Histogram Analysis Using a Scale-Space Approach

MARK J. CARLOTTO

Abstract—**A new application of scale-space filtering to the classical problem of estimating the parameters of a normal mixture distribution is described. The technique involves generating a multiscale description of a histogram by convolving it with a series of Gaussians of gradually increasing width (standard deviation), and marking the location and direction of the sign change of zero-crossings in the second derivative. The resulting description, or *fingerprint*, is interpreted by relating pairs of zero-crossings to *modes* in the histogram where each mode or component is modeled by a normal distribution. Zero-crossings provide information from which estimates of the mixture parameters are computed. These initial estimates are subsequently refined using an iterative maximum likelihood estimation technique. Varying the scale or resolution of the analysis allows the number of components used in approximating the histogram to be controlled.**

Index Terms—**Estimating the parameters of a normal mixture, fingerprints, histogram analysis, image segmentation, mode finding, scale-space filtering.**

I. Introduction

Recently, the idea of filtering across a continuum of scales using Gaussian filters has been explored for the purpose of constructing symbolic descriptions of signals [14] and shape [1]. Stansfield [11] motivated by the use of multiple Gaussian filters for edge detection [7] observed that the zero-crossing locations of 1-D signals traced out contours in a 2-D space later termed *scale-space* by Witkin. Stansfield alluded to the possible use of Gaussian filters for constructing multiscale representations of 1-D signals. Witkin [14] actually developed a multiscale representation based on ternary trees and described a methodology for extracting the perceptually salient features of the signal. Babaud *et al.* [2] proved that the Gaussian is the only linear filter which has the desirable property of not creating zero-crossings as the scale decreases. (This allows the signal to be represented, for example, in terms of a ternary tree.) Yuille and Poggio [12] showed that 2-D signals (images) smoothed by Gaussian filters have similar properties. They also proved that the positions of zero-crossings in the second-derivative capture all the information that is needed to reconstruct the signal up to a scale factor plus a constant and harmonic term [13].

Multiscale descriptions in terms of the location of zero-crossings of the derivatives of the signal in scale-space have become known as fingerprints. By using fingerprints, events can be *detected* at coarse scales and *localized* by tracking zero-crossing contours in scale-space down to fine scales. For example, if one is interested in extracting peaks and valleys in a 1-D waveform, zero-crossings in the first derivative can be detected at a scale where only the more significant variations in the waveform remain. The detected zero-crossings can then be tracked down to a lower scale where the precise location of the peak or valley can be determined.

In this correspondence, an application of scale-space filtering to the problem of approximating a signal, in this case the frequency distribution or histogram from a random process, by a sum or mixture of normal distributions is described. Analysis of the fingerprint of a histogram indicates that the histogram may be constructed up to, or at, any scale by a sum of normal distributions. Since the tails

Manuscript received December 4, 1984; revised January 20, 1986. Recommended for acceptance by S. W. Zucker.

The author is with the Analytic Sciences Corporation, Reading, MA 01867.

IEEE Log Number 8608233.

of the distributions will generally overlap, localization by tracking zero-crossing contours down to fine scales is not sufficient to extract the individual components. Instead, initial estimates of the parameters which determine the underlying mixture are computed from the locations of the zero-crossings at the scale in which they are detected, and are refined using an iterative technique. The approximation of the histogram which results is thus scale-dependent in the sense that the number of components used is related to the scale at which the histogram was analyzed.

The organization of this paper is as follows. Section II examines simple mixtures of univariate normal distributions and explores, in particular, how the locations of zero-crossings in the first and second derivative relate to the parameters of the mixtures. In Section III, fingerprints of normal distributions are analyzed and a method for approximating histograms at any scale by normal mixtures is described. Estimation of the mixture parameters themselves is then addressed in Section IV. Experimental results are presented in Section V. Section VI concludes with a discussion of the current limitations of the method and plans for future work.

II. Zero-Crossings in the Derivatives of Normal Distributions

The univariate normal distribution $N(\mu, \sigma)$

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \tag{1}$$

is completely specified by its mean μ and standard deviation σ. The first and second derivatives of the Gaussian are given by

$$\frac{\partial g}{\partial x} = \frac{(\mu - x)}{\sqrt{2\pi}\sigma^3} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \tag{2a}$$

and

$$\frac{\partial^2 g}{\partial x^2} = \left[\frac{(\mu - x)^2}{\sigma^2} - 1\right] \frac{1}{\sqrt{2\pi}\sigma^3} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]. \tag{2b}$$

A zero-crossing occurs in the first derivative, i.e., the sign of the slope changes from positive to negative, at the peak (which is also the mean) of the original distribution. Zero-crossings in the second derivative are points of inflection in the original distribution and occur at $\mu \pm \sigma$.

The two component normal mixture is given by

$$m(x) = \frac{p_1}{\sqrt{2\pi}\sigma_1} \exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right] + \frac{p_2}{\sqrt{2\pi}\sigma_2} \exp\left[\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right]. \tag{3}$$

The first two derivatives of (3) are

$$\frac{\partial m}{\partial x} = \frac{(\mu_1 - x)}{\sqrt{2\pi}\sigma_1^3} \exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right] + \frac{(\mu_2 - x)}{\sqrt{2\pi}\sigma_2^3} \exp\left[\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right] \tag{4a}$$

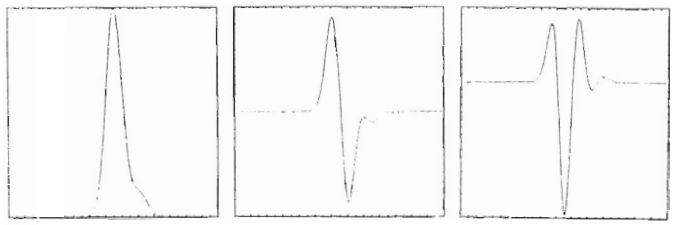
and

$$\frac{\partial^2 m}{\partial x^2} = \left[\frac{(\mu_1 - x)^2}{\sigma_1^2} - 1\right] \frac{1}{\sqrt{2\pi}\sigma_1^3} \exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right] + \left[\frac{(\mu_2 - x)^2}{\sigma_2^2} - 1\right] \frac{1}{\sqrt{2\pi}\sigma_2^3} \exp\left[\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right]. \tag{4b}$$

Three cases illustrating the different ways in which a two component mixture can manifest itself are shown in Figs. 1–3. In Fig. 1(a), the two component distributions are far enough apart so that two distinct peaks are observed. The three zero-crossings in the first-derivative [Fig. 1(b)] correspond to the two peaks and one valley in Fig. 1(a). In the second derivative [Fig. 1(c)], pairs of zero-crossings occur on either side of the peaks. According to (4), if the component densities are far enough apart, the peaks will be close to the means and the distance between turning points on either side of a peak will be approximately equal to twice the standard deviation.

