COMPUTER GRAPHICS AND IMAGE PROCESSING 9, 40–55 (1979)

Feature Detection Using Basis Functions

ROBERT A. HUMMEL

Department of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

Received March 13, 1978

A class of feature-detection operations using orthonormal basis functions is introduced. The basis functions are derived from the Karhunen–Loève expansion of the local image data. This theory is applied to define a new edge detector which can be adjusted to have increased sensitivity in any desired orientation.

1. INTRODUCTION

In his papers on edge and line detection in digital imagery [1, 2], Hueckel introduced a local operator which makes use of a finite set of orthonormal basis functions to reduce the complexity of the local data. His operator may be regarded as an example of a more general class of feature detectors which locate instances of a target set of patterns (such as edges or lines) in digitized pictures. Typically, the target patterns can be parameterized by variables representing the local mean gray level, orientation, size, variance, and possibly other properties, so that the pattern can be described by the values of a few identifying parameters. For example, the edge–line patterns described by Hueckel [2] can be specified by six independent parameters. The object of a simple feature detector is to find the parameter values of the target pattern closest to the local input data, and output some measure of the degree of match between the best pattern and the actual data. One method of performing this function is to execute an exhaustive search of the target patterns to locate the best parameters. An economical feature detector attempts to find the approximately best parameters without an exhaustive search. For this purpose, it is useful to project the initial local data and the set of target patterns onto a finite-dimensional subspace in order to reduce the complexity of the data, and to permit a functional dependence between the parameter values of the best pattern and the coordinate values of the initial data in the finite-dimensional subspace.

Let us consider the components of a general feature detector using orthonormal basis functions. We will eventually give selection criteria for the basis functions and define detectors which take into account initial bias, suitable for iterative implementations such as relaxation techniques [3].

Let $S_{p_1 \ldots p_k}(x, y)$ represent the k-parameter family of target patterns defined

0146-664X/79/010040-16$02.00/0

on the unit disk $\mathfrak{D}$. For an input signal $I(x, y)$ defined on $\mathfrak{D}$, we seek values $p_1, \ldots, p_k$ so that $\|I - S_{p_1 \ldots p_k}\|$ is minimized. The most convenient norm $\| \ \|$ is usually the L^2 norm

$$\|f\|^2 = \int\int_{\mathfrak{D}} f^2(x, y) dx dy.$$

Suppose M is a finite-dimensional subspace of $L^2(\mathfrak{D})$. We may then find an orthonormal basis for M, say $\{\varphi_1, \varphi_2, \ldots, \varphi_N\}$. Denote by I′ the orthogonal projection of I onto M. Thus

$$I'(x, y) = \sum_{i=1}^{N} c_i \varphi_i(x, y),$$

where

$$c_i = \int\int_{\mathfrak{D}} I(x, y) \varphi_i(x, y) dx dy.$$

Similarly, $S'_{p_1 \ldots p_k}$ denotes the projection of the pattern $S_{p_1 \ldots p_k}$. If M is chosen carefully, this projection will be one-to-one on the class of S-patterns. This permits us to find values for $p_1, \ldots, p_k$ so that $\|I' - S'_{p_1 \ldots p_k}\|$ is minimized.

Since $\|I' - S'\| \leq \|I - S\|$, we have $\min \|I' - S'\| \leq \min \|I - S\|$. Thus if $p_1, \ldots, p_k$ minimizes $\|I' - S'_{p_1 \ldots p_k}\|$ we have a lower bound for the distance from I to the mainfold of target patterns. This inequality has two important consequences:

(1) *A failure to find a target pattern on the M subspace fitting I', i.e., $\min \|I' - S'\|$ is large, signifies the nonexistence of a pattern fitting the input data.*

(2) *If I is exactly one of the S's, say $I = S_{p_1, \ldots, p_k}$, then $\|I' - S'_{p_1 \ldots p_k}\| = 0$ will be the unique minimum for $\|I' - S'\|$, so that the process of minimizing $\|I' - S'\|$ will find the exact parameters, $p_1, \ldots, p_k$.*

Thus the feature operator acts as a screen for possible pattern matches, since (1) ensures that eliminated patterns do not contain a match, whereas (2) ensures that no true match is missed. In addition, if M is chosen judiciously, we may assume that

(3) *The parameters minimizing $\|I' - S'_{p_1 \ldots p_k}\|$ are close in value to the parameters minimizing $\|I - S_{p_1 \ldots p_k}\|$, whenever I is fairly close to an S.*

The value of k unknowns $p_1, \ldots, p_k$ minimizing $\|I' - S'_{p_1 \ldots p_k}\|$ satisfy the k equations

$$\frac{\partial}{\partial p_i} \|I' - S'_{p_1 \ldots p_k}\| = 0, \qquad i = 1, \ldots, k. \tag{1}$$

In Hueckel's operator, M is chosen to be a nine-dimensional subspace so that the resulting equations may be reduced to a sequence of polynomials, which may be solved successively using the equations for roots to polynomials of degree four and less. To do this, he first isolates the angular dependence:

$$\|I' - S'_{\theta, p_2 \ldots p_6}\|^2 = f(\theta) + g(\theta, p_2, \ldots, p_6),$$

where

$$f(\theta) = \min_{p_2,\ldots,p_6} \|I' - S'_{\theta,p_2\ldots p_6}\|^2.$$

Thus g is nonnegative, and for a fixed θ one can find values $p_2(\theta), \ldots, p_6(\theta)$ solving $g(\theta, p_2(\theta), \ldots, p_6(\theta)) = 0$.

