MULTISCALE IMAGE SEGMENTATION BY DISHOMOGENEITY EVALUATION AND LOCAL OPTIMIZATION

by

Stefano Casadei

Laurea in Fisica, Universitá degli Studi di Pisa, Italy
(1988)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1991

Author .....
Department of Electrical Engineering and Computer Science
May, 1991

Certified by .....
Sanjoy K. Mitter
Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

MULTISCALE IMAGE SEGMENTATION BY DISHOMOGENEITY EVALUATION AND LOCAL OPTIMIZATION

by

Stefano Casadei

Submitted to the Department of Electrical Engineering and Computer Science
on May, 1991, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

We study the problem of image segmentation in the presence of texture information at several scales. We propose to model homogeneous textured regions as ergodic random functions and to model images as piecewise ergodic random functions. Image properties can then be retrieved by filtering the image at all scale of resolution with a bunch of image descriptors and then averaging this description over regions of different sizes. The result is a representation embedded in a "scale-scale-space" which is an extension of the usual scale-space obtained by adding to it a statistical scale dimension. The ergodicity assumption then is equivalent to assume that the representation becomes deterministic as the statistical scale goes to infinity. Each type of texture can then be characterized by a deterministic "signature" which can be approximately retrieved from the image by using large enough averaging windows.

We propose a method of segmentation based on evaluating the dishomogeneity at all statistical scales and on optimization of a cost function which is basically the dishomogeneity function plus a complexity term. We show that for piecewise ergodic one dimensional images, in the limit of infinitely large regions, this method allows to recover exactly the discontinuity set of the image. Computer experiments with one dimensional images are shown.

Thesis Supervisor: Sanjoy K. Mitter
Title: Professor of Electrical Engineering

Acknowledgements

I am greatly indebted to my advisor Professor Sanjoy K. Mitter for the fascinating and rewarding experience that I am living as an MIT student, for his constant support and encouragement and for his valuable suggestions and ideas which had a strong beneficial influence on the research presented here. In particular, I owe to him the idea of looking at a multiscale representation as a sequence approaching a fixed probability distribution. This suggestion has triggered a dramatic change in my way of thinking to this problem and made possible a more formal presentation of otherwise vague ideas. Also, I am grateful to him (and partly to Marr's book), for pushing me to look beyond the algorithmic aspects of vision and, instead, to pay attention to the computational structure of the problem.

To my friend-and-master Professor Pietro Perona I owe a debt of gratitude for the huge amount of time he spent with me on discussing all sorts of topics in vision, for initiating me to the use of many software packages, and for being always there to fix all the bugs and problems which were continously arising during my sessions at the terminal. It was both a pleasure and a very advantageous experience for me to benefit from his extraordinary professional skills and competence and his generous, communicative way of dealing with people.

I'd like to thank Kathleen O'Sullivan, Sheila Hegarty, Betty Lou McClanahan and all the other administrative staff for making LIDS such a pleasant place to work at.

I am very grateful to Giampiero Sciutto for his continous and valuable support with the computer facility.

Contents

Chapter 1

Introduction

1.1 Image Segmentation: Edge Detection and Image Partitioning

Image segmentation ia one of the most widely studied problems in vision. It is generally agreed upon that image segmentation is a necessary first step for most goals of vision.

We can formulate the problem of image segmentation according to two different paradigms: edge detection and region-based segmentation. In the *edge detection* approach the goal is to identify locations in the image where abrupt changes in image properties — such as brightness and texture — are present. Edges found in this way are not required to be closed curves and usually are detected by means of local operators. These operators work (conceptually) in two steps: 1) They smooth the image to some extent in order to attenuate the noise and 2) Compute some sort of spatial derivative (gradient, laplacian) to detect the presence of a sharp discontinuity. For a theory of edge detection see [34, 35]. A popular edge detection algorithm is Canny's [5]. While most of the edge detection schemes have been developed for *brightness* image segmentation there is also some work done on *texture* edge detection [33, 54, 29].