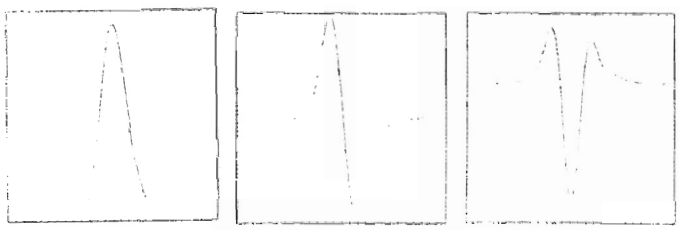
(a) (b) (c)

Fig. 1. Two-component normal mixture and derivatives (distinct peaks).



(a) (b) (c)

Fig. 2. Two-component normal mixture and derivatives (buried mode).



(a) (b) (c)

Fig. 3. Two-component normal mixture and derivatives (single peak).

In general, the individual components in the mixture will not be well-separated as in Fig. 1, and the means and variances will not correspond to the locations of zero-crossings in the first and second derivatives. For example, in Fig. 2(a) the standard deviation of each density has been increased, causing the smaller of the two peaks and the valley to disappear. A single peak remains, producing one zero-crossing in first derivative [Fig. 2(b)]. However, the presence of the second, smaller mode still manifests itself as a pair of zero-crossings in the second derivative [Fig. 2(c)]. In Fig. 3(a) the standard deviations are further increased so as to cause the larger mode to completely dominate and obliterate the smaller mode. One zero-crossing occurs in the first derivative [Fig. 3(b)] and only two zero-crossings occur in the second derivative [Fig. 3(c)].

The location of peaks, valleys, and turning points are plotted in Fig. 4(a)–(c) for the distributions in Figs. 1–3. The two modes observed in Fig. 1 produce two pairs of turning points (a). Although only one peak is present in Fig. 2, two pairs of turning points remain (b). Only in Fig. 3 does the larger component completely dominate the smaller (c). Therefore, it appears that if the components are spaced far enough apart to be resolvable [5] each will give rise to a pair of zero-crossings.

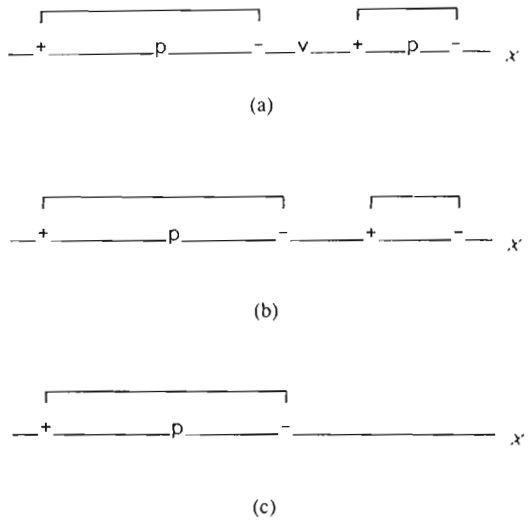


Fig. 4. Zero-crossings in previous three mixtures ("p"—peak, "v"—valley, "+"—plus-to-minus sign change in second derivative, "−"—minus-to-plus sign change in second derivative.

III. Fingerprints of 1-D Signals

The apparent similarity in structure between Fig. 4 and slices from fingerprints of 1-D signals suggests a possible use of scale-space filtering in decomposing histograms into sums of normal distributions. In the following subsections, a review of scale-space filtering is provided (Section III-A), followed by an examination of fingerprints of normal distributions (Section III-B), and the discussion of an approach for approximating the histogram by a normal mixture (Section III-C).

A. Scale-Space Filtering

For a continuous signal $f(x)$, Witkin [14] defines its scale-space image to be

$$F(x, \tau) = f(x) * g(x, \tau) = \int_{-\infty}^{\infty} f(u) \frac{1}{\sqrt{2\pi}\tau} \exp\left| \frac{-(x-u)^2}{2\tau^2} \right| du \qquad (5)$$

where "*" denotes convolution. A slice through the scale-space image at $\tau = \tau_0$ is the signal smoothed by a Gaussian of standard deviation τ_0. Fig. 5 shows a 1-D signal (a) and its scale-space image (b). The scale varies from one to 100 and is plotted logarithmically in (b).

Fingerprints are constructed by plotting the locations of zero-crossings (usually, of the second derivative of F) in scale-space. Yuille and Poggio [13] have shown that fingerprints capture all the information that is needed to represent the signal up to a scale factor plus a constant and harmonic term. It has also been shown that signals smoothed by convolution with Gaussians create fingerprints having nice topological properties [2], [12]. For one, zero-crossings are never created as the scale increases. Witkin [14] developed a symbolic description based on ternary trees which results from this topological property of fingerprints.