The parameter values $\theta, p_2, \ldots, p_6$ minimizing $\|I' - S'\|$ can then be found from

$$\frac{\partial f}{\partial \theta}(\theta) = 0, \qquad f(\theta) \text{ minimized;}$$

$$p_2 = p_2(\theta); \ldots;$$

$$p_6 = p_6(\theta).$$

It works out that finding θ minimizing $f(\theta)$ is equivalent to solving a quartic polynomial, while the remaining variables can be solved using the quadratic formula and solutions to linear equations. Unfortunately, this method does not always yield values of $p_2, \ldots, p_6$ which are within the permitted ranges, nor can one guarantee that the parameters will be real (as opposed to complex) valued [4]. Nonetheless, the error $\|I' - S'\|$ which results is a *lower bound* for the error $\|I - S\|$ minimized over all admissible patterns. As mentioned earlier, this effectively screens the data for potential pattern matches.

In this way, the parameters for the best edge/line projected onto M can be found by using formulas which depend only on the coefficients $c_1, \ldots, c_9$ of the basis functions in the expansion of I'. In fact, a large class of feature detectors can be regarded as functionals depending on a few coefficients in the expansion of the local gray level values in terms of a fixed set of basis functions. For example, the gradient operator depends on the samples obtained from differences of averages in the horizontal and vertical directions. The two corresponding basis functions have a region of positive support opposite a region of negative support. As another example, consider a linear operator

$$\mathcal{O} = \int\int_{\mathcal{D}} T(x, y) I(x, y) dx dy.$$

Clearly, this operator simply samples the coefficient of $T(x, y)$, a single basis function generating a one-dimensional subspace.

We would like a general method to derive a Hueckel-like operator for any specified feature detector. What basis functions should we use so as to best represent the feature information in the local data? The sample values obtained by integrating the local data against the basis functions must determine the parameter values of the best feature pattern. So the basis functions should attempt to preserve, as well as possible, the information about parameter values in the feature patterns. In addition, the equations governing the optimal parameter values will become too complicated for an economical algorithm if M is chosen incorrectly or too many basis functions are used to define M.

In Section 2, we describe a general method for finding a sequence of basis functions which preserve the feature information, and thus permit the development of a Hueckel-like feature detector. This method allows one to vary the basis functions according to initial expectations in the probability distribution of target patterns. Unfortunately, there is no guarantee that the resulting minimization problem will admit a functional solution, or even a fast iterative approximate solution. However, if the basis functions are kept small in number, and perhaps cleverly modified, the resulting minimization will sometimes be equivalent to finding the roots of low-order polynomials. In Section 3, we apply this theory to a simplified edge-detection problem, where the orientation of the edge is treated as the principal unknown. Section 4 extends this edge detector to the case where prior knowledge biases the expected orientation distribution of the edges. This extension demonstrates how the choice of basis functions can be used as a tool in a process of iterative refinement of feature parameters in scene analysis. Sections 5 and 6 present some experimental results and a summary of the principal results.

2. OPTIMAL BASIS FUNCTIONS

The basis functions which define a finite-dimensional subspace of local patterns and thereby determine the feature operator should preserve the feature information. One way to preserve information is to require that S' be a good approximation of S. In general, the discriminability of the parameters will be maximized providing that the projection of I to I' maps onto a subspace M which is "parallel" to the manifold of patterns S. These two observations coincide if we make certain assumptions about the set of target patterns S. Specifically, we assume that all patterns S have zero mean and unit mean square norm. This may be assumed without loss of generality, since each local input pattern can be normalized, thereby eliminating two parameters (which should be independent of the others) at the very start of the problem.

The primary selection criterion for M, accordingly, is that the expected value of $\|S - S'\|$ should be minimal. This expectation, of course, is taken over the set of target patterns parameterized by $p_1, \ldots, p_k$, and is weighted according to the probability density of occurrence of the patterns. If we regard the S's as a random field with probability density $f(p_1, \ldots, p_k)$, then the problem is to find orthornormal basis functions $\varphi_1, \varphi_2, \ldots$ such that the expected truncation error is minimized, i.e.,

$$E\{\|S - \sum_{i=1}^{N} a_i \varphi_i\|\} \text{ minimized,}$$

where

$$a_i = \int\int_{\mathfrak{D}} S(x, y)\, \varphi_i(x, y)\, dxdy.$$

This problem is solved by the classical theory of optimal sampling using the Karhunen–Loève basis functions [5, 6]. The solution yields the φ_i as solutions to the integral eigenvalue problem

$$\iint R(x, y, x', y') \varphi_i(x', y') dx'dy' = \lambda_i \varphi_i(x, y), \qquad \lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots > 0, \quad (2)$$

where

$$R(x, y, x', y') = E\{S(x, y) \cdot S(x', y')\}.$$

The autocorrelation function R can be computed:

$$R(x, y, x', y') = \int_{p_1} \int_{p_2} \cdots \int_{p_k} S_{p_1, \ldots, p_2}(x, y) S_{p_1, \ldots, p_k}(x', y') \times f(p_1, \ldots, p_k) dp_1 dp_2 \ldots dp_k. \quad (3)$$

Since R is an autocorrelation, it forms the kernel of a positive definite compact operator, so that the eigenvalues will be positive real and countable. The desired subspace M used to define the feature operator is obtained from the first few eigenfunctions corresponding to the largest eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$.

Sometimes Eq. (2) will admit an exact analytical solution. Sometimes the equation must be solved numerically by using a quadrature representation of the integral and standard matrix eigenanalysis routines. Alternatively, (2) can be transformed into a boundary value differential equation, which can be solved numerically. Once basis functions are determined, the complete feature detector can be defined.

