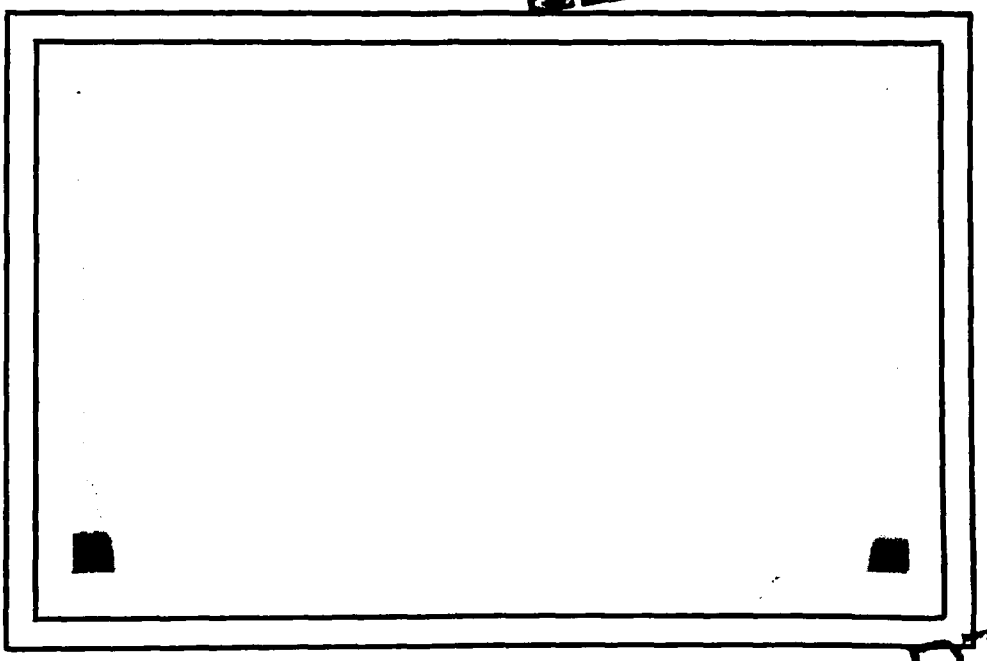


UNIVERSITY OF MARYLAND
COMPUTER SCIENCE CENTER

COLLEGE PARK, MARYLAND
20742

TR-860
DAAG-53-76C-0138
January 1980

FAST, HIERARCHICAL CORRELATIONS WITH GAUSSIAN-LIKE KERNELS

Peter J. Burt
Computer Vision Laboratory
Computer Science Center
University of Maryland
College Park, MD 20742

ABSTRACT

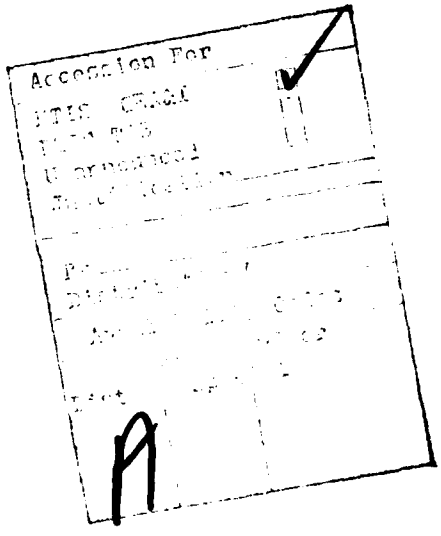
A hierarchical procedure is described for computing the discrete correlation, $h_\ell(x,y)$ o $f(x,y)$, between a function f and a kernel h_ℓ. For the class of h considered here, this correlation is equivalent to a weighted sum of correlations of f(x,y) with the kernel $h_{\ell-1} \equiv r^2 h_\ell(x/r, y/r)$. Here r>1, so $h_{\ell-1}$ is narrower than h_ℓ by 1/r. The correlations of f with $h_{\ell-1}$ are themselves computed as the weighted sums of correlations with still narrower kernels. The narrowest kernel, h_0, has unit width, and its correlation with f is f itself.

Kernels which can be computed hierarchically in this way closely approximate the Gaussian probability distribution. Hierarchical discrete correlation (HDC) is more efficient than direct correlation or correlation using the FFT, by two to three orders of magnitude.

The HDC has immediate applications to computer image processing. Samples of the correlations obtained at nearby image positions can be summed to obtain band pass Laplacian

41 ∠ 074

and oriented first and second derivative operators ("Mexican hat", edge and bar masks). The correlation of an image with many operators and at many scales requires little computation beyond a single HDC.



The support of the Defense Advanced Research Projects Agency and the U.S. Army Night Vision Laboratory under Contract DAAG-53-76C-0138 (DARPA Order 3206) is gratefully acknowledged, as is the help of Kathryn Riley in preparing this paper. The author also wishes to acknowledge discussions with Ted Adelson of New York University, whose ideas have added to several aspects of the work reported here.

1. Introduction

A fundamental and frequent task in computer image analysis is the computation of image correlations. Typically one image is correlated with a number of others in order to identify objects, or, in the case of stereopsis and motion analysis, to detect object displacements. Still more often relatively small operators, or masks, are correlated with larger images in order to extract local properties such as bar- and edge-like features and zero crossings. Unfortunately correlations are computationally expensive: many elementary computational steps (adds and multiplies) are required for each location in the image at which the correlation is computed. The expense is compounded in tasks such as texture or motion analysis where it is appropriate to process the image with operators of many sizes, some of which cover hundreds or even thousands of image pixels [1,2].

This paper describes a new method for computing correlations which is particularly well suited for image processing. The method, called hierarchical discrete correlation, or HDC, is computationally efficient, typically requiring two or three orders of magnitude fewer computational steps than direct correlation or correlation computed in the spatial frequency domain using the Fast Fourier transform. In addition the method simultaneously generates correlations for kernels (operators) of many sizes. These kernels closely approximate the Gaussian

probability distribution, so that the correlation is equivalent to low pass filtering. The operators commonly used in image processing can be readily obtained from sums and differences of Gaussian-like correlations at nearby image points.

The principle underlying the HDC is that the correlation of a function with a large kernel can be computed as a weighted sum of correlations with smaller kernels, and these in turn can be computed as weighted sums of correlations with still smaller kernels. The kernels at each iteration of the HDC computation differ in size by a factor r, which we call the order of the hierarchical correlation, but not in shape.

We begin by defining three types of HDC. Types 1 and 2 are for integer values of r, while type 3 is for fractional values. Type 2 differs from type 1 only in that an even rather than an odd number of correlations with small kernels are summed to obtain the correlation with the next larger kernel.

Definitions and analyses are initially given for HDC in one dimension, and then it is shown that these can be directly generalized to a second dimension. Next we show that bandpass Laplacian and first and second derivative image operators can be obtained from the HDC. Finally we compare the computational cost of the HDC with standard correlation and FFT methods.

2. Hierarchical Discrete Correlation, Type 1 (Odd)

Suppose f(x) is a single valued function of x which is known only at integer values of x. Suppose also that w(x) is a discrete weighting function which is defined at integral x and which is non-zero only for $-m \leq x \leq m$. We then define the hierarchical discrete correlation of type 1 (HDC1) as a set of correlation functions $g_\ell(x)$ which are obtained from f and w as follows:

$$\text{(Type 1)} \quad g_0(x) = f(x)$$
$$g_\ell(x) = \sum_{i=-m}^{m} w(i)\, g_{\ell-1}(x+ir^{\ell-1}) \quad \text{for } \ell>1 \qquad (1)$$

Here ℓ, m and r are positive integers.

The function g_ℓ is obtained from f through ℓ recursions of a correlation-like operation using the weighting function w(x). Thus we say that ℓ is the level of $g_\ell(x)$ in the HDC and w(x) is the generating kernel.

Note that $g_\ell(x)$ is defined as a sum of k=2m+1 values of $g_{\ell-1}(x)$ which are separated by multiples of the distance $r^{\ell-1}$. This sample distance grows geometrically by the factor r from level to level, so we say r is the order of the HDC; k is called the width of the generating kernel.

The HDC1 is illustrated graphically in Figure 1.

Equivalent Kernel

The function g_ℓ which is defined recursively in Eq. 1 may also be defined as a standard correlation of f with an

equivalent kernel h_ℓ:

$$g_\ell(x) = h_\ell(x) \circ f(x)$$
$$= \sum_{i=-M_\ell}^{M_\ell} h_\ell(i) f(x+i) \qquad (2)$$

The equivalent kernel may itself be defined recursively, and is independent of f:

$$h_0(x) = \begin{cases} 1 & \text{for } x=0 \\ 0 & \text{otherwise} \end{cases}$$

$$h_\ell(x) = \sum_{i=-m}^{m} w(i)\ h_{\ell-1}(x-ir^{\ell-1}) \qquad (3)$$

We can show by induction that the definition of g_ℓ given in Eq. 2 is equivalent to that given in Eq. 1. First observe that the limit of the sum in Eq. 2, M_ℓ, is just that interval over which h_ℓ is non-zero. To simplify our derivations we can extend the range of the sum without changing its value. (The actual value of M_ℓ will be obtained in the next subsection.)

For $\ell=0$ we have from Eqs. 2 and 3

$$g_0(x) = h_0(x) \circ f(x)$$
$$= \sum_{i=-M_0}^{M_0} h_0(i) f(x+i)$$
$$= \sum_{i=-\infty}^{\infty} h_0(i) f(x+i)$$
$$= f(x)$$

This result agrees with the first part of Eq. 1. Now suppose

that Eq. 2 is true for level ℓ-1. We need to show that it is also true for level ℓ.

We assume:

$$g_{\ell-1}(x) = h_{\ell-1}(x) \text{ o } f(x)$$

$$= \sum_{j=-\infty}^{\infty} h_{\ell-1}(j) f(x+j)$$

Now substitute this into Eq. 1:

$$g_{\ell}(x) = \sum_{i=-m}^{m} w(i) \left[\sum_{j=-\infty}^{\infty} h_{\ell-1}(j) f(x+ir^{\ell-1}+j) \right]$$

If we let $\hat{j}=ir^{\ell-1}+j$ we get:

$$g_{\ell}(x) = \sum_{i=-m}^{m} w(i) \left[\sum_{\hat{j}=-\infty}^{\infty} h_{\ell-1}(\hat{j}-ir^{\ell-1}) f(x+\hat{j}) \right]$$

Thus we reorder the summations:

$$g_{\ell}(x) = \sum_{\hat{j}=-\infty}^{\infty} \left[\sum_{i=-m}^{m} w(i) h_{\ell-1}(\hat{j}-ir^{\ell-1}) \right] f(x+\hat{j})$$

So from Eq. 3 we find

$$g_{\ell}(x) = \sum_{j=-\infty}^{\infty} h_{\ell}(\hat{j}) f(x+\hat{j})$$

$$= h_{\ell}(x) \text{ o } f(x)$$

This concludes the inductive proof.

While Eqs. 1 and 2 give equivalent definitions, it will be convenient to describe the correlations in terms of equivalent kernels, but compute them with the recursive formulation, Eq. 1.

Width of the equivalent kernel

Suppose M_ℓ is the largest value of x for which $h_\ell(x)$ is non-zero. From Eq. 3 we see that

$$M_\ell = mr^{\ell-1} + M_{\ell-1} \quad \text{for } \ell > 0$$

and $M_0 = 0$

Thus $M_\ell = mr^{\ell-1} + mr^{\ell-2} + \ldots + m$

$$= m \sum_{i=0}^{\ell-1} r^i$$

$$= m\frac{(r^\ell - 1)}{(r-1)} \tag{4}$$

The width k_ℓ of the equivalent kernel h_ℓ is $k_\ell = 2M_\ell + 1$.

Constraints

We now adopt four constraints on the generating kernel which are designed to insure that the equivalent kernels are monophasic, symmetric, and centered at x=0.

Normalization $\sum_{i=-m}^{m} w(i) = 1$ (C1)

Symmetry $w(x) = w(-x)$ for all x (C2)

Monophasic $w(x_1) \geq w(x_2) \geq 0$ for $0 \leq x_1 \leq x_2$ (C3)

Equal Contribution $\sum_{i=-m}^{m} w(j+ir) = \text{constant } (=1/r)$ for all j, $0 \leq j < r$ (C4)

The fourth constraint insures that every sample point of f(x) contributes with equal weight to every level of the HDC when we adopt a reduced form of correlation. (The reduced form will be defined below.)

The properties imposed by the first three constraints on

the generating kernel are transferred through it to the equivalent kernels as well. Thus for each ℓ, $h_\ell(x)$ is normalized, symmetric and monophasic.

Example Type 1 HDC

We now illustrate the hierarchical discrete correlation with an example which is well suited for image processing. We let the width of the generating kernel be 5 (m=2) and the order, r, be 2.

Let w(0)=a, w(1)=b, and w(2)=c. Then the constraints become

w(-1) = w(1) = b	(symmetry)
w(-2) = w(2) = c	
a+2b+2c = 1	(normalization)
$a \geq b \geq c \geq 0$	(monophasic)
a+2c = 2b	(equal contribution)

From these constraints we find

$1/4 \leq a \leq 1/2$

b = 1/4

c = 1/4-a/2

One free parameter remains and we take this to be a. For each value of a within the designated range we obtain a different generating kernel, and from it in turn a set of equivalent kernels. For example, when a = .4, b = .25 and c = .05, the kernels h_1, h_2, and h_3 are as shown in Figure 2. The scales have been adjusted in these graphs to aid comparison: in each case the amplitude of h_ℓ has been increased by 2^ℓ while the x scale has been reduced by the same factor. Thus the areas

under the curves have been preserved (area=1).

Notice that $h_1(x)=w(x)$, and that $h_2(x)$ and $h_3(x)$ are symmetric and monophasic. In addition these three equivalent kernels illustrate a critical property of HDC:

Observation 1 (Constant shape): The shape of the rescaled equivalent kernel converges towards a characteristic terminal form as the level, ℓ, is increased.

Shape convergence is very rapid. If we let h_∞ designate the characteristic form of the rescaled kernels then, in this example, h_3 deviates from h_∞ by less than 1%, while h_4 deviates less than 0.1%. We use h_5 as an estimate of h_∞ here (Fig. 2d) and in the remaining examples in this paper.

The shape of the characteristic kernel depends on the generating kernel and hence on the value assigned to the free parameter a. Examples for a=0.5, 0.4 and 0.3 are given in Figure 3.