For the purpose of analyzing histograms, we shall be interested in the location of zero-crossings in the second derivative

$$F_{xx}(x, \tau) = \frac{\partial^2}{\partial x^2} [F(x, \tau)], \qquad (6)$$

and the sign of the third derivative at the zero-crossings; i.e., whether the sign of the second derivative is changing from plus to minus ($F_{xxx} < 0$) or minus to plus ($F_{xxx} > 0$). Fig. 5(c) shows the sign of the second derivative of the scale-space image in Fig. 5(b), where white (black) denotes $F_{xx} \geq 0$ ($F_{xx} < 0$). Zero-crossings in the second derivative of a waveform are the positions where

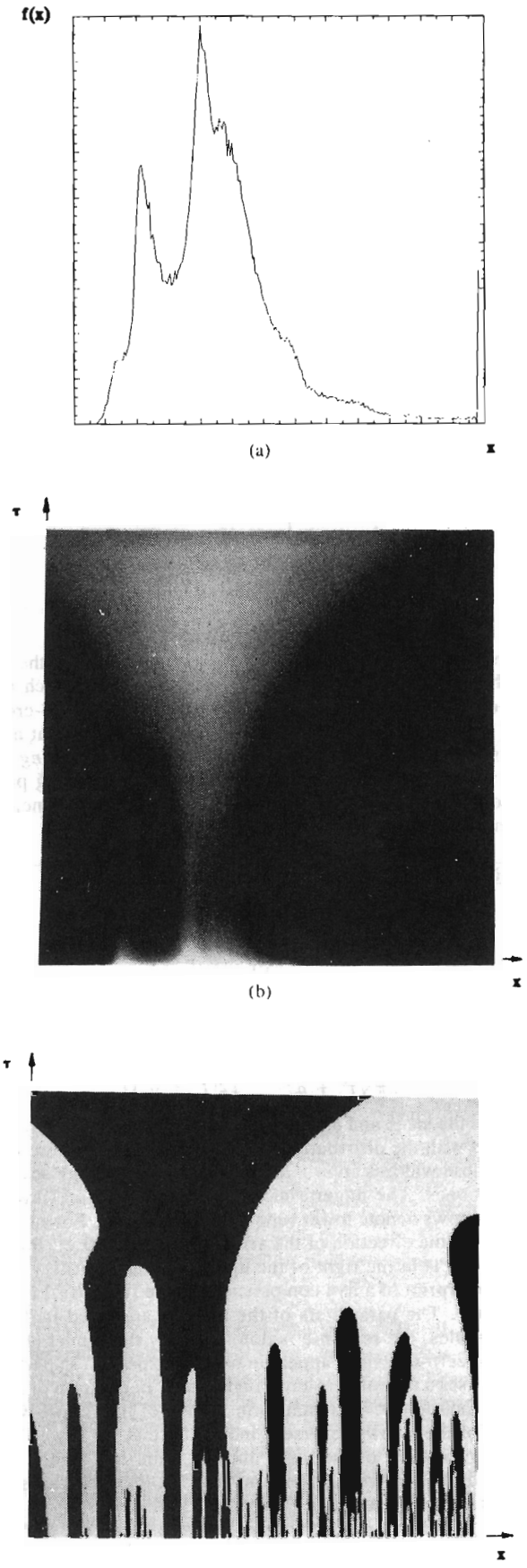


Fig. 5. (a) Histogram, (b) its scale-space image, and (c) the sign of the second derivative.

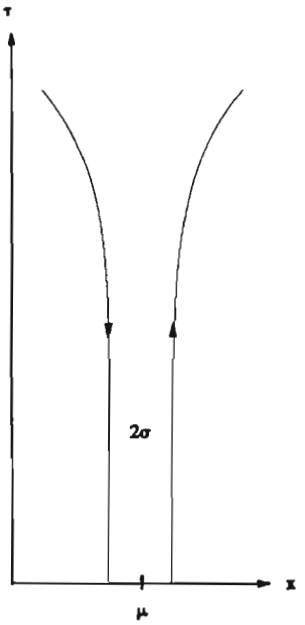


Fig. 6. Fingerprint of a normal distribution.

the curvature of the waveform changes sign. Sections of the signal over which $F_{xx} < 0$ are convex, while sections over which $F_{xx} > 0$ are concave. If $F_{xxx} < 0$ at a zero-crossing, the zero-crossing will be labeled a *lower turning point*, and if $F_{xxx} > 0$ at a zero-crossing, the zero-crossing will be labeled an *upper turning point*. Thus for the normal distribution in (1), the lower turning point is to the left of the upper turning point. If the distribution is negative, the turning point order is reversed.

B. Fingerprints of Normal Distributions

The impulse response of the scale space filter $g(x, \tau)$ is given by

$$\frac{1}{\sqrt{2\pi}\tau} \exp\left[\frac{-x^2}{2\tau^2}\right]. \tag{7}$$

A normal distribution thus produces the following response

$$\frac{1}{2\pi\sqrt{\tau^2 + \sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2(\tau^2 + \sigma^2)}\right] \tag{8}$$

where μ is the mean and σ is the standard deviation. The fingerprint associated with the distribution is shown in Fig. 6 and resembles a funnel whose width is approximately equal to 2τ at high scales and 2σ at low ones. The fingerprint is centered at $x = \mu$. Down (up) pointing arrows denote lower (upper) turning points. For a negative distribution, the direction of the arrows is reversed (i.e., the lower turning point is to the right of the upper turning point).

The fingerprint of a five component mixture [Fig. 7(a)] is shown in Fig. 7(b). The parameters of the mixture are listed in Table I. At large scales, the response is dominated by the funnel (1). As τ decreases, zero-crossings appear in pairs as expected. In particular, as τ decreases a pair of zero-crossings appear in the form of an arch closed at the top (2). The arch forms inside of the funnel with the direction of the arrows reversed, indicating that the lower turning point is to the right of the upper turning point and signifying the presence of a negative distribution. As τ further decreases, three more pairs of zero-crossings (3–5) form. It is interesting to note at this point that even though the signal represented in the fingerprint is a sum of five positive Gaussians, a top-down decomposition of it suggests that is can be constructed at certain scales by sums of positive and negative Gaussians.