Feature detection using orthonormal basis functions as proposed in Section 1 has three main components. First, one must find the coefficients of the local expansion

$$c_i = \frac{1}{|\mathfrak{D}|} \iint_{\mathfrak{D}} I(x, y) \varphi_i(x, y) dxdy, \quad (4a)$$

where $I(x, y)$ specifies the local data. Next, one must find the values of the parameters $p_1, \ldots, p_k$ which minimize the error

$$\epsilon^2(p_1, \ldots, p_k) = \iint_{\mathfrak{D}} [S'_{p_1}, \ldots, {}_{p_k}(x, y) - \sum c_i \varphi_i(x, y)]^2 dxdy. \quad (4b)$$

Finally, some measure of the quality of the match must be made to determine the output of the feature detector. The most logical measure is

$$Q = \iint [S_{p_1, \ldots, p_k}(x, y) - I(x, y)]^2 dxdy. \quad (4c)$$

The third step, which computes $\|I - S\|^2$, is unnecessary if the value of $\epsilon^2 = \|I' - S'\|^2$ is large enough to indicate the lack of a match. However, if the third step is ignored and ϵ^2 is used as an estimate for Q, then the operator may incorrectly respond when no true match is present.

The parameter values minimizing Eq. (4b) will satisfy Eq. (1). This crucial step of the feature detector may pose mathematically intractable problems, especially if too many basis functions are used. In the example described in Sections 3 and 4, however, no difficulties arise.

We should mention that the idea of employing Karhunen–Loève basis functions to assist in feature selection is a standard technique in the field of pattern recognition [11]. In pattern recognition, however, the goal is to recognize an instance of one of the feature samples without specifying the parameter values of that sample. Our purpose in using Karhunen–Loève basis functions is different: We wish to preserve the feature parameter information in order that the solution to (4b) should yield the most likely feature pattern candidate.

3. EDGE DETECTION

Let us restrict our attention to the detection of edges which pass through the central point of a circular neighborhood. This restriction will require the application of edge detectors at every point, and may require the use of thinning after edges are detected. However, these detectors are uniformly sensitive across the image, which is desirable for parallel implementations. Moreover, we avoid the complex-root objections to the Hueckel operator [4]. The more general case, including parameters for edge displacement, could be treated numerically using a more complicated version of the techniques demonstrated here.

For comparison, a survey of common edge techniques can be found in [7]. In particular, Griffith's operator [8] and Chow's operator [9] (in addition to Hueckel's) attempt to find an optimal edge based on a model of ideal edges. Other approaches to edge detection are discussed in [5].

Our model of ideal edges is a one-parameter family of functions defined on the unit disk whose representation in polar coordinates (with the origin at the center) is given by

$$S_{\Psi}(r, \theta) = \chi(\theta - \Psi), \tag{5}$$

where χ is a periodic function which is $+1$ on the upper half circle and -1 on the lower half circle. Recall that these target patterns have been normalized to have zero mean and unit mean square norm.

Suppose that the expected probability density of these edges is given by $f(\Psi)$. We assume that $f(\Psi + \pi) = f(\Psi)$, which says that S_{Ψ} and $S_{-\Psi}$ $(= -S_{\Psi})$ occur equally often. By permitting nonuniform densities, the edge detector may be designed to be more sensitive to certain orientations (Section 4).

The autocorrelation function (Eq. (3)) can be obtained by integrating with respect to $f(\Psi)d\Psi$. In polar coordinates, the function $R(r_1, \theta_1, r_2, \theta_2) = R(\theta_1, \theta_2)$ satisfies

$$R(\alpha, \beta) = 1 - 4|P(\alpha) - P(\beta)|, \qquad \text{for} \qquad 0 \leq \alpha, \quad \beta \leq \pi,$$

$$R(\alpha, \beta) = R(\beta, \alpha), \qquad \text{and} \qquad R(\alpha + \pi, \beta) = -R(\alpha, \beta),$$

where

$$P(\alpha) = \int_0^{\alpha} f(\Psi)d\Psi, \qquad 0 \leq \alpha \leq 2\pi. \tag{6}$$

The basis functions are eigenfunctions of the eigenvalue problem in Eq. (2).

Since the autocorrelation function is independent of the radial components r_1, r_2, the eigenfunction $\varphi_i(r, \theta)$ will likewise be independent of r. This reduces the eigenvalue problem to finding a sequence of basis functions $\varphi_i(r, \theta) = \varphi_i(\theta)$, $i = 1, 2, \ldots$, satisfying

$$\int_{-\pi}^{\pi} R(\alpha, \beta)\varphi_i(\beta)d\beta = \lambda_i\varphi_i(\alpha), \qquad \lambda_1 \geq \lambda_2 \geq \cdots.$$

By differentiating twice, one can transform the integral eigenvalue problem into a differential equation

$$(\varphi'_i(\alpha)/f(\alpha))' + (4/\lambda_i)\varphi_i(\alpha) = 0. \tag{7}$$

Boundary conditions must be imposed and can be derived from the property $R(\alpha + \pi, \beta) = -R(\alpha, \beta)$ and the integral equation. One obtains $\varphi_i(\alpha + \pi) = -\varphi_i(\alpha)$. Note that φ_i is periodic with period 2π.