On the other hand, *region based* image segmentation aims at partitioning the

image into a set of homogenoeus regions. Therefore, the resulting boundaries are closed curves. The simplest way to do this is, for instance, to classify each pixel independently into one of two sets depending on whether its gray level is above or below a given threshold. Else, whole blocks of pixels can be classified together according to more global properties.

For instance, in [58, 49] classification of pixel blocks is performed at a very coarse scale to take advantage of the amount of statistics which can be gathered from large regions of the image. Then boundary resolution is improved by a coarse-to-fine strategy implemented in a pyramid datastructure.

In *pyramid linking* [42, 21, 30], which is another region based method, nodes associated with blocks of pixels are "linked" together whenever they share similar image properties. To compute these properties a smart averaging procedure is adopted which merges information only between blocks of pixels which are linked together. In this way averaging over dissimilar regions is avoided. Linking and averaging are alternated repeatedly until equilibrium is reached.

A similar tecnique (*split and merge*) is to recursively split dishomogeneous regions into smaller components and merge similar regions together until no more changes are needed [38, 22, 6, 25, 46, 45].

In *region growing* methods a region is progressively expanded until its homogeneity would no longer be mantained by a further expansion [62, 18, 31].

It is not clear which of edge detection and image partitioning is better. The choice may depend on the domain of application and in general the two methods can be used in conjunction [39] or treated in an unified fashion [20].

A drawback of image partitioning is that the properties of every region (brightness, texture) must be discontinuous enough at their boundaries in order for a reasonable partition of the image to exist, let alone be found. This assumption does not always hold. In any case, edge detection schemes based on oriented filters can acheive very high signal to noise ratios by accumulating "signal" along boundaries. This makes possible the extraction of very detailed and accurate information which can not be recovered by purely region based methods.

On the other hand, if the above assumption holds to a sufficient degree, it may be useful to constraint the output of image segmentation to be a set of regions, since this format of representation is easier to be used by successive stages of processing. Indeed, most existing vision systems rely on an image partitioning scheme.

1.2 Image Models

Images are noisy projections of the real world and therefore possess to a certain degree the same regularities of the physical world [34]. Indeed, the usefulness of image segmentation comes from the assumption that portions of the image corresponding to the same surface or object are in some sense homogeneous while, most of the time, sharp changes occur between abutting objects.

In making this more precise, edge detection and region based image segmentation enphasize different aspects of modeling, namely, while the former enphasizes the need of good models for discontinuities, the latter depends strongly on good models of homogeneous regions. Discontinuities in brightness, for instance, can be modeled by combimations of step, impulse, ramp and roof waveforms [41].

Let us turn our attention to some models of homogeneous regions. The simplest model of a homogeneous (not textured) region is a constant function. Thus the whole image can be described as a "piecewise constant" function. To take into account the dishomogeneity of the illumination and the variations in the surface orientation "piecewise smooth" functions can be used. For instance, "piecewise polynomial" functions have been employed [20, 31, 25].

Variational approaches to image segmentation [37, 1, 47] embody the piecewise smooth assumpution by evaluating the departure from smoothness locally by means of differential operators and then trying to minimize the output of these operators by introducing boundaries at the appropriate locations.

Modeling of textured regions is a much more complex problem and many frameworks have been proposed. According to the *structural* approach [19] a textured region is created by the regular repetition of a primitive element across the image.

Many different primitives have been proposed: blobs [3, 54], edges [53, 14], peaks and ridges [10], run length primitives [15], Voronoi polygons [51], Julesz's textons [26, 27].

In the statistical approach texture is described by the joint probability distribution (or moments thereof) of neighbouring pixels. For instance the following have been used: co-ocurence matrices [6, 11], correlation [7], gray level difference [55]. Unser has proposed to use the projection of the N-th order probability density on appropriatetely chosen axis. These projection can be computed by using linear filters [52].