A slice from Fig. 7 just below the point where the third pair of zero-crossings appears is shown in Fig. 8. The pairing of zero-crossings obtained from top-down analysis of the fingerprint is shown in (a). The five alternate pairings (b) were obtained by considering all ways to pair upper and lower turning points. (For K pairs of zero-crossings there are $K!$ possible pairings.) All pairings in (b) except for the bottom one imply a decomposition of the signal into both positive and negative components. (Pearson [8] showed that there are at most two ways to decompose a two component mixture, one consisting of a sum of two positive components, the other consisting of one positive and one negative com-

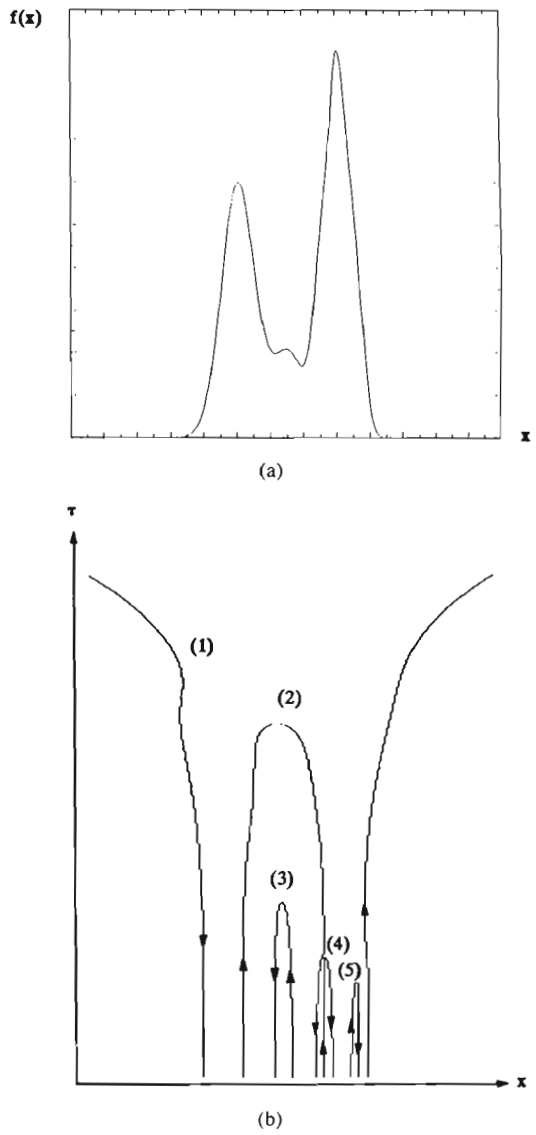


Fig. 7. (a) Five-component normal distribution and (b) its fingerprint.

TABLE I
EXAMPLE FIVE-COMPONENT MIXTURE

Mode	Probability	Mean	Standard Deviation
1	0.375	100	10
2	0.125	130	10
3	0.125	150	5
4	0.250	160	5
5	0.125	170	5

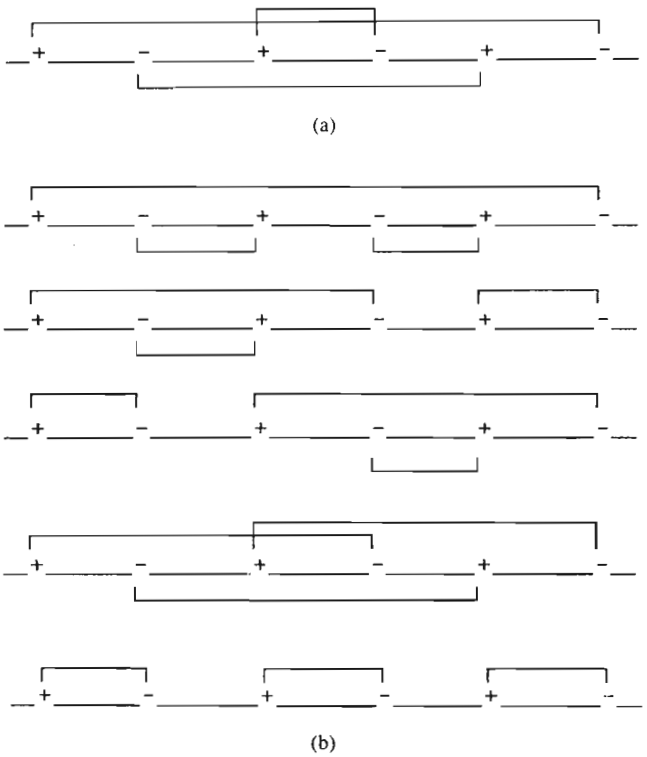


Fig. 8. Possible pairings of zero-crossings have the opposite sign.

ponent.) The six pairings in Fig. 8 represent the six possible solutions to (i.e., decompositions of) the three component problem. Only the last one represents a decomposition of the signal into strictly positive components. (Note that there is only one such decomposition into strictly positive components.) Thus to detect positive modes in the histogram, the fingerprint is parsed left to right (i.e., in the direction of increasing x) at a particular scale. Each pair of zero-crossings detected is associated with a positive distribution present at that scale.

C. Interpreting Fingerprints of Histograms

The fingerprint of the histogram in Fig. 5 is shown in Fig. 9. It contains the same kinds of features found in the fingerprints in the preceding section, namely a funnel which dominates the fingerprint as large scales, pairs of zero-crossings of opposite sign change forming as the scale decreases, and alternations of the sign change in going left to right. (Arrows denoting the direction of zero-crossings were omitted from this figure for clarity.) The similarity in structure is due to the fact that smoothing a signal by convolving it with a Gaussian produces a result which is a superposition of Gaussian responses (i.e., a sum or mixture of normal distributions). This then suggests that the histogram may be constructed top-down by adding positive and negative Gaussians together up to any scale, or alternately, at any scale.

Fig. 10 shows an earlier [Fig. 5(a)] histogram (a) approximated at three scales (b)–(d). (The three distributions (b)–(d) are obtained by the procedure described in the next section.) The three scales chosen are indicated in the fingerprint in Fig. 9. It is apparent in Fig. 9 that as the scale decreases (the histogram is smoothed by decreasing amounts), additional components, i.e., pairs of zero-crossings, form in the fingerprint. At $\tau = 5$, $K = 8$ pairs of zero-crossings (modes) are detected. This approximation captures most of the major features of the histogram, e.g., the lower mode corresponds to shadows in the image over which it was computed, and the higher mode corresponds to specular areas. By further increasing τ in (c) and (d), the number of modes detected, and correspondingly, the accuracy of the resulting approximation, decreases.