For example, suppose that the probability density function is uniform, so that $f(\alpha) \equiv 1/2\pi$. Then (7) becomes

$$\varphi''_i(\alpha) + \frac{8\pi}{\lambda_i}\varphi_i(\alpha) = 0,$$

with the condition $\varphi_i(\alpha + \pi) = -\varphi_i(\alpha)$. Solutions are obtained for $\lambda_n = 8\pi/n^2$, $n = 1, 3, 5, \ldots$; each eigenspace is spanned by the functions $\varphi_n(\theta) = \cos(n\theta)$ and $\varphi_{n+1}(\theta) = \sin(n\theta)$. Thus the first four basis functions for optimal edge detection with unbiased initial density (see Fig. 1) are given by

$$\begin{aligned}
\varphi_1(r, \theta) &= \cos\theta,\\
\varphi_2(r, \theta) &= \sin\theta,\\
\varphi_3(r, \theta) &= \cos 3\theta,\\
\varphi_4(r, \theta) &= \sin 3\theta.
\end{aligned}$$

It is of interest to compare these basis functions with those used by Hueckel. The first two functions are essentially Hueckel's H_2 and H_3 (without feathering the functions to zero at the edges). The Hueckel edge detector must be able to detect laterally displaced edges within the neighborhood and accordingly includes basis functions suitable for that purpose which are not present here. However, none of his basis functions measure the $\cos 3\theta$ and $\sin 3\theta$ components. By including these basis functions, our operator should be more orientation specific.

The complete operator is derived from Eqs. (4a), (4b), and (4c). The value of θ_o which minimizes (4b) can be shown to satisfy the equation

$$\sum_{i=1}^{N} c_i\varphi_i(\theta_o) = 0. \tag{8}$$

The coefficients c_i are obtained from Eq. (4a), using normalized input data $I(x, y)$. Note, however, that the normalization process simply transforms the coefficients by a constant multiplicative factor when the above basis functions are used.

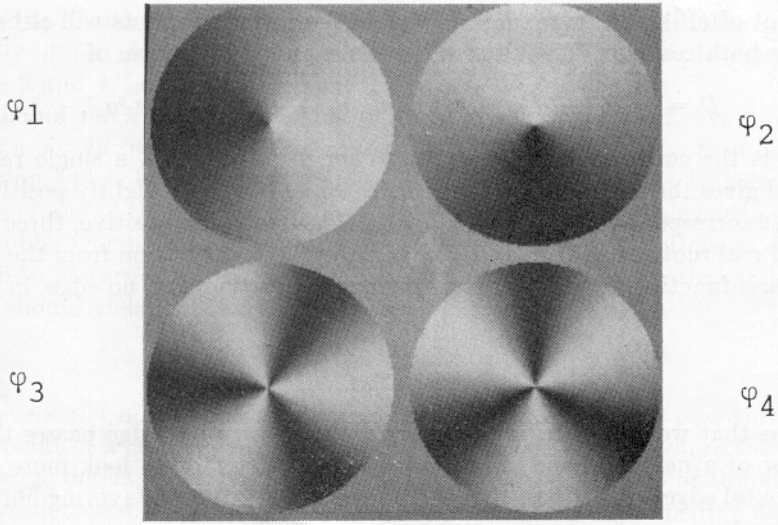

FIG. 1. Templates for first four basis functions, unbiased edge detection. Background gray level = 0.

Thus the solution of (8) remains unchanged, and so normalization of local input data is unnecessary.

For $N = 2$, (8) becomes

$$c_1 \cos \theta_o + c_2 \sin \theta_o = 0, \tag{9a}$$

and for $N = 4$, we have

$$c_1 \cos \theta_o + c_2 \sin \theta_o + c_3 \cos 3\theta_o + c_4 \sin 3\theta_o = 0. \tag{9b}$$

The case $N = 2$ is solved by the formula $\theta_o = \arctan(-c_1/c_2)$, or by noting that the vector (c_1, c_2), points in a direction normal to the proposed edge. The $N = 4$ case may be reduced to a cubic polynomial equation, with either a single real root θ_o (modulo π), or three distinct roots, one of which minimizes (4b). However, if three roots are found, then the function $\sum c_i \varphi_i(\theta)$, which approximates $I(\theta)$, certainly does not represent a "good" edge in any orientation. If three real roots occur, the edge detector may return immediately with the response that no edge was found.

The reduction to a cubic equation may be done in several ways. One method is to expand the $3\theta_o$ terms in (9b), divide by $\cos^3 \theta_o$, and substitute $1 + \tan^2 \theta_o$ for $\sec^2 \theta_o$ to obtain

$$(c_2 - c_4)x^3 + (c_1 - 3c_3)x^2 + (c_2 + 3c_4)x + (c_1 - c_3) = 0, \tag{10}$$

where $x = \tan \theta_o$. If $\cos \theta_o = 0$, this derivation is improper. However, this case arises only when $\theta_o = \pi/2$ is a solution of (9b), which requires that $a_2 = a_4$. Thus if (10) reduces to a quadratic equation, the "missing root" is $\theta_o = \pi/2$.

Note also that the first two basis functions have attenuated support regions in the diagonal directions, i.e., $\varphi_1(r, \pi/4) = 1/2^{\frac{1}{2}} < 1$. In the past, essentially intuitive arguments have been used to derive edge detectors with the same qualitative form (see, e.g., the Sobel edge detector in [12, p. 271]).

One root of (10) will always be real, while the remaining roots will either both be real or both complex. The discriminant is a positive multiple of

$$D = 4a_3a_1^2 - a_2^2a_1^2 - 18a_3a_2a_1a_0 + 4a_2^3a_0 + 27a_3^2a_0^2,$$

where a_i is the coefficient of x^i in (10). When D is negative, a single real root exists and gives the optimum value for θ_o. When D is zero or slightly positive, the optimum θ corresponds to the isolated root. If D is large and positive, three widely separated real roots exist. This case arises when the contribution from the second pair of basis functions is greater than the first, indicating that no edge, in fact, is present.