Texture can be also characterized by the Fourier spectrum [2], or spatially localized Fourier-like transforms [46] such as Gabor filters [4, 8], Wigner distributions [46, 45], prolate sferoidal sequences [58, 57], or by "texture energy masks" [23, 24, 32, 43]. Filtering methods based on models of the human brain have been proposed [33].

Images can also be modeled as two dimensional random functions. While it is difficult to think of a unique model which describes all properties of images, probabilistic models which capture some of the basic properties of images have been proposed. For instance we can model an ensamble of images made of brightness uniform regions by assigning high probablity to images in which neighbouring pixels have the same gray level [13, 17, 36]. This can be done by using the language of statistical mechanics. An energy function can be defined on all possible images so that interaction among neighbouring pixels encourage them to be in the same state. The resulting probability density is called *Gibbs distribution*.

Models constructed in this way are equivalent to *Markov Random Fields*. A Markov Random Field is a random function defined by specifying the conditional probabilities of the state of a pixel on the state of neighbouring pixels. They have been used to model textured images [9, 13, 12]. In these models, the parameters which specify the probabilistic dependency of a pixel on neighbouring pixels are used as a "signature" of the texture type and therefore can be used to discriminate among different textures.

Autoregressive models have also been used to model texture [48, 25, 28, 29, 50].

1.3 Image Segmentation with Probablistic Models

A number of image segmentation algorithms have been developed within the probablistic setting. In [17], it is assumed that the observed image is a corrupted version of an image drawn from an ensemble of piecewise constant images with a known set of possible gray levels. Also, it is assumed that the probability density of the noise-free image (the "a priori" density) and the conditional distribution of the observed image on the original one are known. The problem then reduces to finding the image which maximizes the "a posteriori" density (MAP). Simulated annealing is employed to this purpose.

In [13, 12] a random function which describes texture is embedded in another random function which specifies the partition of the image into homogenoeus regions. The parameters of the texture random function are switched by the second function. The problem is then to find the partition of the image which maximizes the MAP distribution of the second function given the input image. A dynamic programming tecnique is used to acheive this maximum.

The scheme proposed in [25] allows for different types of models (e.g. polynomial and stochastic) to be present in the image. A decision rule is used to classify pixels into one of a fixed set of models.

In [16] an energy function is defined by using a disparity measure between neighbouring blocks of pixels. Large disparity encourage intervening boundaries and distinct partition labels. Differently from previous Bayesian methods, the set of possible models is not fixed a priori. However, some kind of ad-hoc training is needed to set a number of thresholds. Also, an approximate estimate of the number of regions and the scale parameter must be preset externally.

1.4 Multiscale Image Segmentation

Scenes of the world contain objects of many sizes containing features of many sizes. Also, objects can be viewed at several distances. Therefore, the images that we see

contain features at different scales. For example, the coat of a zebra can be described at several different scales — at the finest level the individual hairs are visible, at the next level collections of hairs of the same color generate a number of stripes, and at a still coarser level the stripes appear to belong to a larger regular "textured" pattern (from [14]). Thus, a useful scheme for image segmentation must be able to deal with information extracted at several scales.

In edge detection methods based on linear filtering, the scale parameter is given by the size of the filter. The outputs of filters having different scales can be put together to generate a unique representation embedded in the so called *scale-space* [59]. As one increases the size of the filter the smoothing increases and the number of edges decreases [60]. Unfortunately, while on one hand this allows more global information to be extracted, on the other hand it makes the location of boundaries poorer and poorer making it impossible to retrieve interesting features such as corners and T-junctions.

Similarly, variational approaches have been shown unable to retrieve corners and T-junctions. A solution to this problem within the variational framework has been proposed in [47] where it has been shown that in the microscopic scale limit the optimal solution converges to the discontinuity set of the image.

Another method to generate coarse description of the image without deteriorating the localization of boundaries is by means of anisotropic diffusion [40]. The basic idea is to smooth the image iterativelly using local averages where the brightness gradient is not too big while averaging is inhibited at location where the gradient is too strong, i.e. where an edge is very likely to exist.

