successful matches for alphabetic characters and relatively complex images.

This log-polar matching procedure falls down when no evidence of a match occurs--that is, no discernible peak occurs in the cross-correlation function. In other attempts to solve this problem[9] an adaptive procedure is used to establish the center. However, this is time consuming in that the procedure involves changing between polar and cartesian coordinates. The second limitation of the above procedure is the simple use of signal sampling to keep the number of signal pan and scroll values (coordinates) below 255. This does not necessarily result in the signal features which are more likely to remain invariant to light fluxuations, image distortions, etc. However, these problems are currently under investigation, and the matching procedure and criteria are being analyzed from a signal detection perspective. Indeed, if pipe-line pixel processes could enact log-polar cross-correlations for every possible center, then these problems would be solved.

## References

1. Marr, D. (1982). Vision. San Francisco: Freeman.
2. Rosenfeld, A. & Kak, A. (1982). Digital Picture Processing: Vol. 1. New York: Academic Press.
3. Watt, R. & Morgan, M. (1983). The recognition and representation of edge blur: evidence of spatial primitives in human vision. Vision Research, 23, 1465-1477.
4. Pollen, D., Andrews, B. & Felden, S. (1978). Spatial frequency selectivity of periodic complex cells in the visual cortext of the cat. Vision Research, 18, 665-682.
5. Marr, D. & Hildreth, E. (1980). Theory of edge detection. Proceedings of the Royal Society (London), B207, 187-217.
6. Leclerc, Y. & Zucker, S. (1983). The local structure of image discontinuities in one dimension. Technical Report TR-83-19R, Department of Electrical Engineering, McGill University.
7. Hsu, Y., Arsenault, H. & April, G. (1982). Rotation-invariant digital pattern recognition using circular harmonic expansion. Applied Optics, 21(22), 4012-4015.