When a=0.5, then c=0, so the width of the generating kernel becomes 3. In this case the equivalent kernel is triangular, as shown in Figure 3a. For other values of a, h_∞ assumes a "bell" shape, which is broader for smaller values of a. This illustrates a second critical property of the HDC:

Observation 2 (Gaussian-like): Equivalent kernels obtained when the generating kernel has width greater than 3 approximate the Gaussian probability distribution.

The closeness of this approximation depends on the value of a. For a=0.4 the approximation of h_∞ to a Gaussian is particularly good, as is shown graphically in Figure 4. As a is made smaller the width of h_∞ becomes greater. To characterize this tendency we have found the best fit (LSE) Gaussian to h_∞ as a function of a. The standard deviation, σ, of this Gaussian is shown as a function of a in Figure 5a. (σ is shown normalized by $1/r^{\ell}$, so does not change with level of the HDC.) The squared error between h_∞ and the best fit Gaussian is shown in Figure 5b. Note that the error is minimum for $a \cong 0.4$.

3. HDC Type 2 (Even)

The width of the generating kernel for type 1 hierarchical discrete correlations was always odd, k=2m+1. In order to permit kernels of even width we must change the definition of the correlation slightly:

(Type 2) $g_0(x) = f(x)$

$$g_\ell(x) = \sum_{i=-m}^{m-1} w(i + \tfrac{1}{2}) g_{\ell-1}(x+(i + \tfrac{1}{2})r^{\ell-1}) \qquad (5)$$

In this case $g_\ell(x)$ is defined for integer x when $(r^\ell-1)/(r-1)$ is even but for x = ...-1/2,1/2,3/2 ... when it is odd. The generating kernel is also defined at intermediate values of x.

To illustrate type 2 HDC's we consider an example in which r=2 and m=2. The structure of the computation is shown graphically in Figure 6. Let w(1/2)=a and w(3/2)=b. Then applying the constraints we find:

w(-1/2) = w(1/2) = a w(-3/2) = w(3/2) = b	symmetry
2a+2b = 1	normalization
a≥b≥0	monophasic

(the fourth constraint, equal contribution, is automatically satisfied).

Thus we find:

1/4≤a≤1/2

b = 1/2-a

Characteristic kernels are shown in Figure 7 for a = 0.5,

0.4, and 0.3. For a = 0.5, b = 0, and the generating kernel degenerates to width 2, and the characteristic kernel, h_∞, is square shaped (Fig. 7a). For smaller a h_∞ becomes bell shaped and Gaussian-like, although these kernels approximate the Gaussian function less closely than the kernels obtained in the previous example when the generating kernel width was 5. This point is made in Figure 8 where we plot σ and the mean squared error for the best-fit Gaussian to h_∞ as a function of a. The error is relatively small only over a narrow range of a near a≈.37.

4. Reduced HDC

We now introduce a modification to the hierarchical discrete correlation in which the number of samples of $g_\ell(x)$ is reduced by a factor of r from level to level. The arrangement of nodes in the graph which represents the correlation for a type 1 HDC, with r=2, is shown in Figure 9a. If we compare this to Figure 1 we see that every other node has been retained at level 1, every fourth node at level 2, and so on. However the correlations defined at the remaining graph nodes (sample points of $g_\ell(x)$) are not changed, nor is the generating kernel, w, or the equivalent kernels, h_ℓ. It will be convenient to change the index of the samples of $g_\ell(x)$ as follows: Let $g_{n,\ell}$ be the nth sample of $g_\ell(x)$ measured from $g_\ell(0)$ in the positive x direction. Then $g_{n,\ell} = g_\ell(nr^\ell)$, and Eq. 1 becomes:

(Type 1, reduced) $$g_{n,0} = f(n) \quad n = \ldots -1,0,1,2,\ldots$$

$$g_{n,\ell} = \sum_{i=-m}^{m} w(i) g_{rn+i,\ell-1} \qquad (6)$$

For Type 2 HDC let

$$g_{n,\ell} = g\left(\frac{1}{2}\frac{r^\ell - 1}{r-1} + nr^\ell\right).$$

(See Figure 9b.) Then Eq. 5 becomes:

(Type 2, reduced) $$g_{n,0} = f(n)$$

$$g_{n,\ell} = \sum_{i=-m}^{m-1} w(i+\tfrac{1}{2}) g_{rn+i+1,\ell-1} \qquad (7)$$

These reduced forms of the HDC have a distinct advantage over the original forms in that they are obtained with

substantially fewer computation steps and require less computer storage. Fortunately no information is lost in adopting the reduced sample density provided the sample interval ($=r^{\ell}$) satisfies the relationship

$$r^{\ell} \leq \frac{1}{2w_{\ell}} \tag{8}$$

where w_{ℓ} is the highest spatial frequency component of $g_{\ell}(x)$ [3, p. 70].

Let G_{ℓ}, H_{ℓ} and F be the Fourier transforms of g_{ℓ}, h_{ℓ} and f respectively. Then

$$G_{\ell}(s) = H_{\ell}(s)F^{*}(s)$$

(The '*' indicates the complex conjugate.) We have observed that $h_{\ell}(x)$ closely approximates a Gaussian distribution. Therefore H_{ℓ} can be reasonably approximated by the transform of the Gaussian. In particular, if the Gaussian has standard deviation σ_{ℓ}, then its transform is itself a Gaussian and has standard deviation $\sigma_s = \frac{1}{2\pi\sigma_{\ell}}$. As a fairly conservative estimate of the high frequency limit of H_{ℓ}, and hence also of G_{ℓ}, we say $w_{\ell} = 2\sigma_s$. Then Eq. 8 is satisfied when

$$\frac{\sigma_{\ell}}{r^{\ell}} \geq \frac{2}{\pi}$$

In fact this is approximately the ratio obtained when the generating kernel has width 4 or 5 and a is within the Gaussian-like region (see Fig. 5 and 8). We conclude that the correlation functions are adequately sampled in the reduced form of the HDC.

5. Type 3 HDC (Fractional Order)

An inherent restriction in hierarchical discrete correlations of types 1 and 2 is that the order, r, must be a positive integer. The smallest order of interest is r=2 since r=1 results in a process which is recursive but not hierarchical. In this case, the distance between level ℓ samples which are summed to obtain each level $\ell+1$ sample does not increase from level to level. It is possible to obtain fractional orders, with r between 1 and 2, in the following way. Suppose $r=k_1/k_2$ where k_1 and k_2 are integers and $k_2<k_1<2k_2$. We restrict our attention to the reduced form of the correlation so that for every k_1 samples of g_ℓ, there will be k_2 samples of $g_{\ell+1}$. These samples are computed at regular intervals (r^ℓ and $r^{\ell+1}$) so that the distance from $g_{\ell+1,i}$ to the nearest sample point of g_ℓ changes with i. However the same nearest neighbor distance is obtained when i is increased modulo k_2. To accommodate these k_2 distance relationships we must define k_2 different generating kernels $w_t(x)$, $t=0,\ldots,k_2-1$. If we position the samples in the x coordinate so that $g_{\ell,i}= g_\ell(i\cdot r^\ell+x_{\ell,0})$ where $x_{\ell,0}= \frac{1}{2}(r^\ell-1)/(r-1)$, then the distances between $g_{\ell+1,i}$ and the nearest g_ℓ sample will be the same as the distance between $g_{\ell+1,k_2-i-1}$ and the nearest g_ℓ sample, but in the opposite direction. In this case the t^{th} generating kernel is simply a mirror reflection of the (k_2-t-1)th kernel:

$$w_t(x) = w_{k_2-t-1}(-x).$$

We do not give the general definition of the type 3 HDC since indexing of the correlation functions and generating kernels is rather tedious. Instead we illustrate the construction with an example in which r=3/2. In this case there are two generating kernels, one of which is the mirror reflection of the other. We let the width of the generating kernels be 5, and say

$$w_0(x_1) = w_1(-x_1) = a$$
$$w_0(x_2) = w_1(-x_2) = b$$
$$w_0(x_3) = w_1(-x_3) = c$$
$$w_0(x_4) = w_1(-x_4) = d$$
$$w_0(x_5) = w_1(-x_5) = e$$

where

$$x_1 = -7/4$$
$$x_2 = -3/4$$
$$x_3 = 1/4$$
$$x_4 = 5/4$$
$$x_5 = 9/4$$

(See Figure 10.)

The constraints on w are the same as in previous HDC's except that the symmetry constraint is replaced by one or more <u>balance</u> constraints. Let $x_{t,i}$ be the position of the ith member of the kernel $w_t(x)$. Then the nth order balance constraint is

$$\sum_{i,\ \text{for } x_{t,i}<0} |x_{t,i}|^n w_t(x_{t,i}) = \sum_{i,\ \text{for } x_{t,i}>0} |x_{t,i}|^n w_t(x_{t,i})$$

In the present example we have

$a+b+c+d+e = 1$	normalization
$a+c+d = 2b+2e$	equal contribution
$c \geq b \geq d \geq a \geq e \geq 0$	monophasic
$7a+3b = c+5d+9e$	1st order balance
$49a+9b = c+25d+81e$	2nd order balance

From these we can determine the variables b, c, d, and e in terms of one free variable, a:

$$.021 \leq a \leq .127$$

$$b = \frac{17}{54} - \frac{a}{9}$$

$$c = \frac{23}{36} - \frac{8a}{3}$$

$$d = \frac{1}{36} + \frac{5a}{3}$$

$$e = \frac{1}{54} + \frac{a}{9}$$

The generating kernels obtained subject to these constraints are irregular in appearance (Fig. 11a) but the equivalent kernels are Gaussian-like and nearly symmetric (Fig. 11b). The standard deviations and errors of best fit Gaussians to h_∞ are given in Figure 12. Note that type 3 HDC with r = 3/2 and kernel width 5 can approximate Gaussians more closely than the type 1 and 2 examples described earlier (compare Figure 12 with Figures 5 and 8).

6. HDC in two dimensions

The three types of hierarchical discrete correlations which we have defined thus far in one dimension can be extended directly to a second dimension. Assume the function f(x,y) is sampled for x,y = ...-1,0,1,2,.... Then Eq. 6 for the reduced type 1 HDC becomes:

$$g_{n_x,n_y,\ell} = g_\ell(r^\ell n_x, r^\ell n_y)$$

$$g_{n_x,n_y,0} = f(n_x,n_y)$$

$$g_{n_x,n_y,\ell} = \sum_{i=-m}^{m} \sum_{j=-m}^{m} w(i,j) g_{rn_x+i,rn_y+j,\ell-1} \qquad (9)$$

The sample positions in two dimensions for types 1 and 2 HDC with r=2 and type 3 with 4=3/2 are given in Figure 13. (Other sample arrangements in two dimensions are shown in the appendix.)

Constraints on the generating kernel also extend in the natural way:

Normalization $\sum_{i=-m}^{m} \sum_{j=-m}^{m} w(i,j) = 1$

Symmetry (Type 1,2) $w(i,j) = w(-i,j) = w(i,-j) = w(-i,-j)$

Monophasic $w(i,j) \geq w(k,\ell)$ for $|i| \leq |k|$ and $|j| \leq |\ell|$

Equal Contribution $\sum_{i=-m}^{m} \sum_{j=-m}^{m} w(\hat{i}+ir,\hat{j}+jr) = \text{constant}(=1/r^2)$ for $0 \leq \hat{i},\hat{j} < r$

An important special case is that in which the generating kernel is separable:

$$w(i,j) = w_x(i) \cdot w_y(j)$$

Here the generating kernel w satisfies the two-dimensional constraints just when the one dimensional kernels, w_x and w_y, satisfy the one-dimensional constraints. Furthermore, when w is separable then the equivalent kernels are also separable:

$$h_\ell(x,y) = h_{\ell x}(x)h_{\ell y}(y)$$

where $h_{\ell x}$ and $h_{\ell y}$ are the equivalent kernels obtained in one dimension with the generating kernels w_x and w_y. Thus the data summarized in Figures 5, 8, and 12 can be applied to two dimensions when the kernel is separable.

Suppose $w_x = w_y$ and the equivalent kernel $h_{\ell x}(=h_{\ell y})$ is approximately Gaussian. Then $h_\ell(x,y)$ will approximate a two-dimensional Gaussian function and we may anticipate that $h_\ell(x,y)$ will be nearly circularly symmetric. This is confirmed in the type 1 HDC shown in Figure 14a, for which r=2, m=2, and a=.4. On the other hand, if $w_x \neq w_y$ we obtain an equivalent kernel which is elongated, such as that in Figure 14b. Again r=2 and m=2, but in this case w_x and w_y are determined by parameters a_x=.45 and a_y=.25.

7. HDC for Image Processing

We now show how the HDC can be used to construct the local operators commonly used in image processing.

Bandlimited Laplacian, L

Laplacian-like operators have long been defined as 3x3 weighting functions in which the central weight is positive (say 8) while the surrounding weights are negative (say -1). Often, however, it is appropriate to convolve operators of various sizes with an image. Such operators are said to be band-limited since they respond to details of the image which contain a limited range of spatial frequencies. The band-limited Laplacian may be formally defined as the Laplacian of a Gaussian, $\nabla^2 G$, but the result is generally approximated as the difference between two Gaussian functions which have different standard deviations:

$$L(x,y) = \frac{1}{2\pi} \frac{e^{\frac{-(x^2+y^2)}{2\sigma_1^2}}}{\sigma_1^2} - \frac{1}{2\pi} \frac{e^{\frac{-(x^2+y^2)}{2\sigma_2^2}}}{\sigma_2^2}$$

Typically the ratio σ_2/σ_1 is in the range 1.5 to 2.5.

A band-limited Laplacian with a ratio $\sigma_2/\sigma_1=r$ may be obtained from the type 1 HDC of order r simply by subtracting $g_{\ell+1}(x)$ from $g_\ell(x)$. (If the reduced form of the HDC is used, samples of $g_{\ell+1}(x)$ are not computed for every sample of $g_\ell(x)$. Missing $g_{\ell+1}$ are obtained by applying w(x,y) to the neighborhood of $g_\ell(x)$.)

To obtain ratios $\sigma_2/\sigma_1<r$ two HDC (any type) are computed with different generating kernels. These are determined from graphs such as those in Figures 5, 8, and 12 to obtain the desired σ. The L operator is then simply a difference between corresponding samples in the two HDC.