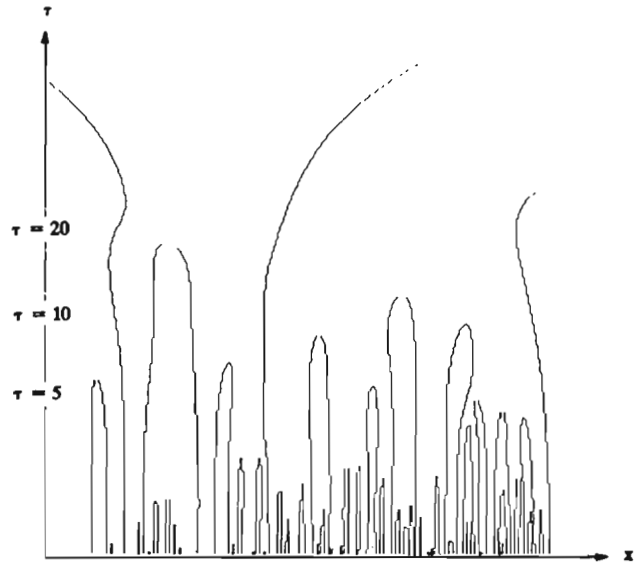


Fig. 9. Fingerprint of histogram.

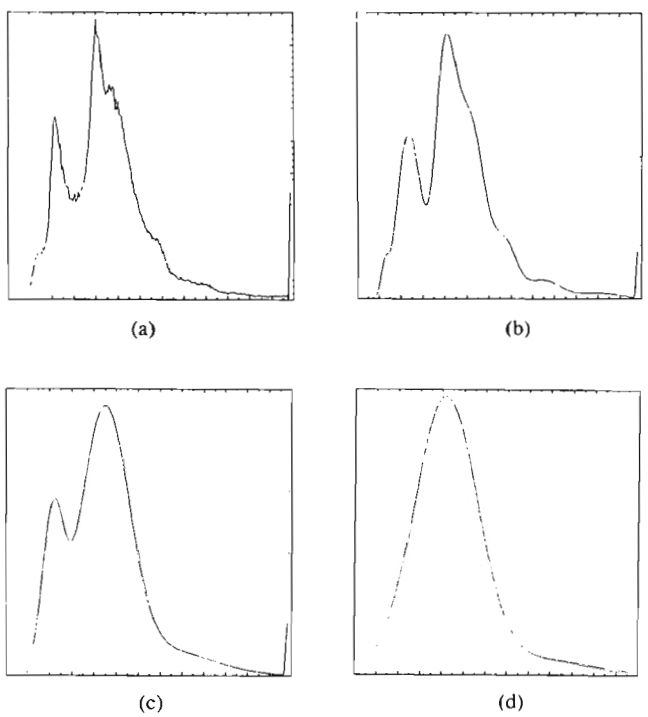


Fig. 10. Analysis of histogram (a) at three resolutions (b)–(d).

IV. Estimating the Parameters of a Normal Mixture

Redner and Walker [9] provide an overview of various methods developed for estimating the parameters of mixture densities. Many are concerned with the problem of determining the parameters of normal mixtures. The original work by Pearson [8] addressed the problem of determining the parameters of two univariate normal densities. His approach, known as the method of moments, equates expected values to sample moments; e.g., in the two component case, five parameters must be determined by solving a set of five (generally) nonlinear equations. Due to the computational complexity of this approach, evident even in the two component case, many other methods (both direct and iterative) have since been developed.

A graphical solution for the parameters of a normal mixture in

cases where the component distributions are well-separated was described by Bhattacharya [3]. His approach, adequate for the situation illustrated in Fig. 1, would fail in the situations depicted in Figs. 2 and 3. In all three distributions, the ability to resolve the two modes is dependent on the variance, the distance between the means, and on the relative strengths of the components. Sammon [10] developed a method for resolving individual components based on a deconvolution approach. Its applicability is limited to cases where the individual components have approximately the same variance.

In general, to determine a K components normal mixture, $(3K - 1)$ parameters must be estimated. A direct method for computing these parameters as a function of the location of zero-crossings is to form a system of $(3K - 1)$ simultaneous nonlinear equations. Consider determining the parameters which make up the mixture in Fig. 2. If z_1 is the location of the one zero-crossing in the first derivative, and z_2 through z_5 are the locations of zero-crossings in the second derivative, one must solve the system

$$\frac{p_1(\mu_1 - z_1)}{\sqrt{2\pi}\sigma_1^3}\exp\left[\frac{-(z_1-\mu_1)^2}{2\sigma_1^2}\right] + \frac{p_2(\mu_2 - z_1)}{\sqrt{2\pi}\sigma_2^3}\exp\left[\frac{-(z_1-\mu_2)^2}{2\sigma_2^2}\right] = 0 \tag{9a}$$

$$\left[\frac{(\mu_1 - z_2^2)}{\sigma_1^2} - 1\right]\frac{p_1}{\sqrt{2\pi}\sigma_1^3}\exp\left[\frac{-(z_2-\mu_1)^2}{2\sigma_1^2}\right] + \left[\frac{(\mu_2 - z_2^2)}{\sigma_2^2} - 1\right]\frac{p_2}{\sqrt{2\pi}\sigma_2^3}\exp\left[\frac{-(z_2-\mu_2)^2}{2\sigma_2^2}\right] = 0 \tag{9b}$$

$$\left[\frac{(\mu_1 - z_3^2)}{\sigma_1^2} - 1\right]\frac{p_1}{\sqrt{2\pi}\sigma_1^3}\exp\left[\frac{-(z_3-\mu_1)^2}{2\sigma_1^2}\right] + \left[\frac{(\mu_2 - z_3^2)}{\sigma_2^2} - 1\right]\frac{p_2}{\sqrt{2\pi}\sigma_2^3}\exp\left[\frac{-(z_3-\mu_2)^2}{2\sigma_2^2}\right] = 0 \tag{9c}$$