4. ORIENTATION BIAS

Suppose that we have reason to believe that a horizontal edge passes through the center of a neighborhood, and want the edge detector to look more closely for horizontal edges. We choose a nonzero density distribution favoring horizontal edges given by

$$f_r(\alpha) = \frac{1}{2\pi}\left(\frac{1 - r^4}{(1 + r^2)^2 - 4r^2\cos^2\alpha}\right), \tag{11}$$

where $0 \leq r < 1$ is a parameterization of the degree of bias for horizontal edges. This distribution was chosen using the following considerations:

(i) $f_0(\alpha) \equiv 1/2\pi$;
(ii) $f_1(\alpha) = 1/2(\delta_0(\alpha) + \delta_\pi(\alpha))$;

(iii) $\int_{-\pi}^{\pi} f_r(\alpha)d\alpha = 1, \qquad 0 \leq r \leq 1$;

(iv) Thus $u(re^{i\alpha}) = f_r(\alpha)$ is harmonic at 0.

Equation (11) represents the unique set of functions f_r satisfying (i), (ii) such that $u(z)$ defined in (iv) is harmonic in the disk $|z| < 1$.

By expanding $\varphi_i(\alpha)$ in terms of cosines and sines, the differential equation (7) becomes an infinite matrix problem, which may be solved numerically to any degree of accuracy desired (see Appendix). Alternatively, one could choose to find a numerical solution to the integral eigenvalue problem directly, using the kernel obtained by substituting (11) into (6): $R(\alpha, \beta) = 1 - 4|P(\alpha) - P(\beta)|$, with $P(\alpha) = 1/2\pi \arctan((1 + r^2)/(1 - r^2)\tan\alpha)$. Here the arctan branch must be chosen so that the angle returned is in the same quadrant as α.

Figure 2 displays the first four basis functions (corresponding to the largest eigenvalues $\lambda_1, \ldots, \lambda_4$) for optimal sampling of edges, as computed numerically for $r = 0.8$. Note how the bias for horizontal edges is reflected by larger magnitudes in the vertical directions, thus using the values in the normal directions to the edge to a greater extent. Basis functions which are designed to favor edges in any particular orientation may be obtained by simply rotating these solutions by the appropriate angle. The angular dependence of φ_1 of Fig. 2 is shown in Fig. 3.

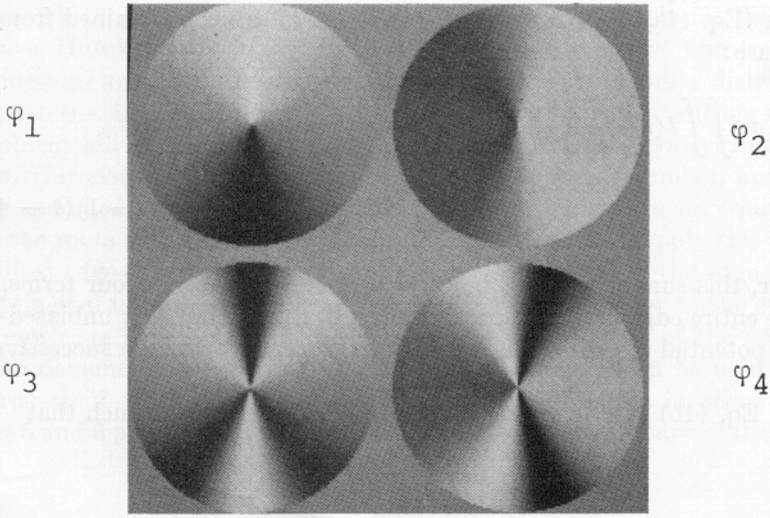

FIG. 2. Templates for first four basis functions, horizontal edge bias, $r = 0.8$. Note that φ_1 is the sine series and so measures the step across the horizontal axis.

Thus, as shown in the Appendix, for bias in the Ψ direction $\varphi_i(\theta)$ is given by

$$\varphi_i(\theta) = \sum_{k=1}^{\infty} a_k{}^i \cos\,(2k-1)(\theta - \Psi) + b_k{}^i \sin\,(2k-1)(\theta - \Psi),$$

where for fixed i, either $a_k{}^i = 0$ for all k or $b_k{}^i = 0$ for all k. The coefficients $\{a_k{}^i\}$ and $\{b_k{}^i\}$ are obtained as the eigenvectors of infinite matrix problems.

The biased edge detector can now be defined. First, the coefficients c_i are obtained from a local sum of the input data against the basis functions. Note that

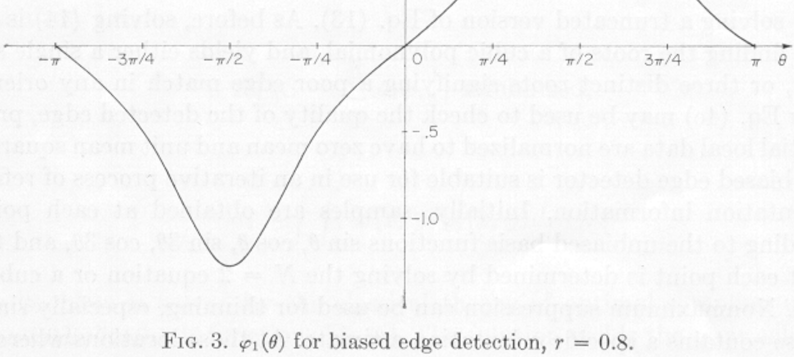

FIG. 3. $\varphi_1(\theta)$ for biased edge detection, $r = 0.8$.

this sum (Eq. (4a)) can be rewritten as a sum of samples obtained from the unbiased case:

$$c_i = \sum_{k=1}^{\infty} a_k{}^i \int\int I(r, \theta) \sin ((2k - 1)(\theta - \Psi)) r dr d\theta$$
$$+ \sum_{k=1}^{\infty} b_k{}^i \int\int I(r, \theta) \cos ((2k - 1)(\theta - \Psi)) r dr d\theta.$$

However, this sum should not be truncated to include only four terms, because then the entire edge detection process depends on the first four unbiased samples, and the potential advantage of iterative improvement through successive biasing is lost.