A third procedure yields ratios $\sigma_2/\sigma_1>r$. First an HDC is obtained (any type) to form the central Gaussian. Then the surround Gaussian is obtained by applying the generating kernel in reverse. For type 1 we have

$$L_{n_x,n_y,\ell} = g_{n_x,n_y,\ell} - r^2 \sum_{i=-m}^{m} \sum_{j=-m}^{m} w(i,j) g_{\frac{n_x-i}{r}, \frac{n_y-j}{r}, \ell+1}$$

The sums in this expression are understood to include only those terms for which $\frac{n_x-i}{r}$ and $\frac{n_y-j}{r}$ are integer valued. The equal contribution constraint on w ensures that the sum will have a total weight of $1/r^2$, hence the r^2 factor normalizes the sum in the above definition. An example using this procedure is shown in Figure 15. Here a ratio $\sigma_2/\sigma_1=2.5$ is obtained with a Type 1 HDC, order r=2, width k=2m+1=5, and with a separable generating kernel for which a=.4.

Bandlimited Derivatives, D1 and D2

To obtain band-limited operators for first and second derivatives in the x direction we simply compute the difference of $g_\ell(x)$ samples which are separated by the distance r^ℓ in the x direction. Using the reduced form

$$D1_{i,j,\ell} = g_{i,j,\ell} + g_{i,j-1,\ell} - g_{i-1,j,\ell} - g_{i-1,j-1,\ell}$$

and

$$D2_{i,j,\ell} = \sum_{k=-1}^{1} g_{i,j+k,\ell} - 1/2(g_{i-1,j+k,\ell} + g_{i+1,j+k,\ell})$$

Contour maps for these operators are shown in Figures 16 and 17. Again, these are computed from a type 1 HDC with r=2, k=2m+1=5, and a separable w, with a=.4.

8. Computational Efficiency

We have already encountered several potential advantages that the HDC has over other methods of computing correlations with Gaussian-like kernels. These include the fact that correlations are simultaneously obtained for a number of kernels which differ in size, and the fact that only a small generating kernel need be specified (or stored in computer memory) even to compute correlations with very large equivalent kernels. In addition the structure of the HDC provides a convenient mechanism for adjusting the sample interval to the scale of the equivalent kernel. We will now show that the HDC is also computationally more efficient than either direct correlation or correlations computed in the frequency domain using the Fast Fourier transform (FFT).

The computational expense of the HDC relative to other methods depends in part on whether only the correlation at one level is to be used, or all correlations up to and including that level are to be used, and on whether one generates the full or reduced form of the HDC. We consider first the most costly case, in which a full HDC is computed to level ℓ, and only the correlation at level ℓ is of value. We wish to determine N_{HCD}, the number of elementary operations (additions, multiplications) per sample of $g_\ell(x)$. Suppose the generating kernel has width k, and the order is r. There will be k^2 elementary operations from level 1 to ℓ for each sample. Thus

$$N_{HCD} = \ell \cdot k^2$$

This number should be compared to N_{COR}, the number of operations per sample using a standard correlation with the equivalent kernel h_ℓ. The width of h_ℓ is $k_\ell = 2M_\ell+1$, where M_ℓ is given in Eq. 4. Then

$$N_{COR} = k_\ell^2 = k^2 r^{2\ell}$$

(Here we assume $r \gtrsim 2$.) The relative efficiency of the hierarchical to standard correlations is given by the ratio

$$\frac{N_{HDC}}{N_{COR}} \cong \frac{\ell}{r^{2\ell}}$$

Typical values of r and ℓ are 2 and 6 respectively. In this instance the standard computation requires roughly 650 times as many elementary operations as the HDC.

Next we compare N_{HDC} to N_{FFT}, the per location cost of computing a single level ℓ correlation using the FFT. A one-dimensional FFT requires $N \log_2 N$ elementary operations, where N is the number of sample elements over which the transform is computed. The standard method for computing a two-dimensional FFT requires N^2 one-dimensional transforms, or one per element of the transformed array:

$$N_{FFT} = N \log_2 N$$

Notice that the number of computations per location depends on the array dimensions in the FFT but not in the HDC. The relative cost is given by

$$\frac{N_{HCD}}{N_{FFT}} = \frac{\ell\, k^2}{N \log_2 N}$$

As an example assume k=5, ℓ=6, and N=512. Then the FFT will require more elementary operations by a factor of about 40.

Further reduction in the cost of the HDC relative to the other methods is realized when one wishes to obtain correlations for kernels at a number of scales. No additional computational steps are required for the HDC, while a separate computation is required for each scale with the other methods.

The cost of the correlation can be cut by an additional factor of almost ℓ when the reduced form of the HDC is used. As we have seen, essentially no information is lost by using the reduced form.

9. Summary and Discussion

We have shown that correlation of a function f(x,y) with certain kernels can be computed hierarchically: the correlation with a large kernel, $h_\ell(x,y)$, is a weighted sum of correlations with a smaller kernel $h_{\ell-1}(x,y)$, which in turn are weighted sums of correlations with still smaller kernels. The kernels in this sequence differ in scale by factors of r but not in shape, so that $h_{\ell-1}(x,y) \cong r^2 h_\ell(x/r,y/r)$. The parameter r indicates the order of the hierarchy: variations on the hierarchical discrete correlation (HDC) have been defined for fractional r between 1 and 2, and for integer r of 1 or greater.

The members of the class of kernels which can be computed hierarchically closely resemble the Gaussian probability distribution. This means that correlation is equivalent to low pass filtering. Also the two-dimensional kernel is very nearly circularly symmetric.

The principal advantage of the HDC method is that it is computationally more efficient than the direct correlation and FFT methods. In addition, correlations for a set of scaled kernels are computed at once, without any need to construct and store large kernels or kernels of different shapes and sizes.

These properties make the HDC particularly well suited for computer image analysis. The correlations of an image with sets of operators which differ in size and shape can be obtained from a single HDC. These include the bandpass Laplacian

and oriented first and second derivative operators.

Several other correlation methods which have been developed for image processing may now be regarded as special cases of the HDC. Rosenfeld and Thurston [1] describe an elegant method for computing averages within square regions which differ in scale by powers of 2. This is equivalent to the type 2 HDC with order r=2 and generating kernel width k=2 (see Fig. 7a). Pratt et al. [4] also obtain convolutions with large kernels by repeated convolution with small kernels. (Convolutions and correlations are equivalent under a reflection of the kernel.) This method is the same as an HDC in which the order r=1. The result is not hierarchical in the sense we describe here since the equivalent kernels do not increase geometrically in scale from level to level (or even arithmetically), nor do the kernels obtained at each iteration agree in shape.

The reduced form of the type 2 HDC with r=2 and generating kernel width 2 is equivalent to the pyramid structures proposed by Tanimoto and Pavlidis [5] and others. The cone structure of Hanson and Riseman [6] is like a reduced Type 1 HDC in which the generating kernel width is 3 or 5, and the order is 2. However, no constraints are imposed on the generating kernel in the cone so equivalent kernels are not Gaussian-like.

Finally, the reduced HDC, or Gaussian-weighted pyramid, defined here overcomes an awkward property of traditional pyramids, that the information represented at each level changes with image position. In the Gaussian-weighted pyramid values at each

level represent discrete samples of the continuous correlation function. We have shown that this function is band-limited and that the sample interval is sufficiently small so that no information is lost in the sampling process. This is true regardless of image position.

Appendix

Type 4 HDC (hexagonal sampling and rotation)

There is considerably greater flexibility in the structure of the HDC in two dimensions than in one. Two extensions will be described here: the use of hexagonally arranged samples, and the possibility of rotating the axes of the generating kernel from level to level of the HDC. We shall only consider cases in which there is a single generating kernel (unlike type 3 HDC), and shall only illustrate the computational structure graphically by showing the arrangement of sample and generating kernel nodes at two successive levels. The same constraints apply as with previous HDC: generating kernels should be normalized, symmetric, monophasic, and provide for equal contribution.

The permissible orders of the HDC are determined uniquely by the sampling patterns. In particular, r is a possible order only if r^2 is an integer and r is the distance between some pair of points in the sampling grid [7]. Permissible orders for the hexagonal grid are $r^2 = 3,4,7,...$ and for the rectangular grid $r^2 = 2,4,5,....$ Examples with $r^2 = 3$ and 4 (hexagonal) and $r^2 = 2$ and 5 (rectangular) are shown in Figure 18. In these diagrams small dots show the positions of sample points at level ℓ while open circles show sample points at level $\ell+1$. The generating kernels are centered at level $\ell+1$ nodes, and the boundary of the smallest such kernel is shown by line segments. Extension to larger kernels is simply a matter of expanding

these bounds; the arrangement of samples is unchanged. Notice that in all cases shown except Figure 18a, the grid of level ℓ+1 nodes is rotated with respect to level ℓ. Rotations are unavoidable with these orders.

References

1. A. Rosenfeld and M. Thurston, Edge and curve detection for visual scene analysis, IEEE Trans. on Computers C-20, 1971, 562-569.

2. D. Marr and E. Hildreth, Theory of edge detection, AI Memo 518, AI Lab, MIT, Cambridge, MA, April 1979.

3. R. C. Gonzalez and P. Wintz, Digital Image Processing, Addison-Wesley, Reading, MA, 1977.

4. W. K. Pratt, J. F. Abramatic and S. U. Lee, Two-dimensional small generating kernel convolution, in USCIPI Report 860, Image Processing Institute, Univ. of Southern California, 1979.

5. S. L. Tanimoto and T. Pavlidis, A hierarchical data structure for picture processing, Computer Graphics and Image Processing 4, 1975, 104-119.

6. A. Hanson and E. Riseman, Segmentation of natural scenes, in Computer Vision Systems, A. Hanson and E. Riseman (Eds.), Academic Press, NY, 1978, 129-163.

7. P. J. Burt, Tree and pyramid structures for coding hexagonally sampled binary images, TR-814, Computer Science Center, University of Maryland, College Park, MD, 1979.

FIGURE CAPTIONS

Figure 1. Graph representation of a type 1 hierarchical discrete correlation. Here nodes represent sample points of the correlation function $g_\ell(x)$. The horizontal position of the nodes indicates spatial location, x, while the vertical position indicates level, ℓ. The generating kernel is shown as a pattern of arrows between successive levels: sample values at level ℓ are weighted by a, b, c and summed to obtain the value of a single sample at level $\ell+1$. The same weighting pattern is applied at each x position to compute all $g_{\ell+1}(x)$ samples from $g_\ell(x)$ samples. Note that the distance between the sample points which contribute to a sum increases as r^ℓ, where the order, r, has value 2 in this example.

Figure 2. Equivalent kernels, $h_\ell(x)$, for type 1 HDC. In this example the generating kernel has width 5, order r=2, and weight parameter a=0.4. In general the width of the equivalent kernel increases by a factor of r from level to level, while the amplitude decreases by the same factor. We have compensated for this effect in the graphs by contracting the x scale and expanding the y scale by r^ℓ. Note that $h_1(x)$ is identical to the generating kernel, and that as ℓ is increased, the shape of $h_\ell(x)$ converges rapidly to a continuous function which is similar to the Gaussian probability distribution.

Figure 3. Characteristic forms, $h_\infty(x)$, of the rescaled equivalent kernels for three values of the weight parameter, a. Examples are for a type 1 HDC in which the generating kernel has width 5 and r=2.

Figure 4. Comparison of the equivalent kernel (solid curve) with the Gaussian probability distribution (dashed curve). Here a=0.4, and the Gaussian which most closely approximates the equivalent kernel has standard deviation, $\sigma=0.56r^\ell$. This is the most Gaussian-like equivalent kernel that can be obtained with a type 1 HDC when the generating kernel width is 5 and r=2 (see Fig. 5).

Figure 5. Shape characteristics of type 1 equivalent kernels. In general the equivalent kernel becomes broader as the weight parameter, a, is decreased. This effect is shown in Fig. 5a as an increase in the standard deviation, σ, of the Gaussian function which most closely approximates the equivalent kernel (least squared error). The squared error between this Gaussian and the equivalent kernel is shown in Fig. 5b. Note that the error is minimum, and the equivalent kernel most Gaussian-like, for $a\tilde{=}0.4$. (The error is normalized by rescaling the equivalent kernel so that the best fit Gaussian has unit standard deviation.) Here the generating kernel has width 5 and r=2.

Figure 6. Graph representation of a type 2 HDC. See Fig. 5 for explanation and comparison to type 1 HDC.

Figure 7. Characteristic forms for type 2 HDC. Forms are shown for three values of the weighting parameter, a, when the generating kernel has width 4 and the order r=2. (Compare with Fig. 3.)

Figure 8. Shape characteristics of type 2 equivalent kernels. The standard deviation of the best fit Gaussian is shown in (a) while its squared error is shown in (b). (Compare with type 1, Fig. 5.) Here the generating kernel has width 4 and r=2.

Figure 9a. Reduced type 1 HDC. The reduced form of the HDC is like the standard form (Fig. 1) except that only nodes spaced by r^{ℓ} are included at level ℓ. Computations at these points are identical in both forms.

Figure 9b. Reduced type 2 HDC. Compare with the standard form, Fig. 6.

Figure 10. Type 3 HDC with order r=3/2. This graph illustrates the construction of a fractional order hierarchical correlation. Here there are 2 nodes at level ℓ for every 3 at level ℓ-1, and there are two generating kernels, one of which is the left-to-right reflection of the other.

Figure 11. Generating and equivalent kernels for order r=3/2, type 3 HDC. The generating kernel in Fig. 11a is one of two required for this HDC, which are left-to-right reflections of one another (see Fig. 10). The characteristic form of the equivalent kernel is shown as a solid curve in Fig. 11b while the best fit Gaussian function is shown as a dashed curve. In this example the weighting parameter a=0.085.

Figure 12. Shape characteristics of type 3 equivalent kernels. The standard deviation of the best fit Gaussian is shown in Fig. 12a while its squared error is shown in Fig. 12b. (Compare with type 1, Fig. 5, and type 2, Fig. 8.) Here the generating kernel has width 5 and r=3/2.

Figure 13. Node positions in two dimensions. Dots represent level ℓ nodes while open circles indicate the relative position of level ℓ+1 nodes.

Figure 14a. Contour map for a type 1 two-dimensional equivalent kernel. This example was obtained with a separable generating kernel of width 5 and equal weighting patterns in both X and Y directions, $a_x = a_y = 0.4$. (The magnitude has been scaled so that its maximum is 1.)

Figure 14b. Contour map for an elongated type 1 equivalent kernel. As in Fig. 14a, this example was obtained with a separable generating kernel, but here different weighting patterns were applied in the X and Y directions, $a_x = 0.45$ and $a_y = 0.25$.

Figure 15. Band limited Laplacian operator. Laplacian-like operators are obtained as the difference of Gaussian-like operators computed at successive levels of the HDC. This example is a type 1 HDC with width 5 generating kernel, r=2 and a=0.4.