$$\left[\frac{(\mu_1 - z_4^2)}{\sigma_1^2} - 1\right]\frac{p_1}{\sqrt{2\pi}\sigma_1^3}\exp\left[\frac{-(z_4-\mu_1)^2}{2\sigma_1^2}\right] + \left[\frac{(\mu_2 - z_4^2)}{\sigma_2^2} - 1\right]\frac{p_2}{\sqrt{2\pi}\sigma_2^3}\exp\left[\frac{-(z_4-\mu_2)^2}{2\sigma_2^2}\right] = 0 \tag{9d}$$

$$\left[\frac{(\mu_1 - z_5^2)}{\sigma_1^2} - 1\right]\frac{p_1}{\sqrt{2\pi}\sigma_1^3}\exp\left[\frac{-(z_5-\mu_1)^2}{2\sigma_1^2}\right] + \left[\frac{(\mu_2 - z_5^2)}{\sigma_2^2} - 1\right]\frac{p_2}{\sqrt{2\pi}\sigma_2^3}\exp\left[\frac{-(z_5-\mu_2)^2}{2\sigma_2^2}\right] = 0 \tag{9e}$$

for p_1, μ_1, μ_2, σ_1, and σ_2, where $p_2 = 1 - p_1$. Thus, even for $K = 2$ the solution is difficult, if not impossible, to obtain analytically.

A two-step approach for estimating the mixture parameters which overcomes these difficulties is based on computing a rough estimate of the parameter values from the zero-crossing locations, and then refining the estimate using an iterative maximum likelihood estimation algorithm. An overview of the estimation process is shown in Fig. 11. Steps (a) and (b) have already been described. At any scale, sign changes will alternate left to right. Odd (even) numbered zero-crossings will thus correspond to lower (upper) turning points.

Prior to entering the iterative loop, initial estimates of the mixture parameters are computed from zero-crossing locations in step (c). Let a_k and b_k be the locations of upper and lower turning points for $k = 1, 2, \cdots, K$. The point halfway between the turning point pairs $(a_k + b_k)/2$ is used as an estimate of the mean $\hat{\mu}_k$. Half the distance between turning point pairs $(b_k - a_k)/2$ is used as an estimate of the standard deviation $\hat{\sigma}_k$. The histogram $\{f(x_n)\}$ is partitioned into K segments by assigning a bin to the kth segment if

$$|\hat{\mu}_k - x_n| < |\hat{\mu}_j - x_n| \tag{10}$$

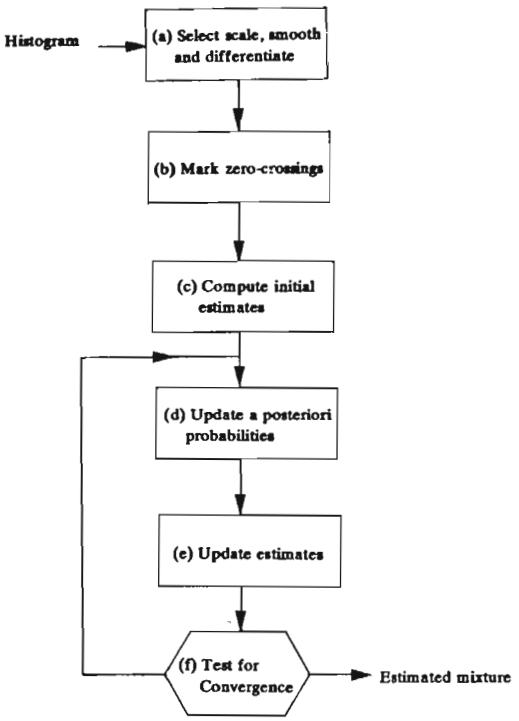


Fig. 11. Overview of parameter estimation procedure.

for $j = 1, 2, \cdots, K$ and $j \neq k$. The fraction of the total area in each segment is taken as an initial estimate of the relative frequency $\hat{p}_k$.

Given these initial estimates of the parameters which determine the mixture, an iterative algorithm described in Duda and Hart [4] can be used to refine the estimates. Specifically, we compute the parameters Θ which maximize the probability of observing the distribution, $f(x_n)$, given Θ. Upon entering the iterative loop in Fig. 11, the *a posteriori* probabilities are computed first in step (d) according to

$$\hat{p}(\omega_k|x) = \frac{\hat{p}(x|\omega_k)\hat{p}_k}{\sum_{k=1}^{K}\hat{p}(x|\omega_k)\hat{p}_k} \tag{11}$$

where

$$\hat{p}(x|\omega_k) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_k}\exp\left[\frac{-(x-\hat{\mu}_k)^2}{2\hat{\sigma}_k^2}\right] \tag{12}$$

is the conditional probability distribution for the kth component. These probability functions are used to compute new estimates in step (e) according to

$$\hat{p}_k = \frac{\sum_{n=1}^{N}\hat{p}(\omega_k|x_n)\,f(x_n)}{\sum_{n=1}^{N}\hat{p}(\omega_k|x_n)}$$

$$\hat{\mu}_k = \frac{\sum_{n=1}^{N}\hat{p}(\omega_k|x_n)\,f(x_n)\,x_n}{\sum_{n=1}^{N}\hat{p}(\omega_k|x_n)\,f(x_n)} \tag{13}$$

and

$$\hat{\sigma}_k = \frac{\sum_{n=1}^{N}\hat{p}(\omega_k|x_n)\,f(x_n)\,(x_n - \hat{\mu}_k)^2}{\sum_{n=1}^{N}\hat{p}(\omega_k|x_n)\,f(x_n)}.$$

TABLE II
ESTIMATED MIXTURE PARAMETERS

Mode	Probability	Mean	Standard Deviation
1	0.380 (+1.6%)	100.22 (+0.2%)	10.12 (+1%)
2	0.116 (-7.2%)	130.14 (+0.1%)	9.15 (-8%)
3	0.093 (-25.6%)	148.68 (-0.8%)	4.67 (-6%)
4	0.314 (+25.6%)	159.83 (-0.1%)	5.79 (+15%)
5	0.097 (-22.4%)	170.88 (+0.5%)	4.77 (-5%)