The choice of an appropriate scale of segmentation is related to the uncertainty principle [56]. That is, a large scale allows a good "class localization" (as long as we do not mix statistics from different regions) but deteriorates the spatial information about the localization of boundaries. In other words, it is possible to rely on accurate measurment of the parameters of the image model (say, the average brightness) only at the price of blurring the boundaries. On the other hand, a small scale improves the boundary resolution but precludes the detection of more global patterns. To overcome

the limitation due to this uncerainty principle the scheme proposed in [58, 49] adopts a pyramidal datastructure. Detection of the model parameters is done at a fixed coarse scale but the final spatial localization of boundaries is determined at the finest scale of the pyramidal representation.

However, the uncertainty principle is not the only issue in choosing a scale for segmentation. Another problem comes from the intrinsic ambiguity in defining what a good segmentation should be. In fact, a given location of the image can be described as belonging to different regions depending on the scale of observation. For instance, a hair in the coat of a zebra can be a region by itself, or belong to a stripe or belong to the black and white striped pattern of the coat. Therefore an ideal segmentation scheme should specify all the scales at which a "reasonable" segmentation exists, along with generating the corresponding segmentation.

1.5 Contribution of this thesis

The goals of this thesis are 1) Propose an image representation scheme based on *two* scale dimensions, namely, the *feature scale* λ and the the *statistical scale* L. 2) Introduce an unsupervised image partitioning method which selects the "best scale" of segmentation at each location.

Many image segmentation algorithms using classification or maximization of the MAP distribution allow only for a few number of region types [25, 13, 12, 17]. To avoid this limitation one can try to detect clusters in feature space [58]. Another approach is to compare image parameters measured from neighbouring regions either to find abrupt changes (edge detection) or to discover homogeneous region (see for instance [16]). It is this last approach that we intend to follow in this thesis.

A limitation of most image segmentation algorithms is that they fix a priori a scale parameter so that only edges or regions within a given range of scales are discovered. In the approach suggested in this thesis a "best" scale of segmentation is assumed to exist at each location but no restriction are posed a priori on it[1]. This means, for

[1]Note that this is not the most general point of view. As we mentioned before, more than one

instance, that we should be able to detect a small white spot on black background with high spatial resolution and, at the same time, be able to detect the coarse boundary between two macrotextures.

A way to do this is to evaluate the "homogeneity" of several subregions having different sizes and to choose at each location the scale having the highest homogeneity. It is also possible to add a bias which favours larger regions, since a partition into a small number of regions is usually desirable.

Since we do not want to commit ourselves to any particular scale, we should employ an image representation embedded in a scale-space. In this way features at all scales can be represented explicitely and can contribute to the segmentation process.

However, in order to evaluate homogeneity reliably at different scales we claim that it is necessary to add another scale dimension to the standard scale-space, namely the *statistical scale*. Roughly speaking, this means that each parameter measured from the image should be averaged spatially over windows of several sizes, or, in other words, every parameter should be analyzed at several statistical scales.

The assumption that underlies the introduction of the statistical scale is that a homogeneous region is a realization of an *ergodic* random function. This means that when one computes the spatial average of any quantity over larger and larger windows, the result converges to the ensemble average of that quantity and therefore becomes independent of the position of the averaging window. Thus the dishomogeneity function, which is computed by comparing averages from neighbouring windows, goes to zero as the statistical scale goes to infinity. This makes possible to detect homogeneous regions of any size, provided that the size of the region is big enough to make the ergodic limit a reasonable approximation at some statistical scale.

interesting scale might exist at the same location and it might be worth to find them all.