Figure 16. Lowpass first derivative operator. This operator is obtained as the weighted sum of four neighboring samples in a type 1 HDC, as explained in the text. The upper figure shows a contour map and the lower figure a horizontal cross-section of the operator.

Figure 17. Band limited second derivative operator. This operator is the weighted sum of nine neighboring samples of the type 1 HDC. The upper figure shows a contour map and the lower figure a horizontal cross section of the operator.

Figure 18a,b. Examples of low order hexagonal sample patterns for type 4 HDC. In each case closed circles represent level ℓ sample points, while open circles represent level $\ell+1$ sample points. Lines indicate the bounds of the smallest allowed generating kernels. One such kernel is shown with heavy lines and the distribution and values of its weights are indicated.

Figure 18c,d. Fractional order rectangular sample patterns for type 4 HDC. (Compare to Fig. 13.)

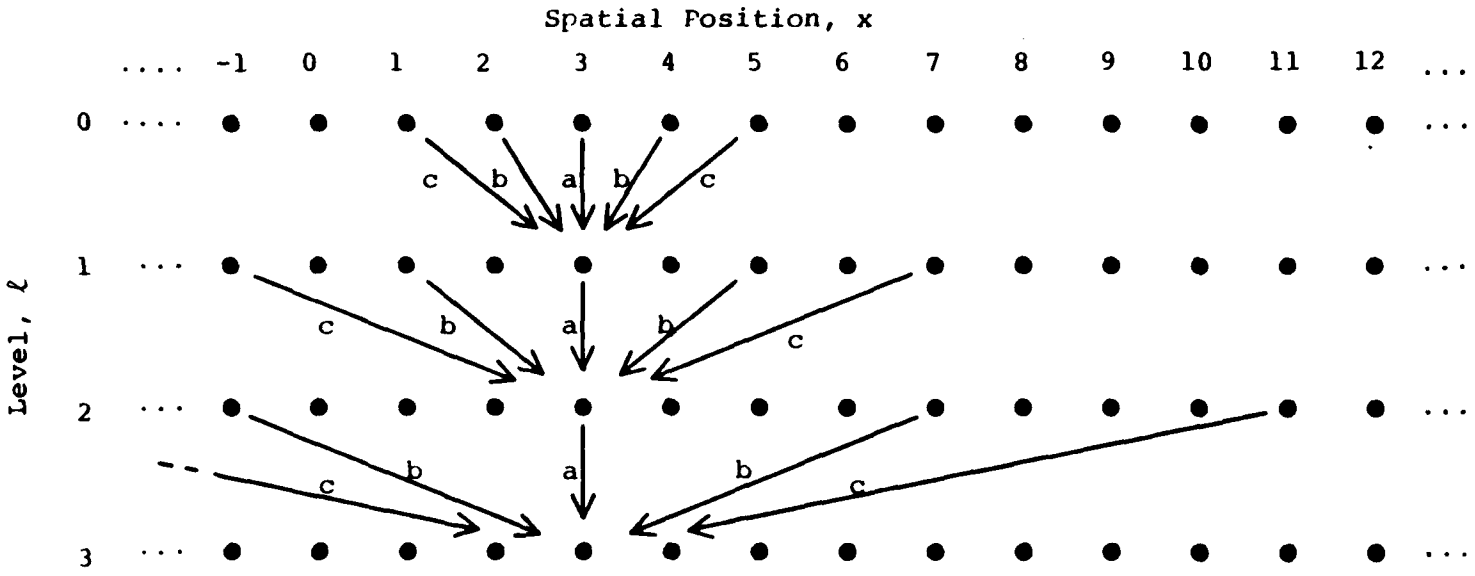

Spatial Position, x
.... -1 0 1 2 3 4 5 6 7 8 9 10 11 12 ...
Level, ℓ
0
1
2
3
c b a b c
c b a b c
c b a b c

Figure 1.

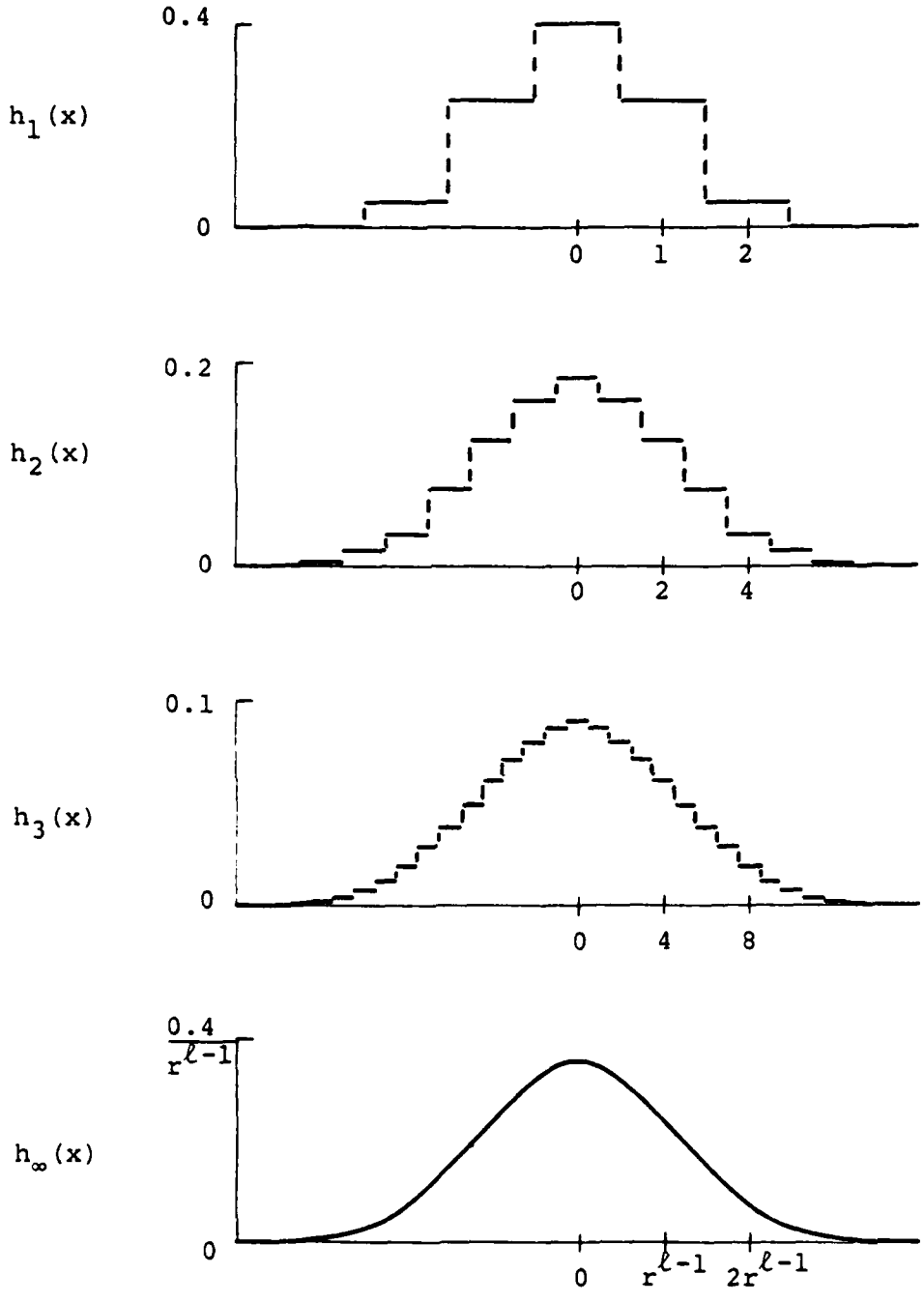

$h_1(x)$
0.4
0
0 1 2
$h_2(x)$
0.2
0
0 2 4
$h_3(x)$
0.1
0
0 4 8
$h_\infty(x)$
$\frac{0.4}{r^{\ell-1}}$
0
0 $r^{\ell-1}$ $2r^{\ell-1}$
Spatial Position, x

Figure 2.

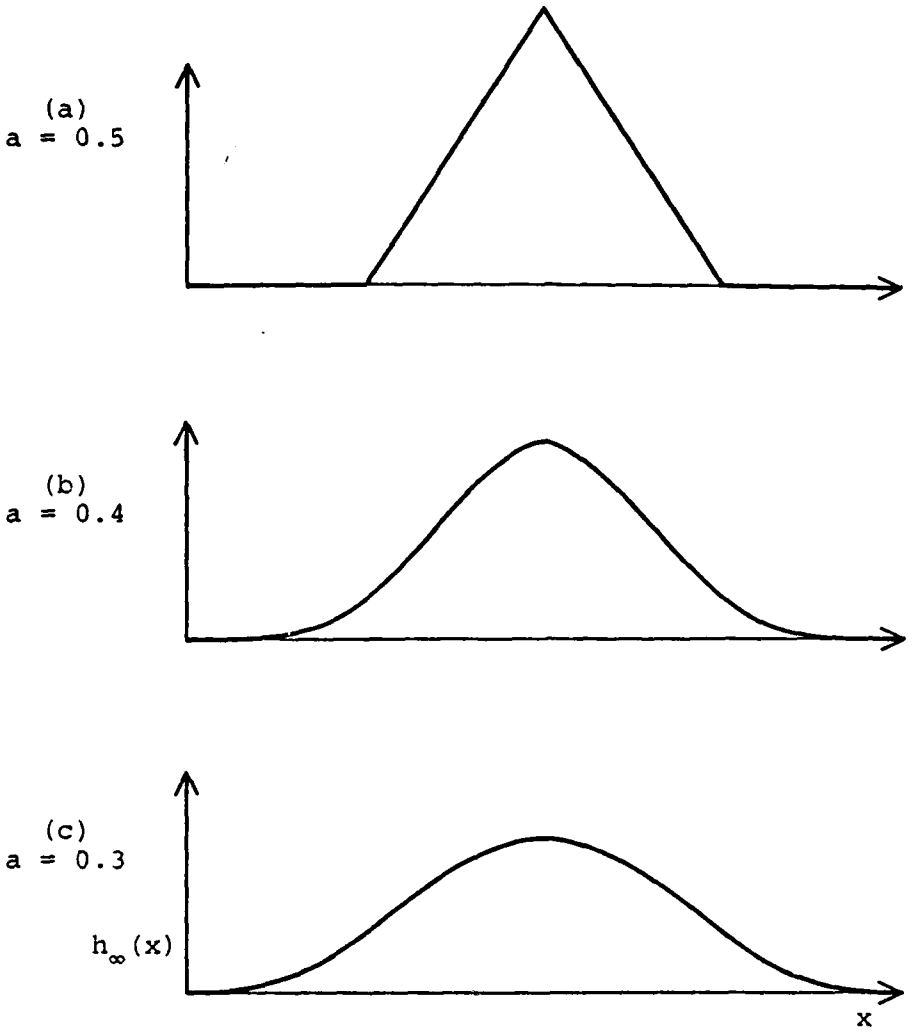


Figure 3.

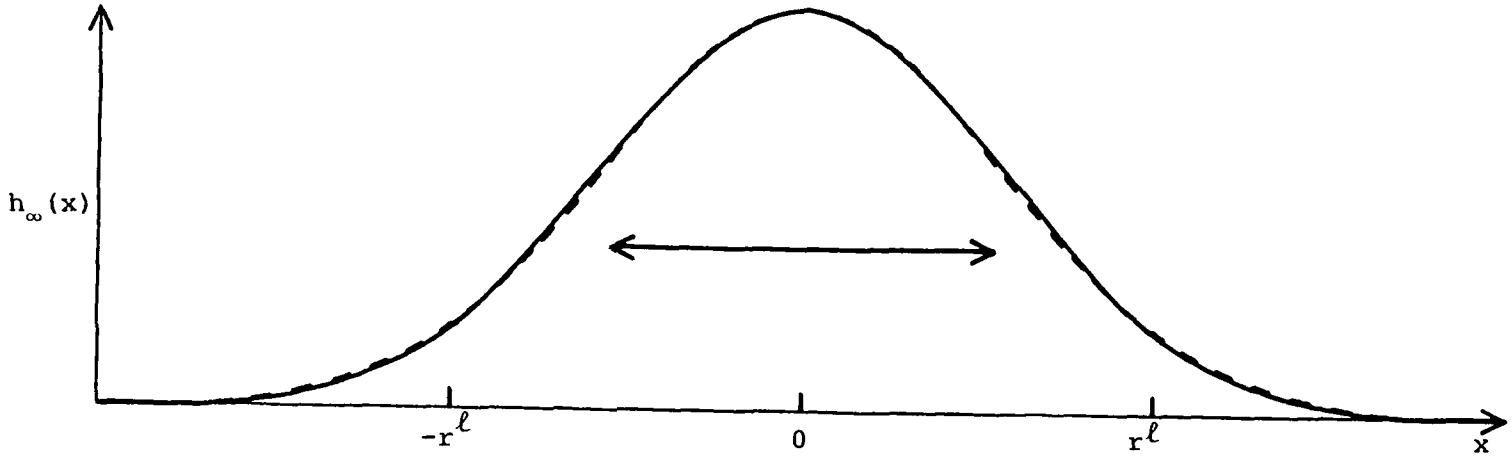


Figure 4.