The iteration involves repeating steps (d)–(f) until the means converge (i.e., the magnitude of the difference between successive means is less than a specified amount). When the iteration terminates, the histogram is approximated by summing the individual components together:

$$\hat{f}(x_n) = \sum_{k=1}^{K} \frac{\hat{p}_k}{\sqrt{2\pi}\,\hat{\sigma}_k} \exp\left[\frac{-(x_n - \hat{\mu}_k)^2}{2\hat{\sigma}_k^2}\right]. \quad (14)$$

It must be pointed out that the maximum likelihood solution is but one of several solutions possible using the iterative algorithm. Duda and Hart emphasize the importance of having good initial estimates for the mixture parameters in order to ensure convergence to the correct solution. A key assumption in the two-step approach is that the initial estimates are close enough to the desired solution for the iterative system to converge to it. It has been our experience that if the number of components detected at a particular scale is sufficient to represent the more significant features in the histogram, convergence to reasonable solutions usually results.

V. EXPERIMENTAL RESULTS

In this section three examples illustrating the use of the histogram analysis technique are presented. In our first example, the parameters of the five component mixture in Fig. 7 are estimated using the two-step procedure described in Section IV. The histogram (Fig. 7) is generated by adding together five normal distributions (Table I). Smoothing was not required since the distributions are ideal. The results of the analysis are summarized in Table II. The means all converged to within 1 percent of their true value. A positive bias in the relative frequencies of the larger, better separated components is also evident. The considerable overlap in the distributions resulted in larger errors in the relative frequencies and standard deviations than in the means.

In the second example, a Gaussian noise generator is used to synthesize two component normal mixtures with known statistics. The three cases depicted in Figs. 1–3 are considered. For each, histograms are computed from 10 000, 1000 and 100 noise generator samples. Results of the analysis using a smoothing factor $\tau = 5$ are summarized in Table III–V. When the components are well separated as in Fig. 1 the estimates are quite close to the true parameter values for as few as 100 samples. Even in the situation where the less dominant mode is partially buried as in Fig. 2, satisfactory estimates were obtained for moderate sample sizes (1000 samples or more) as shown in Table IV. For the unimodal case depicted in Fig. 3 the analysis technique should, and does, fail to detect the smaller mode (Table V) when a large number of samples (10 000 or more) are used. The above suggests that satisfactory results may be obtained if the modes are well-separated and/or a large number of samples are used in computing the histograms; otherwise, spurious peaks may be detected by the technique.

The third example illustrates the manner in which the histogram analysis technique is used within our Multi-Spectral Image Analysis System (MSIAS) described in [6]. MSIAS is a knowledge-based system for classifying surface materials (e.g., vegetation, water, roads) in multispectral (e.g., Landsat TM) imagery. A hierarchical classification strategy is used to recursively partition the image into subclasses. At any node within the classifier, subclassification is performed based either on an analysis of the histogram of a particular band or on the spectral signature. The histogram analyzer is used within MSIAS to identify groups of pixels likely to correspond to surface materials based on their spectral reflectance in a particular band(s).

TABLE III
TWO-COMPONENT MIXTURE (DISTINCT PEAKS)

True Probability	True Mean	True Standard Deviation
0.9	130	5
0.1	160	5
Est. Probability	**Est. Mean**	**Est. Standard Deviation**
(10,000 samples)		
0.9	130	4.23
0.1	160	4.25
(1,000 samples)		
0.9	130	4.25
0.1	160	4.52
(100 samples)		
0.9	129.4	4.50
0.1	161.4	4.29

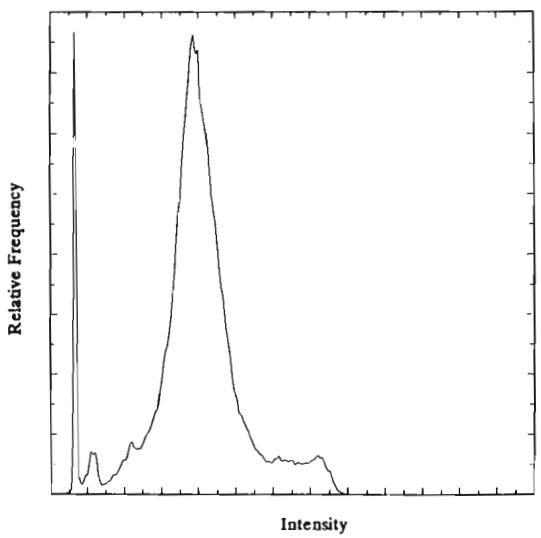
TABLE IV
TWO-COMPONENT MIXTURE (BURIED MODE)

True Probability	True Mean	True Standard Deviation
0.9	130	10
0.1	160	10
Est. Probability	**Est. Mean**	**Est. Standard Deviation**
(10,000 samples)		
0.91	130.2	9.34
0.09	160.5	9.36
(1,000 samples)		
0.86	129.6	9.08
0.14	154.1	12.24
(100 samples)		
0.92	126.1	10.25
0.05	160.2	3.26
0.03	175.4	1.88

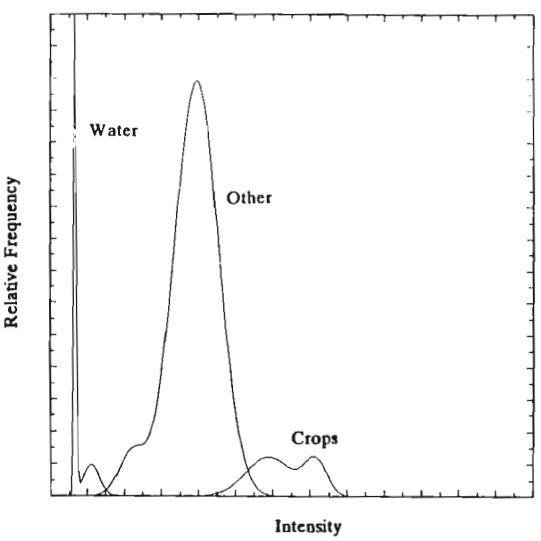
TABLE V
TWO-COMPONENT MIXTURE (SINGLE PEAK)

True Probability	True Mean	True Standard Deviation
0.9	130	15
0.1	160	15
Est. Probability	**Est. Mean**	**Est. Standard Deviation**
(10,000 samples)		
1.0	133.0	16.79
(1,000 samples)		
0.91	130.1	14.08
0.08	157.4	8.72
0.01	183.1	6.36
(100 samples)		
0.11	103.7	6.53
0.26	117.9	2.01
0.46	134.3	7.89
0.14	156.9	5.85
0.03	184.0	3.56

Fig. 12. Landsat Thematic Mapper (TM) image (band 4).