Chapter 2

Multiscale Representations for Piecewise-Ergodic Images

2.1 Homogeneity and the Statistical Scale

What is a homogeneous region ? When dealing with noisy and textured images this is a far from trivial question. Intuitively, a region is homogeneous if its properties do not change from place to place. However, this depends on the (statistical) scale at which these properties are evaluated. For instance, the coat of a zebra is not homogeneous at a scale equal to the size of a stripe. In fact, if we observed the coat through a window of this size then, by translating the window across the image we would observe dramatic fluctuations of the coat's color. Instead, if the observation window is big enough, say three or four stripes large, and we averaged our measurements over this window, then our observations would be stationary under translation of the window.

The following sections will be devoted to give precise definitions of the statistical scale and of homogeneity. First, we will define the class of piecewise ergodic images.

2.2 Piecewise Ergodic Images

In this section we define the class of piecewise ergodic images. This definition will be useful later to show that for images in this class, in the one dimensional case, and in

the limit of infinitely large regions, the discontinuity set of the image can be exactly recovered.

For simplicity, we restrict ourselves to one dimensional "images". Then an image g is a real function defined on the real line: $g : \mathbf{R} \to \mathbf{R}$. We will denote by G the set of all images.

For $x \in \mathbf{R}$ and $\lambda > 0$ let $T_x : G \to G$ and $D_\lambda : G \to G$ be the translation and dilation operators respectively defined by:

$$(T_{x_0} g)(x) = g(x - x_0) \tag{2.1}$$

$$(D_\lambda g)(x) = g(\lambda^{-1} x) \tag{2.2}$$

We require that the set of images be invariant under translation and dilation, i.e. for $x \in \mathbf{R}$ and $\lambda > 0$ we assume: $T_x D_\lambda G = G$.

A *compact support functional* f on G is a real valued function defined on any $g \in G$, such that $f(g)$ depends only on the restriction of g to some interval $[x_1, x_2]$:

$$f(g) = f(g_{|[x_1, x_2]}) \tag{2.3}$$

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and let ψ be a stationary random function in the set of images: $\psi : \Omega \to G$. Then we say that ψ is *ergodic* if, for any compact support functional f, and for almost all $\omega \in \Omega$:

$$\lim_{X \to \infty} \frac{1}{2X} \int_{-X}^{X} f(T_{-x} \psi(\omega)) \mathrm{d}x = \mathrm{E}\left[f(\psi(\omega)) \right] \tag{2.4}$$

That is, ψ is ergodic if the spatial average of any functional converges to the ensemble average when the averaging window becomes infinitely large.

Now, let $\{\psi_0, \ldots, \psi_{N-1}\}$, $\psi_i : \Omega \to G$ be a collection of ergodic random functions. Then a piecewise ergodic random function with *discontinuity set* $D = \{d_0, \ldots, d_N\}$,

$d_i < d_{i+1}$, is the random function:

$$\phi = \sum_{i=0}^{N-1} \psi_i \chi_{[d_i, d_{i+1}]} \tag{2.5}$$

where $\chi_{[d_i,d_{i+1}]}$ is the indicator function of the interval $[d_i, d_{i+1}]$. Also, we say that the image g is piecewise ergodic with discontinuity set D if $g = \phi(\omega)$, i.e. if it is a realization of a piecewise ergodic random function with discontinuity set D.

2.3 Image Representations in Scale-Scale-Space

Given an image $g \in G$, we want to define a description of it embedded in a "scale-scale-space" $\mathbf{R}_L^+ \times \mathbf{R}_\lambda^+ \times \mathbf{R}_x$ containing *two* scale dimensions: feature scale and statistical scale. First, we will define a description embedded in a "standard" scale-space $\mathbf{R}_\lambda^+ \times \mathbf{R}_x$ which contains only the feature scale dimension. This description is obtained by applying an image descriptor to translated and scaled versions of the input image g.