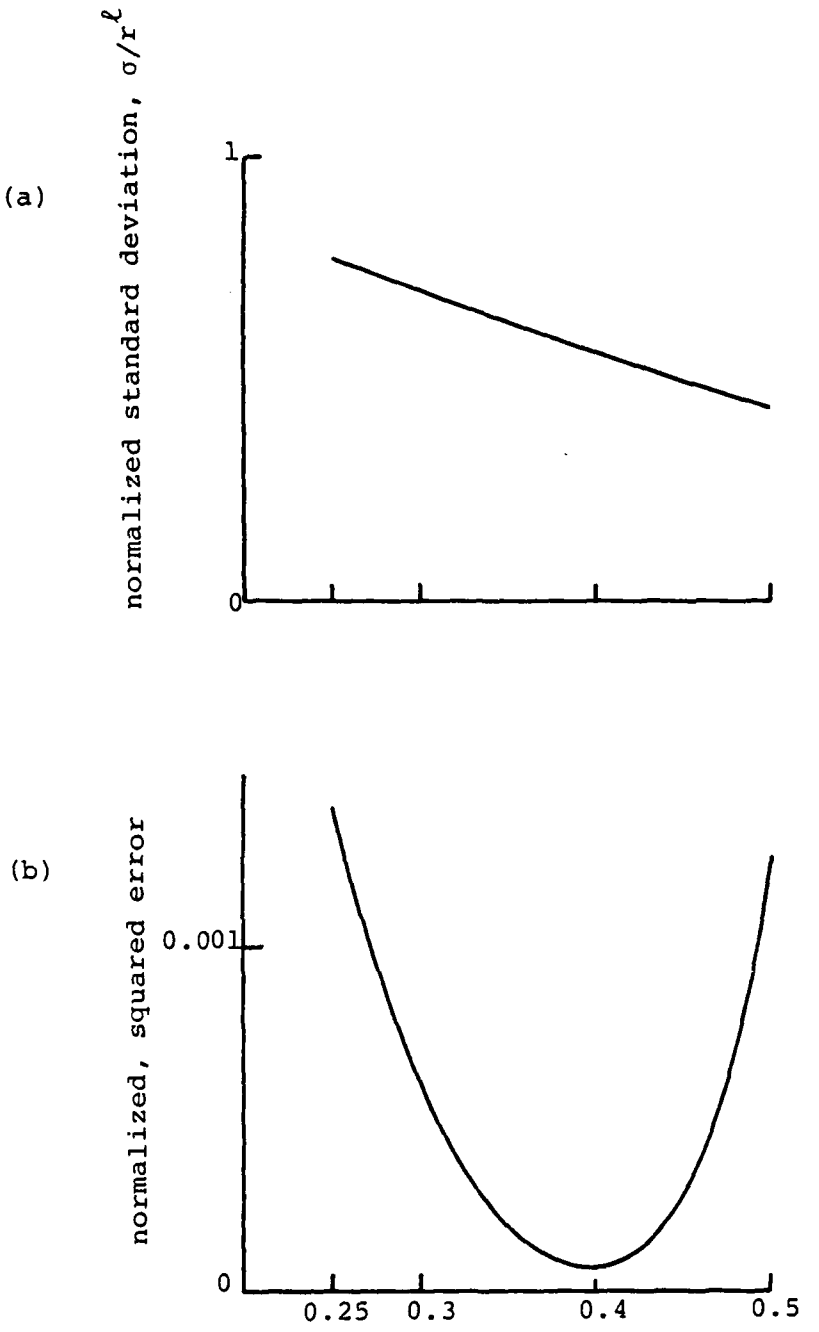

(a)
normalized standard deviation, σ/r^{ℓ}
1
0
(b)
normalized, squared error
0.001
0
0.25
0.3
0.4
0.5
weight parameter, a

Figure 5.

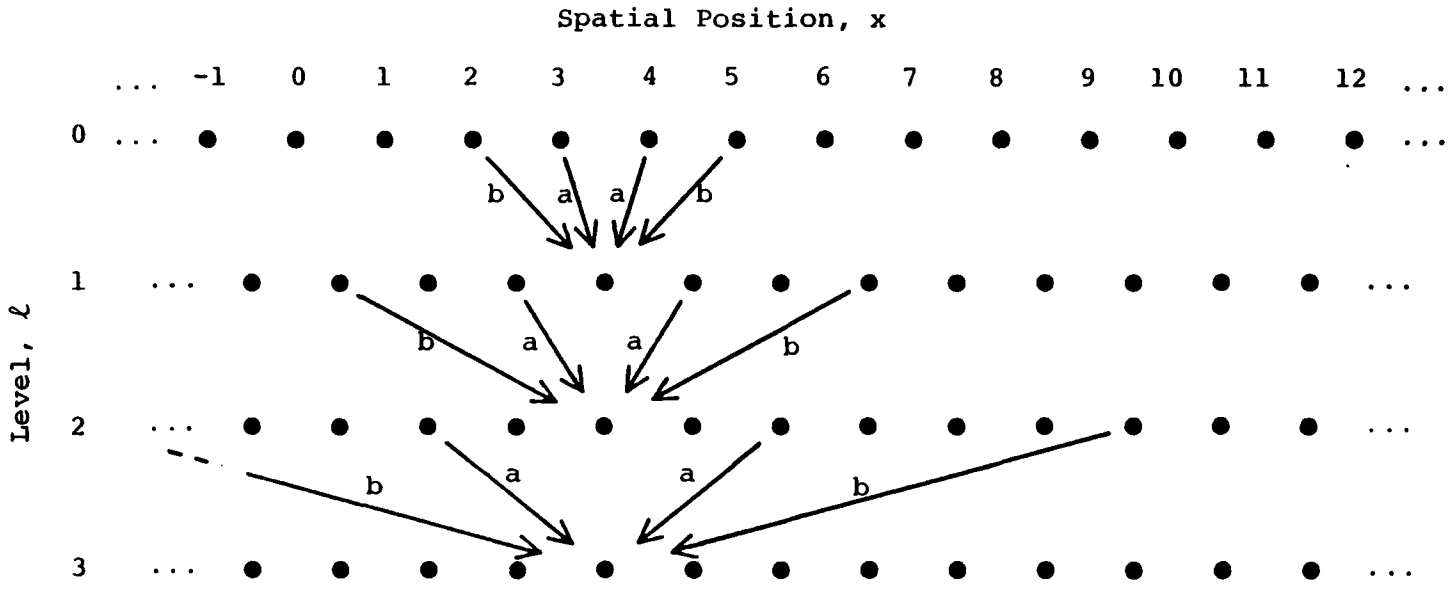

Spatial Position, x
... -1 0 1 2 3 4 5 6 7 8 9 10 11 12 ...
0
1
2
3
Level, ℓ
b a a b
b a a b
b a a b

Figure 6.

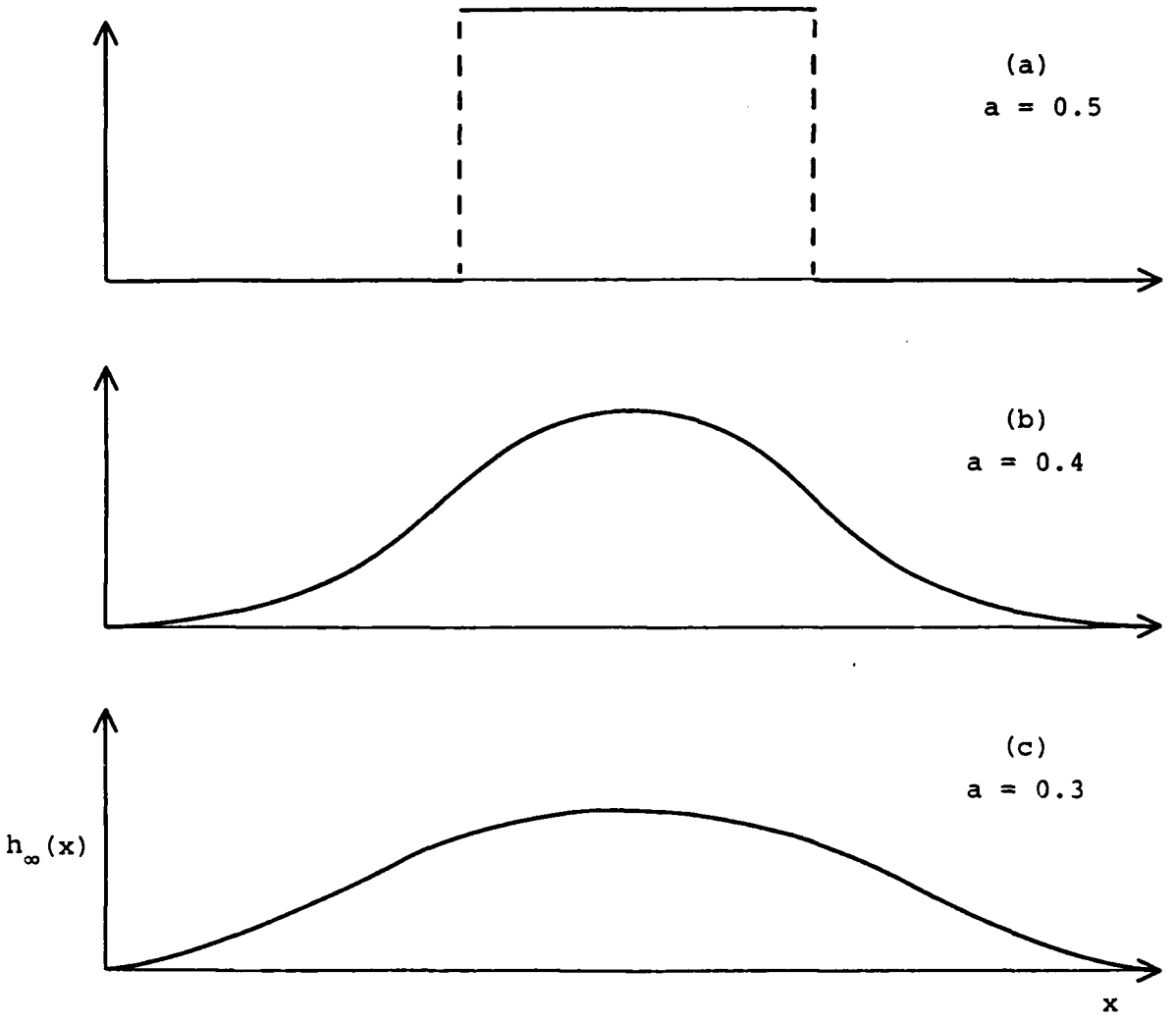


Figure 7.

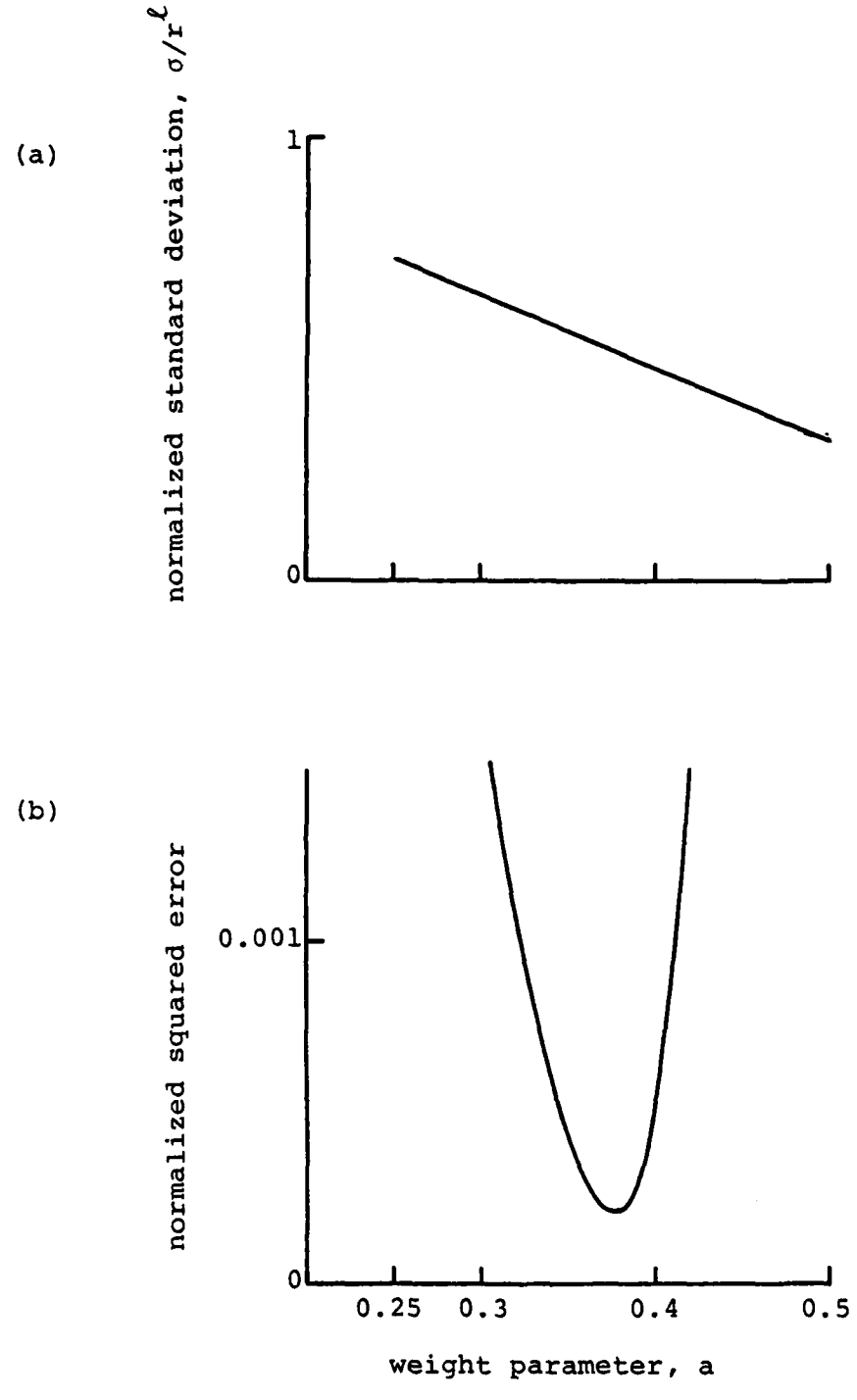

(a)
normalized standard deviation, σ/r^{ℓ}
1
0
(b)
normalized squared error
0.001
0
0.25
0.3
0.4
0.5
weight parameter, a

Figure 8.