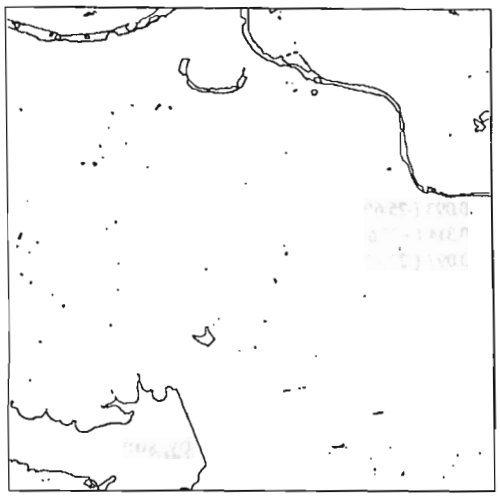


(a)

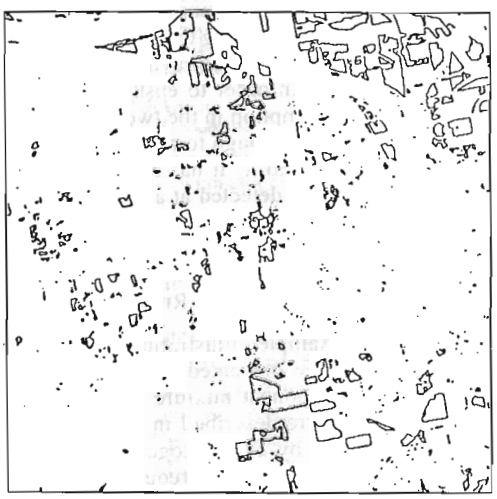


(b)

Fig. 13. (a) Landsat TM band 4 histogram (b) with components due to water, denser vegetation, and other materials identified.



(a)



(b)

Fig. 14. Regions containing (a) the darkest pixels (water) and (b) the brightest pixels (denser vegetation) extracted by thresholding the histogram.

Fig. 12 shows band 4 from the Landsat Thematic Mapper (TM). The histogram shown in Fig. 13(a) is analyzed at a scale which permits the components of interest in the histogram to be easily detected. The six components identified in Fig. 13(b) are detected at a scale $\tau = 2.5$. The two lower components correspond to water (a reservoir and turbid water in a river), and the two higher components correspond to crops and denser vegetation. Other materials present in the image are lumped into the central two modes. Fig. 14 shows bodies of water and areas of denser vegetation extracted by assigning pixels into water and vegetation classes based on a maximum *a posteriori* decision criterion.

VI. Summary

A method for approximating histograms by normal mixtures using a scale-space approach was described. An examination of the fingerprint of a histogram revealed that the histogram could be approximated at any scale by a sum of normal distributions. A two-step approach for estimating the parameters of the underlying distribution was described. It involves a two-step process in which initial estimates are computed from the locations of zero-crossings at the scale of the analysis, and subsequently refined using an iterative maximum likelihood estimation technique. The approxi-

mation of the histogram which results is scale-dependent in the sense that the number of components used is related to the scale at which the histogram was analyzed. The technique is currently limited to analyzing univariate distributions. Extensions to two or more dimensions appears possible but may not be computationally feasible or justified in light of current applications.

References

[1] H. Asada and M. Brady, "The curvature primal sketch," Massachusetts Inst. Technol., Cambridge, AI Lab Memo 758, 1984.

[2] J. Babaud, A. Witkin, and R. Duda, "Uniqueness of the Gaussian kernel for scale-spacing filtering," Fairchild Tech. Rep. 645, 1983.

[3] C. Bhattacharya, "A simple method of resolution of a distribution into Gaussian components," *Biometrics*, vol. 23, 1967.

[4] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[5] I. Eisenberger, "Genesis of bimodal distributions," *Technometrics*, vol. 6, no. 4, Nov. 1964.

[6] R. Ferrante, M. Carlotto, J. Pomarede, and P. Baim, "Multi-spectral image analysis system," in *Proc. 1st Conf. Artificial Intell. Applications*, Denver, CO, 1984.

[7] D. Marr and E. Hildreth, "A theory of edge detection," *Proc. Roy. Soc. London*, vol. 207, 1980.

[8] K. Pearson, "Contributions to the mathematical theory of evolution," *Phil. Trans. Roy. Soc.*, vol. 185a, 1894.

[9] R. Redner and H. Walker, "Mixture densities, maximum likelihood, and the EM algorithm," *SIAM Rev.*, vol. 26, no. 2, 1984.

[10] J. Sammon, "An adaptive technique for multiple signal detection and identification," in *Pattern Recognition*, X. Kanal, Ed. London: Thompson, 1962.

[11] J. Stansfield, "Conclusions from the commodity expert project," Massachusetts Inst. Technol., Cambridge, AI Lab Memo 601, 1980.

[12] A. Yuille and T. Poggio, "Scaling theorems for zero-crossings," Massachusetts Inst. Technol., Cambridge, AI Lab Memo 722, 1983.

[13] ——, "Fingerprint theorems for zero-crossings," Massachusetts Inst. Technol., Cambridge, AI Lab Memo 730, 1983.

[14] A. Witkin, "Scale-space filtering," in *Proc. IJCAI-83*, Aug. 1983, pp. 1019–1022.