An *image descriptor* a is a functional on G with support on $[-\frac{1}{2}, \frac{1}{2}]$:

$$a : G \to \mathbf{R}, \qquad a(g) = a(g_{|[-\frac{1}{2},\frac{1}{2}]}) \tag{2.6}$$

Then, if a is an image descriptor we define the description function $A_g : \mathbf{R}_\lambda^+ \times \mathbf{R}_x \to \mathbf{R}$ as:

$$A_g(\lambda, x) = a(D_{\lambda^{-1}} T_{-x} g) = K_a(\lambda) a(T_{-\lambda x} D_{\lambda^{-1}} g) \tag{2.7}$$

where $K_a(\lambda)$ is a suitable normalization factor which may depend on a.

All wavelet representation can be expressed in this way. In order for (2.7) to be a wavelet expansion it is sufficient (and necessary) that 1) a be a linear functional[1] satisfying the wavelet equation (see [61] for more details) 2) λ, x be discretized in the right way.

Note that $A_g(\lambda, x)$ depends only on $g_{|[x-\frac{\lambda}{2}, x+\frac{\lambda}{2}]}$.

The description of g in "scale-scale-space" is the function $\mathbf{A}_g : \mathbf{R}_L^+ \times \mathbf{R}_\lambda^+ \times \mathbf{R}_x \to \mathbf{R}$

[1] In general a need not be a compact support functional but is sufficient that it vanish for $|x| \to \infty$.

defined, for $L > \lambda$, by:

$$\mathbf{A}_g(L,\lambda,x) = \frac{1}{L-\lambda}\int_{x-\frac{L-\lambda}{2}}^{x+\frac{L-\lambda}{2}} A_g(\lambda,x')\mathrm{d}x' \tag{2.8}$$

For $L \leq \lambda$ the function is undefined (or can be set to zero).

Note that he average is taken over all x' such that $[x'-\frac{\lambda}{2}, x'+\frac{\lambda}{2}] \subset [x-\frac{L}{2}, x+\frac{L}{2}]$ and therefore $\mathbf{A}_g(L,\lambda,x)$ depends only on the restriction of g to the interval $[x-\frac{L}{2}, x+\frac{L}{2}]$. This interval will be referred to as the *averaging window* of $\mathbf{A}_g(L,\lambda,x)$.

We can generalize the definition to averages of "higher order". That is, for $p \geq 1$ let:

$$\mathbf{A}_g^p(L,\lambda,x) = \left[\frac{1}{L-\lambda}\int_{x-\frac{L-\lambda}{2}}^{x+\frac{L-\lambda}{2}} [A_g(\lambda,x')]^p \,\mathrm{d}x'\right]^{\frac{1}{p}} \tag{2.9}$$

For $p = \infty$ we obtain:

$$\mathbf{A}_g^\infty(L,\lambda,x) = \max_{x-\frac{L-\lambda}{2} \leq x' \leq x+\frac{L-\lambda}{2}} A_g(\lambda,x') \tag{2.10}$$

Let us consider a set of image descriptors $\{a_k\}$ and let $\left\{A_g^k(\lambda,x)\right\}$ and $\left\{\mathbf{A}_g^{k,p}(L,\lambda,x)\right\}$ be the corresponding descriptions.

If the image g is a realization of an ergodic random function, ie. if $g = \psi(\omega)$, ψ ergodic, then for the definition of ergodicity it follows, for almost all $\omega \in \Omega$:

$$\lim_{L\to\infty} \mathbf{A}_g^{k,p}(L,\lambda,x) = \lim_{L\to\infty} \mathbf{A}_{\psi(\omega)}^{k,p}(L,\lambda,x) = \left[\mathrm{E}\left[A_g^k(\lambda,x)\right]^p\right]^{\frac{1}{p}} \equiv s_\psi(k,\lambda,p) \tag{2.11}$$

Therefore, for a given set of image descriptors $\{a_k\}$, the function s_ψ of (k,λ,p) may be used as the "signature" of the ergodic random function ψ. Note that, after the limit $L \to \infty$ is taken, (2.11) is independent of the spatial position x.