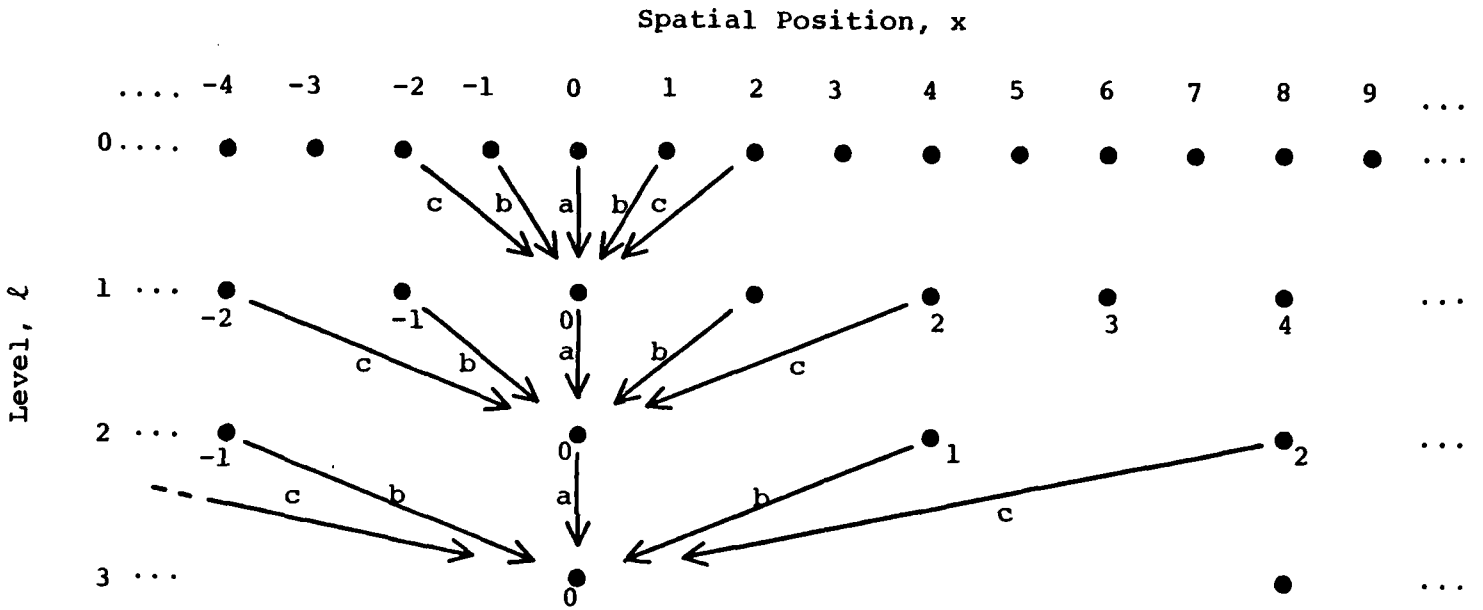

Spatial Position, x
.... -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 ...
0 ....
c b a b c
1 ...
-2 -1 0 2 3 4
a
2 ...
-1 0 1 2
a
3 ...
0
Level, ℓ

Figure 9a.

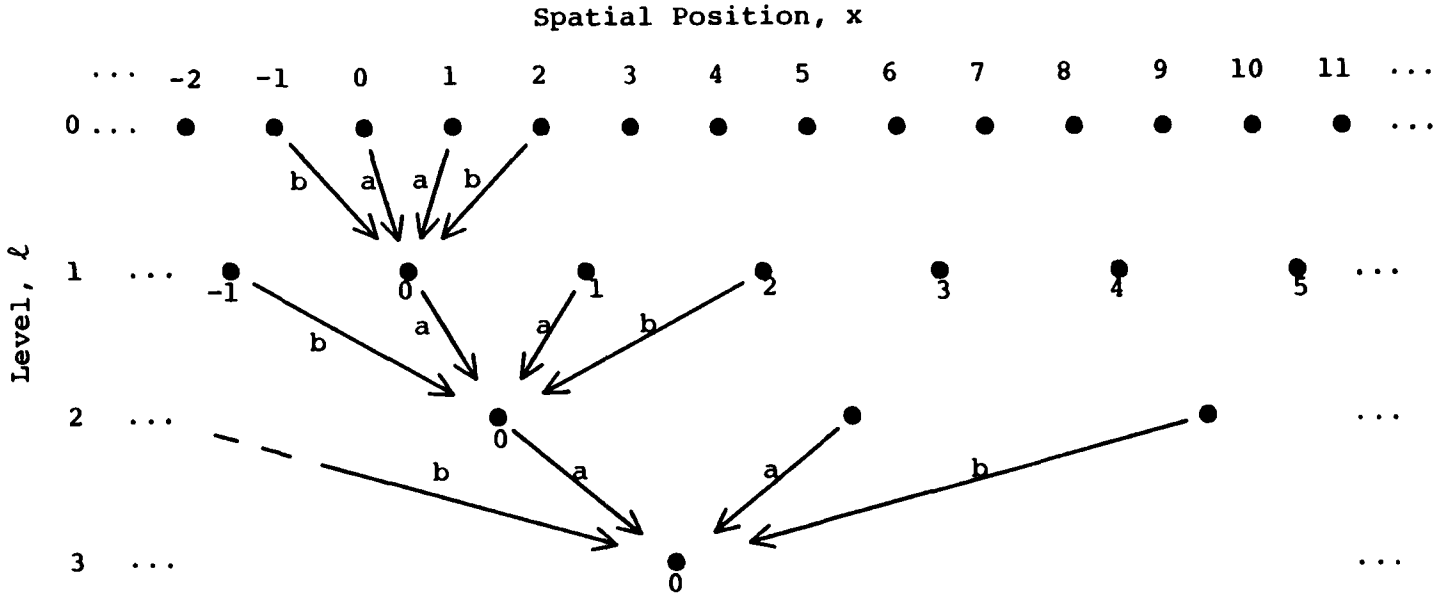

Spatial Position, x
... -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 ...
0 ... ...
b a a b
Level, ℓ
1 ... -1 0 1 2 3 4 5 ...
b a a b
2 ... 0 ...
b a a b
3 ... 0 ...

Figure 9b.

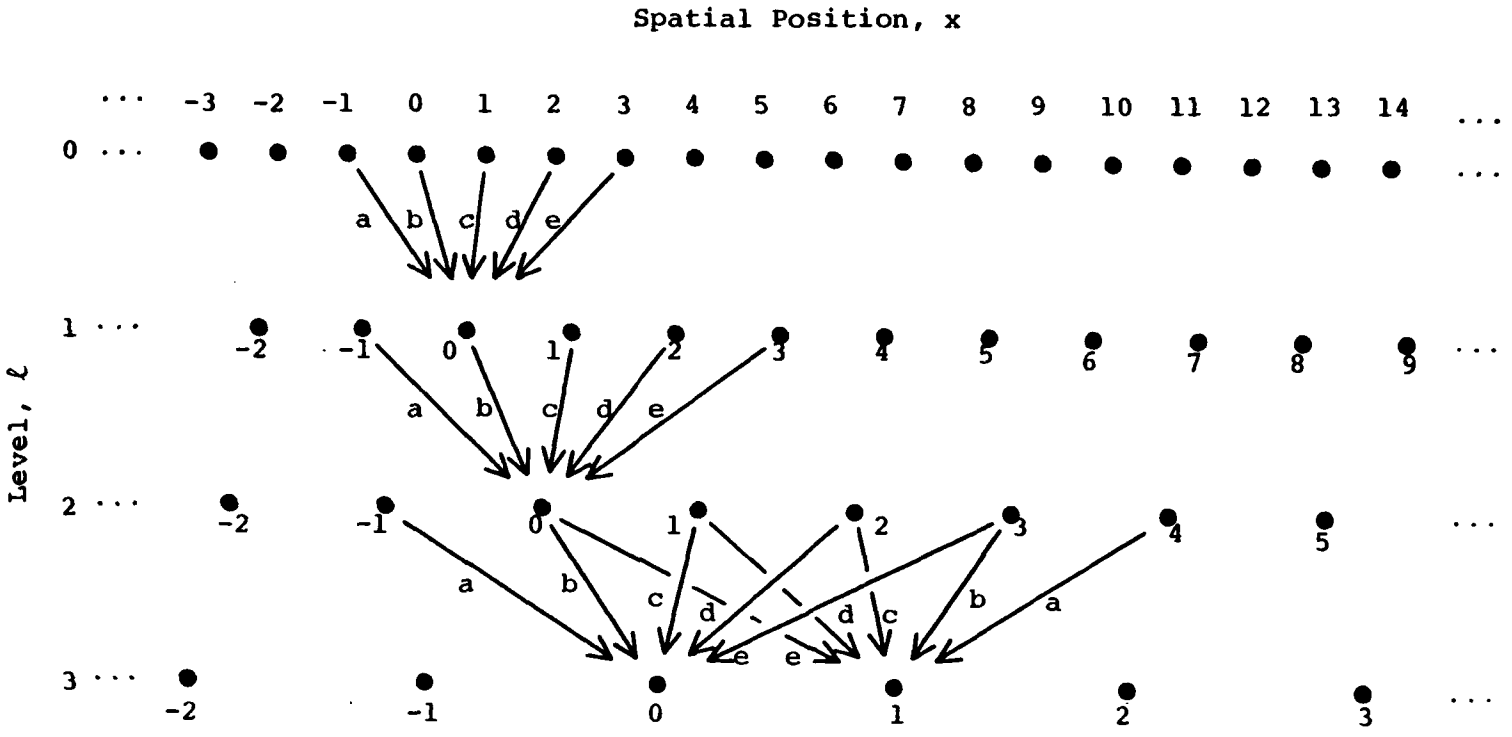

Spatial Position, x
Level, ℓ
··· -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 ···
0 ···
a b c d e
1 ···
-2 -1 0 1 2 3 4 5 6 7 8 9 ···
a b c d e
2 ···
-2 -1 0 1 2 3 4 5 ···
a b c d e d c b a
3 ···
-2 -1 0 1 2 3 ···

Figure 10.

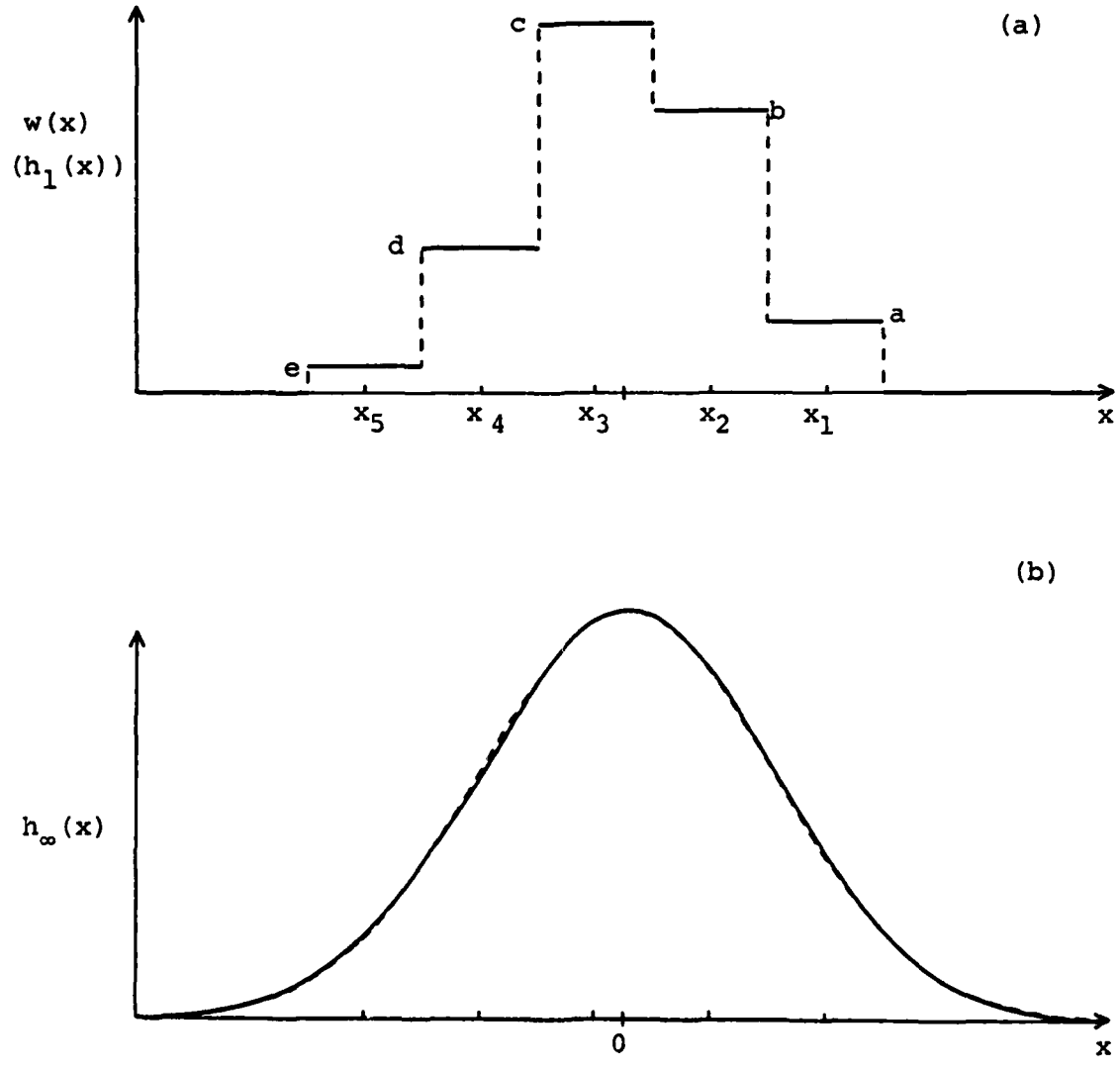


Figure 11.