Next, Eq. (4b) can be approximately solved by finding θ_o such that

$$\sum_{i=1}^{N} c_i \varphi_i(\theta_o) = 0. \tag{12}$$

Actually, this equation corresponds to minimizing $\|S_{\theta_o} - I'\|$; the approximation is very close provided θ_o is close to Ψ. To solve (12), for example, for $N = 4$, one must find θ_o such that

$$\sum_{k=1}^{\infty} (\sum_{i=1}^{4} a_k{}^i c_i) \cos (2k - 1)(\theta_o - \Psi) + (\sum_{i=1}^{4} b_k{}^i c_i) \sin (2k - 1)(\theta_o - \Psi) = 0. \tag{13}$$

The coefficients are easily obtained from the known eigenvectors and the samples c_i. Rather than searching for a zero θ_o of (13), it is easier to find an approximate solution by solving

$$\tilde{c}_1 \cos (\theta_o - \Psi) + \tilde{c}_2 \sin (\theta_o - \Psi) + \tilde{c}_3 \cos 3(\theta_o - \Psi) + \tilde{c}_4 \sin 3(\theta_o - \Psi) = 0,$$

where

$$\tilde{c}_1 = \sum_{i=1}^{4} a_1{}^i c_i, \qquad \tilde{c}_2 = \sum_{i=1}^{4} b_1{}^i c_i,$$
$$\tilde{c}_3 = \sum_{i=1}^{4} a_2{}^i c_i, \qquad \tilde{c}_4 = \sum_{i=1}^{4} b_2{}^i c_i, \tag{14}$$

i.e., by solving a truncated version of Eq. (13). As before, solving (14) is equivalent to finding the roots of a cubic polynomial, and yields either a single solution $\theta_o - \Psi$, or three distinct roots signifying a poor edge match in any orientation. Finally Eq. (4c) may be used to check the quality of the detected edge, providing the initial local data are normalized to have zero mean and unit mean square norm.

The biased edge detector is suitable for use in an iterative process of refinement of orientation information. Initially, samples are obtained at each point corresponding to the unbiased basis functions $\sin \theta$, $\cos \theta$, $\sin 3\theta$, $\cos 3\theta$, and the best edge at each point is determined by solving the $N = 2$ equation or a cubic polynomial. Nonmaximum suppression can be used for thinning, especially since each response contains a specific orientation estimate. At those locations where a good edge is detected, a biased edge detector is applied. The orientation and degree of bias should be determined by the initial response of the edge detector at that point. The response at neighboring points may be allowed to affect the biasing, as in a relaxation labeling process [3]. This process is repeated until the amount of biasing is large enough to justify the application of an orientation-specific edge detector.

5. SOME EXPERIMENTS

Several experiments were conducted to study the advantages of using optimal basis functions for edge detection. In particular, a comparison was made between the results obtained using the unbiased edge operator described above and using an operator described by Meró and Vássy [10], which uses simple step functions in the x and y directions as its basis functions. This comparison is useful, since the basis functions described here are truly optimal only for continuous images, not discrete ones, and are based on a model of ideal edges not always present in real image data.

The comparison experiments used neighborhoods of sizes 5×5, 7×7, and 9×9. The Hueckel operator [2] was also included in the latter two experiments, though its inclusion is somewhat unfair since it was designed for use with slightly larger windows.

Figure 4 shows the results of applying the Hueckel, Meró/Vássy, and "optimal" operators to an artificial image containing edge slopes of 5, 10, 20, 25, and 40° from the horizontal or vertical, with and without noise. (The noise was uncorrelated, approximately binomial, zero-mean, with a standard deviation of 18 gray levels (the contrast is 45 levels).) The Hueckel operator performs more poorly, while the other two operators perform quite comparably. (The optimal operator's output appears to broaden somewhat less when the neighborhood size

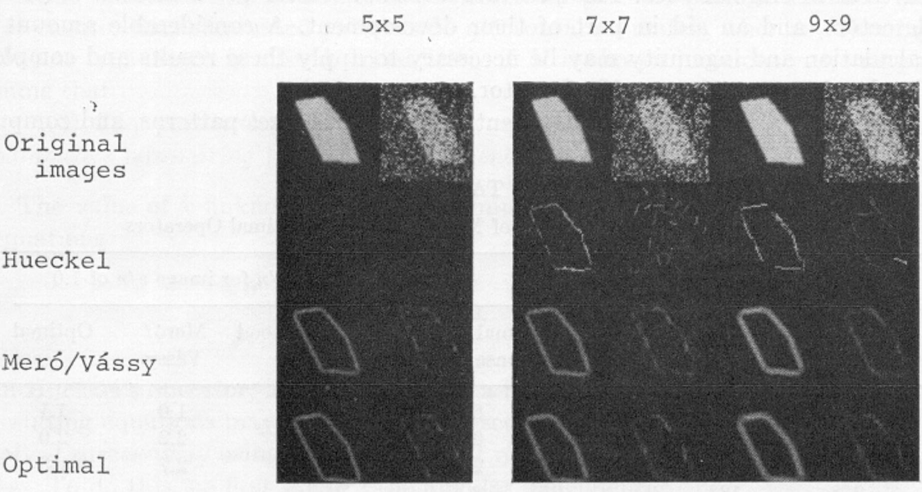

FIG. 4. Results of applying Hueckel, Meró/Vássy, and optimal operators to an artificial image with and without noise, using 5×5, 7×7, and 9×9 neighborhoods.

FIG. 5. Results for an image of a face, using 7 × 7 neighborhoods.

is increased, as a result of the inclusion of two additional basis functions.) Figure 5 shows results for a 7 × 7 neighborhood size on an image of a face.