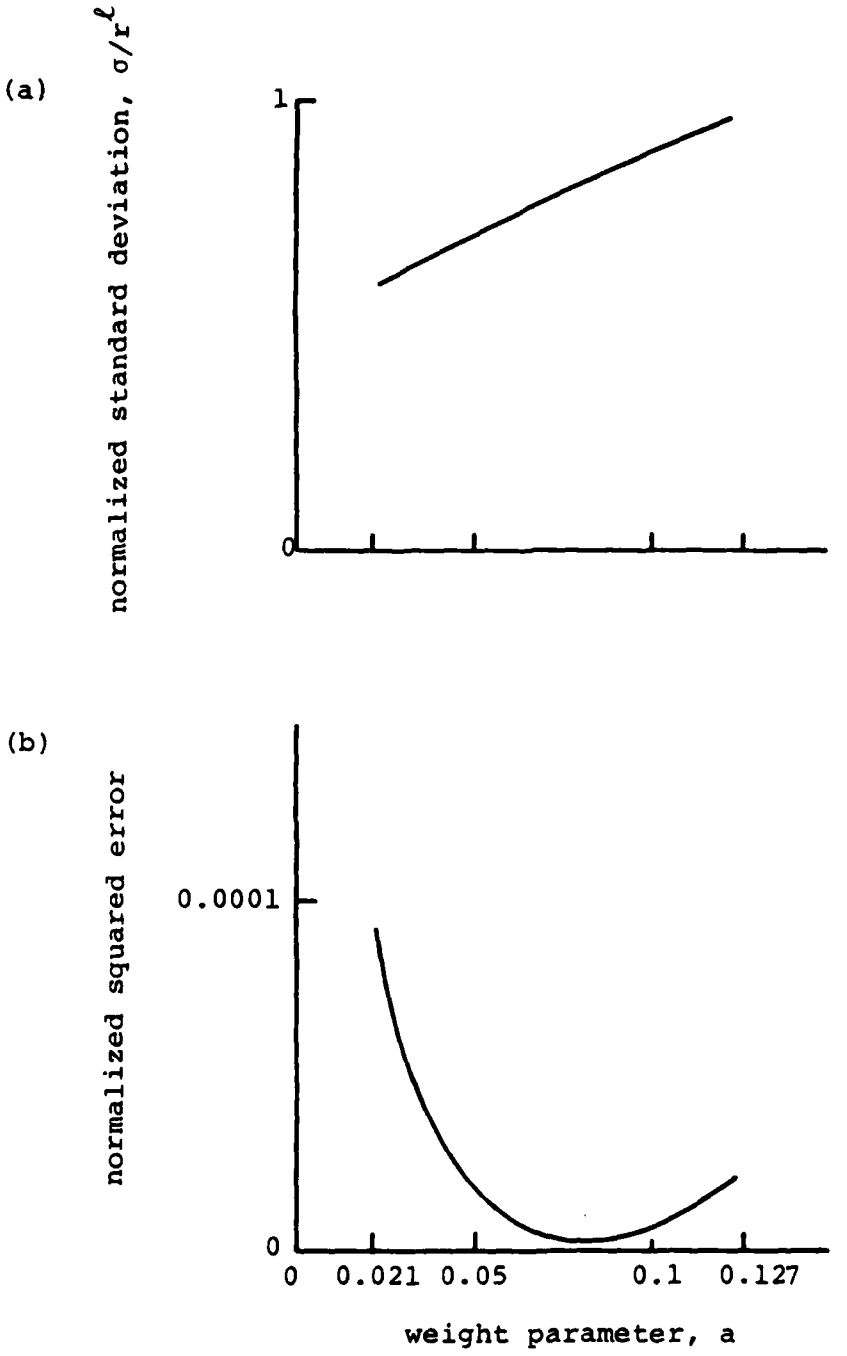
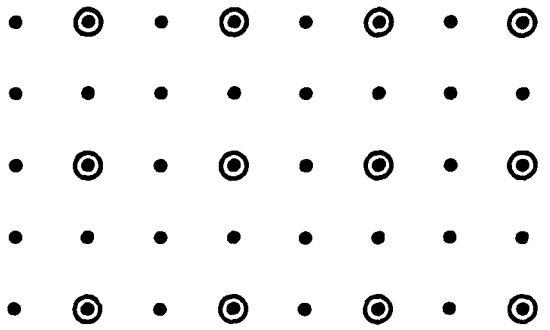
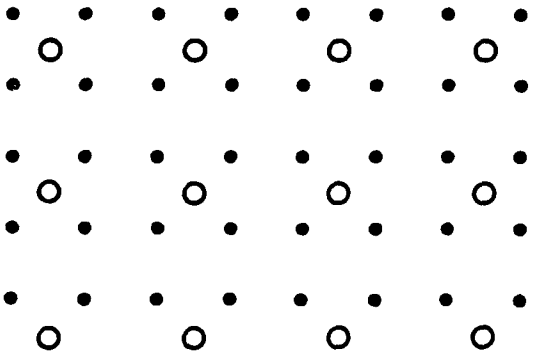

(a)
normalized standard deviation, σ/r^{ℓ}
1
0
(b)
normalized squared error
0.0001
0
0
0.021
0.05
0.1
0.127
weight parameter, a

Figure 12.

(a)
Type 1, order r = 2



(b)
Type 2, order r = 2



(c)
Type 3, order r = 3/2

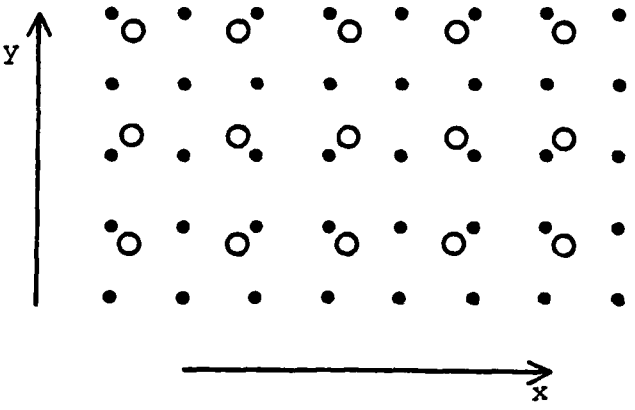


Figure 13.

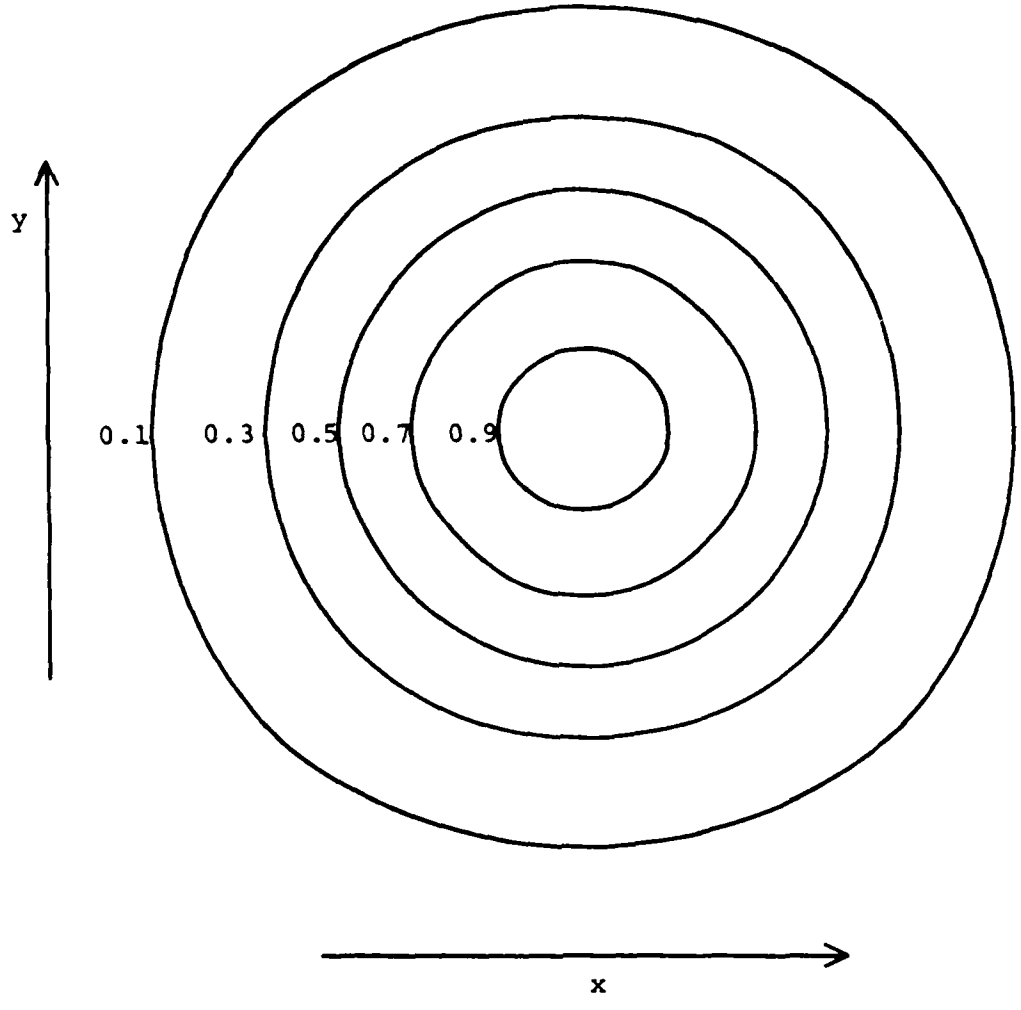


Figure 14a.

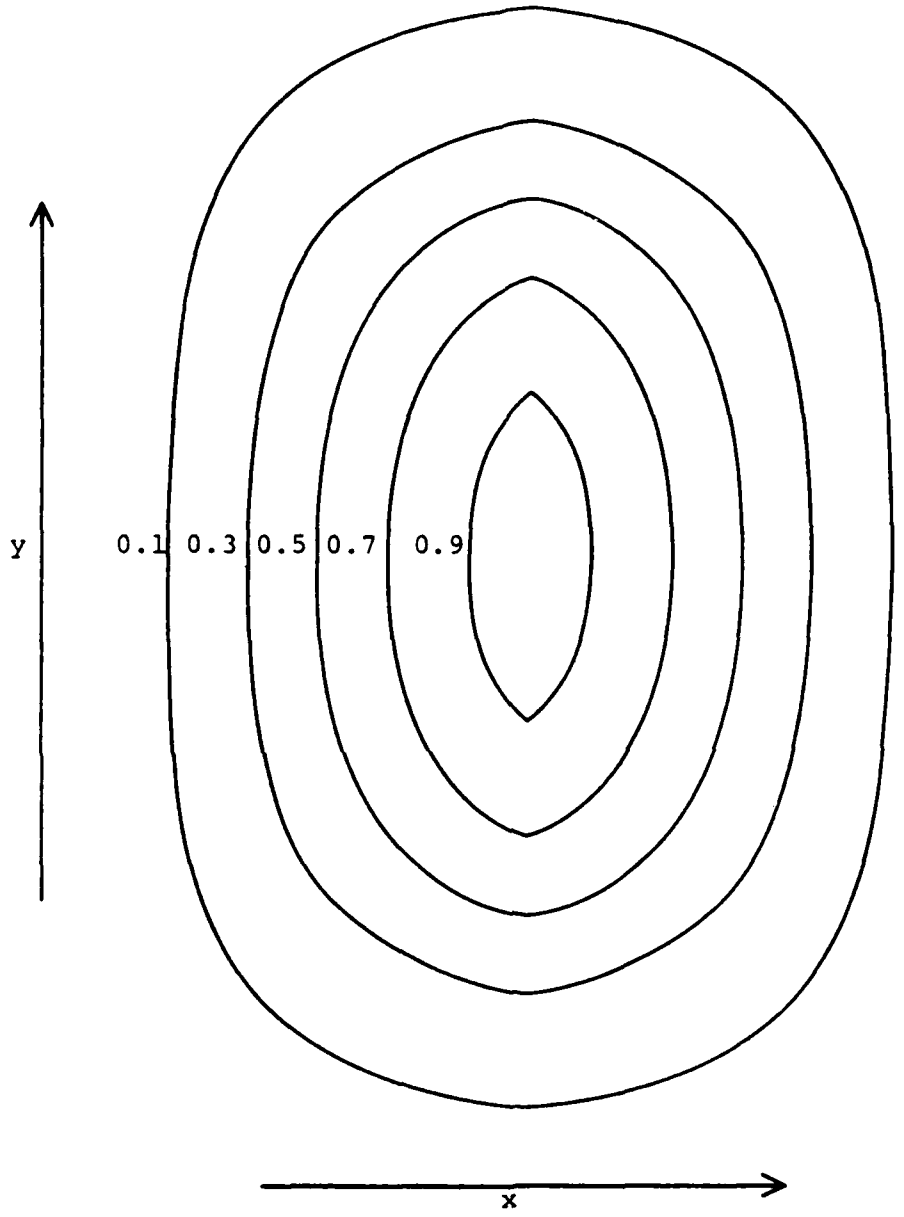


Figure 14b.

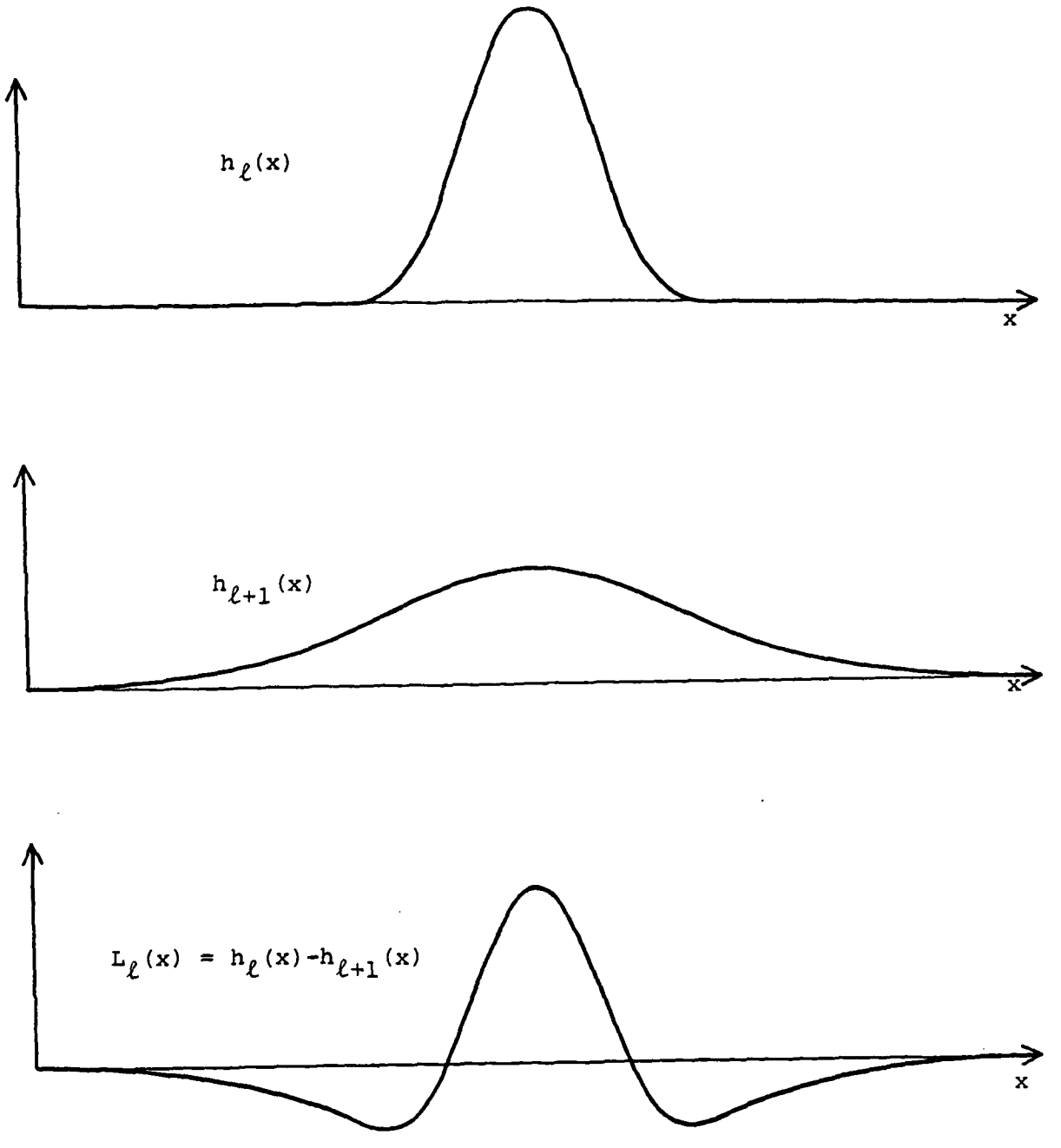


Figure 15.