Quantitative comparison of the (serial) Hueckel operator with the (conceptually parallel) Meró/Vássy and optimal operators is difficult; but the latter two can be quantitatively compared, as in Table 1. The Meró/Vássy operator seems to be at least as good as the optimal operator at detecting edges in the presence of uncorrelated noise, but it is somewhat poorer at estimating the slopes of the edges.

6. SUMMARY

The principal result of this paper is that the Karhunen–Loève sampling theorem may be used to determine basis functions for the design of feature detectors which attempt to locate and identify instances of a class of target patterns in original data. This provides a framework for the definition of feature detectors, and an aid in part of their development. A considerable amount of calculation and ingenuity may be necessary to apply these results and complete the development of a specific detector.

Specifically, the designer must identify his set of target patterns, and compute

TABLE 1

Quantitative Comparisons of Meró/Vássy and Optimal Operators

Edge slope estimation					Edge s/n for image s/n of 1.0		
True slope	Meró/Vássy response		Optimal response		Neighborhood size	Meró/Vássy	Optimal
5°	8°	8	1°	6	5 × 5	1.0	1.1
10°	6°	7	7°	9	7 × 7	2.2	2.0
20°	13°	3	19°	2	9 × 9	2.7	2.6
25°	18°	3	25°	2			
40°	39°	8	40°	5			

their autocorrelation statistics. The application of the Karhunen–Loève theorem may require a numerical solution to the eigenvalue problem. He must then design an operator which identifies that patterns in the data projected onto the span of the basis functions. The latter task has been regarded, in this paper, as a variational problem of determining parameter values for the L^2-nearest projected target pattern to the projected input data. The resulting equations may need to be simplified or approximated in order to allow solvability. Other kinds of operators are possible, however. For example, the designer might simply measure the strength of the projection of normalized input data as in the pattern recognition theory of feature selection using basis functions [11]. If the subspace determined by the basis functions closely approximates the target pattern, this strength should be high when a match is present.

The advantages of the optimal basis function approach to feature detection include (i) detectors which screen for target patterns with a low miss rate, and (ii) detectors which can be biased to "look more closely" for a particular pattern. The operator may incorrectly respond when no match is present, although the use of "optimal" basis functions attempts to minimize this problem. The mathematical ability to bias the operator and reexamine the data in light of a priori knowledge fits into the relaxation paradigm in which different sources of knowledge are allowed to influence the updating of the labeling. Since the synthesis of a decision is based upon cooperating and competing sources, the potential for biasing the examination of one feature represents a possibility for eliminating incongruities. The ability of low-level processes to "look back" at original data or other feature data has sometimes been viewed as an unavailable or inappropriate option, and sometimes simply rejected. While the mathematical capability does not argue that the technique is "visionlike," the availability of classes of biased feature detectors can certainly aid in the construction of vision systems in which many knowledge sources ultimately cooperate in a labeling decision.

APPENDIX: MATRIX DERIVATION OF THE EIGENFUNCTIONS

Expand

$$\varphi_i(\alpha) = \sum_{\substack{n \text{ odd} \\ n \geq 1}} a_n \cos n\alpha + b_n \sin n\alpha.$$

Then since

$$\frac{1}{f(\alpha)} = 2\pi \left[\frac{(1 + r^2)^2}{1 - r^4} - \frac{4r^2}{1 - r^4} \cos^2 \alpha \right],$$

we have

$$\frac{\varphi'_i(\alpha)}{f(\alpha)} = \sum_{n \text{ odd}} \left\{ \frac{2\pi(1 + r^2)^2}{1 - r^4} [-na_n \sin n\alpha + nb_n \cos n\alpha] + \frac{8\pi r^2}{1 - r^4} [na_n \cos^2 \alpha \sin n\alpha - nb_n \cos^2\alpha \cos n\alpha] \right\}.$$

Using $\cos^2 \alpha \sin n\alpha = \frac{1}{4} \sin (n-2)\alpha + \frac{1}{2} \sin n\alpha + \frac{1}{4} \sin (n+2)\alpha$ and $\cos^2 \alpha \cos n\alpha = \frac{1}{4} \cos (n-2)\alpha + \frac{1}{2} \cos n\alpha + \frac{1}{4} \cos (n+2)\alpha$, after combining we obtain

$$\frac{\varphi'_i(\alpha)}{f(\alpha)} = \sum_{n \text{ odd}} \left\{ \frac{2\pi(1+r^4)}{1-r^4} [-na_n \sin n\alpha + nb_n \cos n\alpha] + \frac{2\pi r^2}{1-r^4} [na_n \sin (n-2)\alpha + na_n \sin (n+2)\alpha - nb_n \cos (n-2)\alpha - nb_n \cos (n+2)\alpha] \right\}.$$

Letting $\sigma = 2\pi(1+r^4)/(1-r^4)$, and $\tau = 2\pi r^2/(1-r^4)$, and differentiating, we obtain

$$\begin{aligned}(\varphi'_i(\alpha)/f(\alpha))' &= \sum_{n \text{ odd}} \{\sigma[-n^2 a_n \cos n\alpha - n^2 b_n \sin n\alpha] \\ &\quad + \tau[n(n-2)a_n \cos (n-2)\alpha + n(n+2)a_n \cos (n+2)\alpha \\ &\quad + n(n-2)b_n \sin (n-2)\alpha + n(n+2)b_n \sin (n+2)\alpha]\} \\ &= (-\sigma - \tau)a_1 \cos \alpha + (-\sigma + \tau)b_1 \sin \alpha + 3\tau a_3 \cos \alpha + 3\tau b_3 \sin \alpha \\ &\quad + \sum_{\substack{n \text{ odd} \\ n \geq 3}} [\tau n(n-2)a_{n-2} - \sigma n^2 a_n + \tau n(n+2)a_{n+2}] \cos n\alpha \\ &\quad + [\tau n(n-2)b_{n-2} - \sigma n^2 b_n + \tau n(n+2)b_{n+2}] \sin n\alpha.\end{aligned}$$