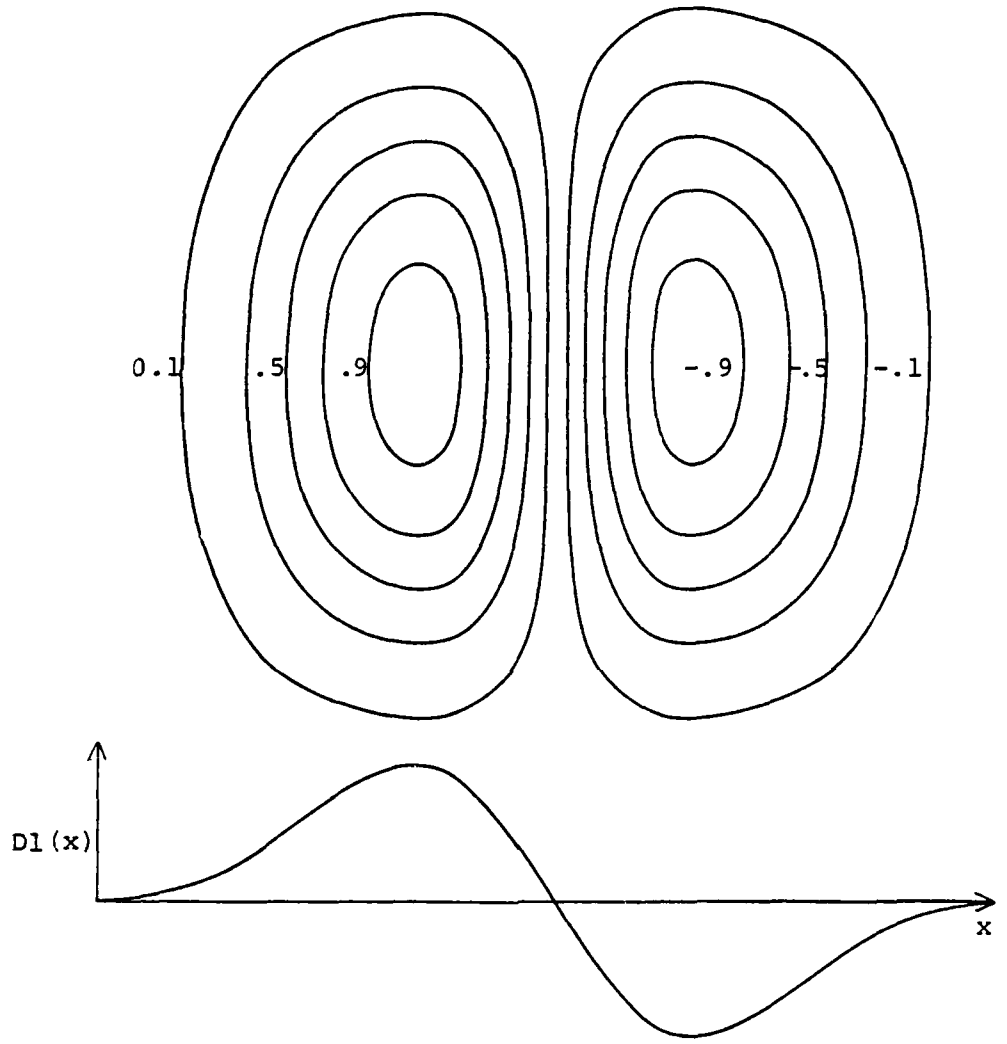
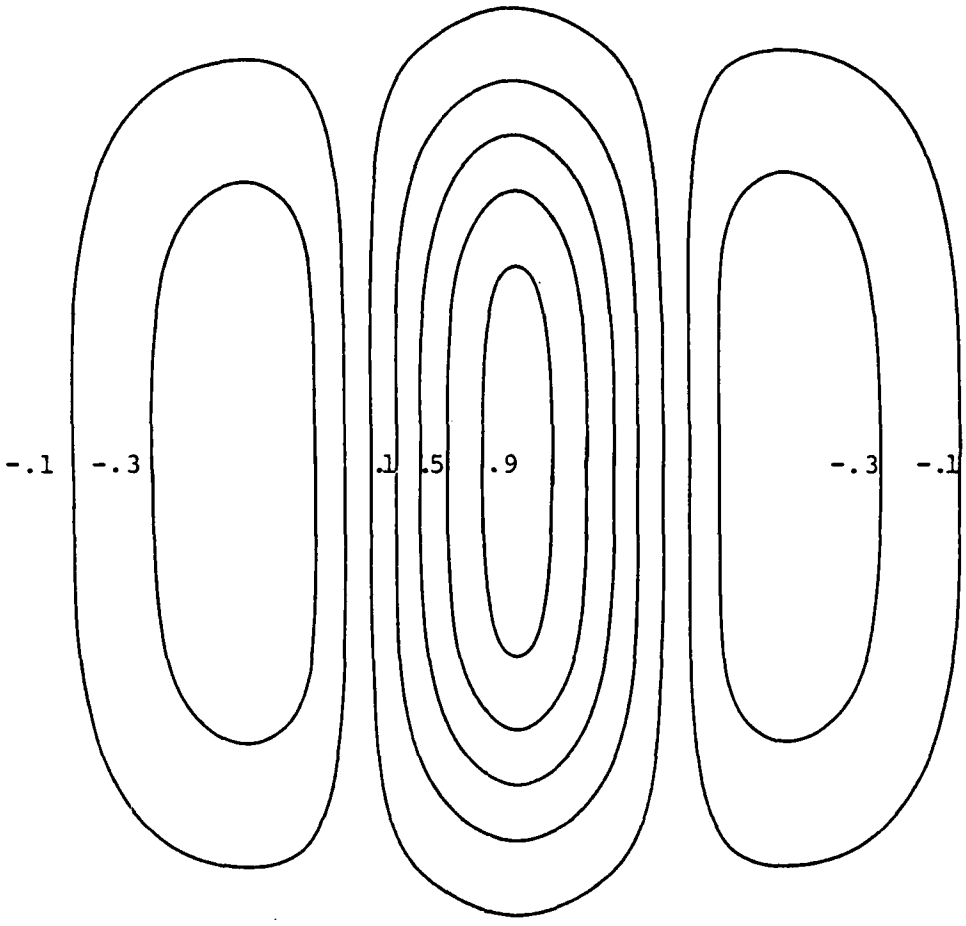


Figure 16.



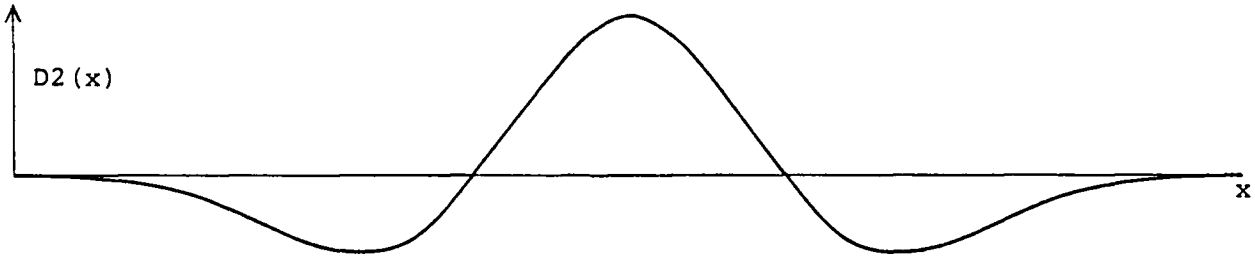


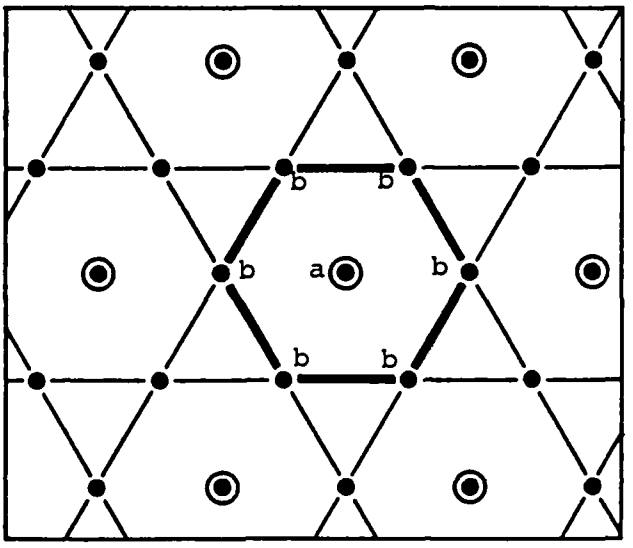
Figure 17.

(a)

Order r =2

a = 1/4

b = 1/8



(b)

Order r = $\sqrt{3}$

a = 1/3

b = 1/9

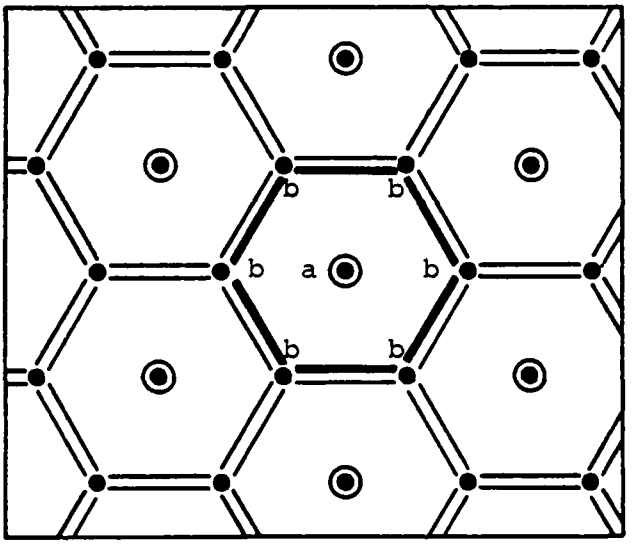


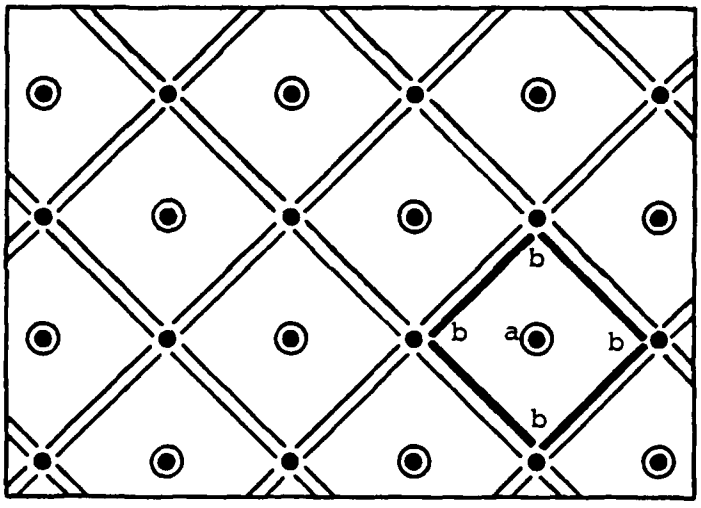
Figure 18a, b

(c)

Order r = $\sqrt{2}$

a = 1/2

b = 1/8



(d)

Order r = $\sqrt{5}$

a = 1/5

b = 1/5

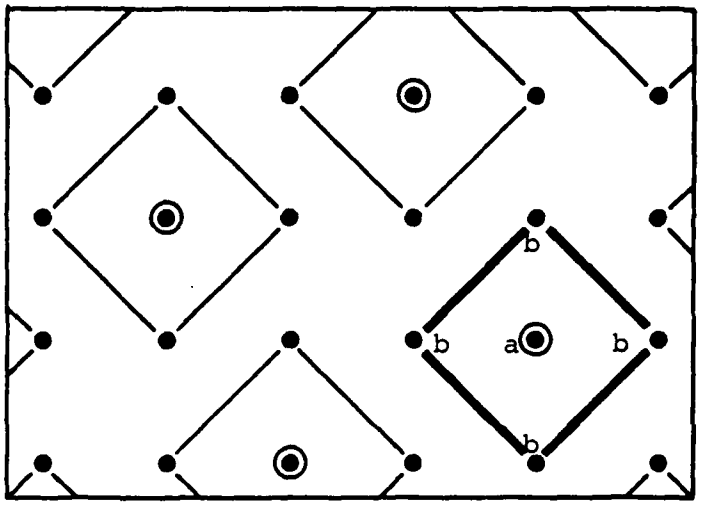


Figure 18c, d

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A090 242	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) FAST, HIERARCHICAL CORRELATIONS WITH GAUSSIAN-LIKE KERNELS	5. TYPE OF REPORT & PERIOD COVERED Technical
	6. PERFORMING ORG. REPORT NUMBER TR-860
7. AUTHOR(s) Peter J. Burt	8. CONTRACT OR GRANT NUMBER(s) DAAG-53-76C-0138
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Vision Laboratory, Computer Science Center, University of Maryland, College Park, MD 20742	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Night Vision Laboratory Ft. Belvoir, VA 22060	12. REPORT DATE January 1980
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Image processing
Signal processing
Convolution
Correlation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A hierarchical procedure is desribed for computing the discrete correlation, $h_\ell(x,y)$ o $f(x,y)$, between a function f and a kernel h_ℓ. For the class of h considered here, this correlation is equivalent to a weighted sum of correlations of $f(x,y)$ with the kernel $h_{\ell-1} \cong r^2 h_\ell(x/r, y/r)$. Here $r>1$, so $h_{\ell-1}$ is narrower than

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

h_ℓ by 1/r. The correlations of f with $h_{\ell-1}$ are themselves computed as the weighted sums of correlations with still narrower kernels. The narrowest kernel, h_0, has unit width, and its correlation with f is f itself.

Kernels which can be computed hierarchically in this way closely approximate the Gaussian probability distribution. Hierarchical discrete correlation (HDC) is more efficient than direct correlation or correlation using the FFT, by two to three orders of magnitude.

The HDC has immediate applications to computer image processing. Samples of the correlations obtained at nearby image positions can be summed to obtain band pass Laplacian and oriented first and second derivative operators ("Mexican hat", edge and bar masks). The correlation of an image with many operators and at many scales requires little computation beyond a single HDC.