Adding $4\varphi_i(\alpha)/\lambda_i$, and setting the result equal to 0, since each term must be 0, the vector $\mathbf{a} = (a_1, a_3, a_5, \ldots)$ satisfies the matrix equation $A\mathbf{a} = (4/\lambda_i)\mathbf{a}$, where

$$A = \begin{bmatrix} (\sigma+\tau) & -3\tau & 0 & 0\ldots \\ -3\tau & 3^2\sigma & -3\cdot5\tau & 0\ldots \\ 0 & -3\cdot5\tau & 5^2\sigma & -5\cdot7\tau \\ & \ldots & & \end{bmatrix},$$

and $\mathbf{b} = (b_1, b_3, b_5, \ldots)$ satisfies the matrix equation $B\mathbf{b} = (4/\lambda_i)\mathbf{b}$, where

$$B = \begin{bmatrix} (\sigma-\tau) & -3\tau & 0 & 0\ldots \\ -3\tau & 3^2\sigma & -3\cdot5\tau & 0\ldots \\ 0 & -3\cdot5\tau & 5^2\sigma & -5\cdot7\tau \\ & \ldots & & \end{bmatrix}.$$

For $r \neq 0$, the eigenvalues will be distinct, so that a simultaneous solution for a nonzero $\varphi_i(\alpha)$ will have either $\mathbf{a} = 0$ or $\mathbf{b} = 0$ (that is, $\varphi_i(\alpha)$ will be a pure cosine series or pure sine series).

A Rayleigh–Ritz method was used to find eigenvalues and eigenvectors of these matrices, although more sophisticated procedures may be employed for greater accuracy. Once an eigenvector for one of the matrices is found, the corresponding eigenfunction of the integral equation can be obtained from the expansion for $\varphi_i(\alpha)$.

For example, the functions in Fig. 2 were obtained by finding eigenvalues and eigenvectors of finite versions of the matrices A and B. For matrices of size 20×20, all values had converged to machine precision. The smallest two eigenvalues of matrix B (corresponding to $4/\lambda_1$ and $4/\lambda_2$) have approximate values 3.625 and 76.415, with corresponding eigenvectors (0.971, −0.230, 0.056, −0.012, ..., 0), and (0.217, 0.769, −0.546, −0.080, ..., 0). If we call these vectors $(b_1^1, b_2^1, b_3^1, \ldots, b_{20}^1)$ and $(b_1^2, \ldots, b_{20}^2)$, respectively, the corresponding eigenfunctions for the integral equation are very nearly

$$\varphi_{s,1}(\theta) = \sum_{n=1}^{20} b_n^1 \cdot \sin((2n-1)\theta),$$

$$\varphi_{s,2}(\theta) = \sum_{n=1}^{20} b_n^2 \cdot \sin((2n-1)\theta).$$

The two smallest eigenvalues of a 20×20 version of matrix B have approximate values 16.324 and 77.291, and give rise to two eigenfunctions $\varphi_{c,1}$ and $\varphi_{c,2}$ from the corresponding cosine series. Since the basis eigenfunctions $(\varphi_1, \varphi_2, \varphi_3, \varphi_4, \ldots)$ should be ordered so that $4/\lambda_1 \leq 4/\lambda_2 \leq \ldots$, the appropriate order of the basis functions is $(\varphi_{s,1}, \varphi_{c,1}, \varphi_{s,2}, \varphi_{c,2}, \ldots)$.

ACKNOWLEDGMENTS

The support of the U.S. Army Night Vision Laboratory under Contract No. DAAG53-76C-0138 (ARPA Order 3206) to the University of Maryland Computer Vision Laboratory is gratefully acknowledged. The author deeply appreciates the assistance and encouragement of Bert Shaw and Azriel Rosenfeld.

REFERENCES

1. M. F. Hueckel, An operator which locates edges in digitized pictures, *J. Assoc. Comput. Mach.* **18**, 1971, 113–125.
2. M. F. Hueckel, A local operator which recognizes edges and lines, *J. Assoc. Comput. Mach.* **20**, 1973, 634–647.
3. A. Rosenfeld, R. Hummel, and S. Zucker, Scene labeling by relaxation operations, *IEEE Trans. Systems, Man, and Cybernetics*, **SMC-6**, 1976, 420–433.
4. G. Shaw, *Computer Graphics and Image Processing* **9**, in press.
5. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, New York, 1976.
6 J. L. Brown, Mean square error in series expansion of random functions, *SIAM Journal* 8, 1960, 28–32.
7. L. S. Davis, A survey of edge detection techniques, *Computer Graphics and Image Processing* **4**, 1975, 248–270.
8. A. K. Griffith, Mathematical models for automatic line detection, *J. Assoc. Comput. Mach.* **20**, 1973, 62–80.
9. C. K. Chow and T. Kaneko, Automatic boundary detection of the left ventricle from cineangiograms, *Comput. Biomed. Res.* **5**, 1972, 388–410.
10. L. Meró and Z. Vássy, A simplified and fast version of the Hueckel operator for finding optimal edges in pictures, *Proceedings, 4th International Conference on Artificial Intelligence, Tbilisi, USSR, Sept. 1975*, pp. 650–655.
11. J. R. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Addison–Wesley, Reading, Mass. 1974.
12. R. O. Duda, and P. E. Hart. *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.