2.4 The Dishomogeneity Function

Let $g \in G$ and let $\{a_k\}$ be a set of descriptors. First, let us define the dishomogenity at location x and scale L for a given choice of (k,λ,p). This can be done by evaluating

the range of variation of the description $\mathbf{A}_g^{k,p}(L,\lambda,x')$ within the interval $[x-L, x+L]$. To do this we define:

$$\mathrm{M}_g^{k,p}(L,\lambda,x) = \max_{x-\frac{L}{2}\le x'\le x+\frac{L}{2}} \mathbf{A}_g^{k,p}(L,\lambda,x') \tag{2.12}$$

$$\mathrm{m}_g^{k,p}(L,\lambda,x) = \min_{x-\frac{L}{2}\le x'\le x+\frac{L}{2}} \mathbf{A}_g^{k,p}(L,\lambda,x') \tag{2.13}$$

Note that, since $\mathrm{M}_g^{k,p}(L,\lambda,x)$ and $\mathrm{m}_g^{k,p}(L,\lambda,x)$ are computed for $x' \in [x-\frac{L}{2}, x+\frac{L}{2}]$ and since $\mathbf{A}_g^{k,p}(L,\lambda,x')$ depends only on $g_{|[x'-\frac{L}{2},x'+\frac{L}{2}]}$, we have that $\mathrm{M}_g^{k,p}(L,\lambda,x)$ and $\mathrm{m}_g^{k,p}(L,\lambda,x)$ depend only on $g_{|[x'-L,x'+L]}$.

Also note that the leftmost averaging window contributing to (2.12),(2.13) is $[x-L, x]$ and the rightmost is $[x, x+L]$. That is, the extrema in (2.12),(2.13) are computed over all averaging windows lying between these two adjacent nonoverlapping windows.

Then we define the dishomogeneity $v_g^{k,p}(L,\lambda,x)$ as:

$$v_g^{k,p}(L,\lambda,x) = \mathrm{M}_g^{k,p}(L,\lambda,x) - \mathrm{m}_g^{k,p}(L,\lambda,x) \tag{2.14}$$

Of course we have $v_g^{k,p}(L,\lambda,x) \ge 0$. Also from (2.11) it follows that if g is a realization of an ergodic random function, $g = \psi(\omega)$, ψ ergodic, then $\lim_{L\to\infty} v_g^{k,p}(L,\lambda,x) = 0$ for almost all $\omega \in \Omega$.

Finally the dishomogeneity at location x and statistical scale L is given by the most dishomogeneous "channel":

$$v_g(L,x) = \max_{(k,\lambda,p)} v_g^{k,p}(L,\lambda,x) \tag{2.15}$$

Again, $v_g(L,x) \ge 0$. Also, if $g = \psi(\omega)$, ψ ergodic, and if the ergodic limit (2.11) is uniformly attained over[2] (k,λ,p) then $\lim_{L\to\infty} v_g(L,x) = 0$ for almost all $\omega \in \Omega$.

[2]Uniformity over x is guaranteed by the stationarity of g.

Chapter 3

Computing a Partition of the Image

In the previous chapter we have described a way of representing images at several scales. Now we turn to the problem of finding an "optimal" partition of the image. In the next two sections we show that if the image g is piecewise ergodic and if the regions are large enough then the discontinuity set of g can be recovered by minimizing an appropriate function.

3.1 Asymptotic Recovery of the Discontinuity Set

Let $g \in G$ and let the support Σ of g be compact. Then a partition of g is identified by a set of consecutive *boundary points* $B = \{b_0, \ldots, b_M\}$, $b_j < b_{j+1}$, such that $[b_0, b_M] = \Sigma$. Note that if $g = \sum_{i=0}^{N-1} \chi_{[d_i, d_{i+1}]} \psi_i(\omega)$ then we must have: $b_0 = d_0$ and $b_M = d_N$.

Given an image g, we define the dishomogeneity of a partition B of g as:

$$v_g(B) = \sum_{j=0}^{M-1} \frac{b_{j+1} - b_j}{b_M - b_0} v_g \left(\frac{b_{j+1} - b_j}{2}, \frac{b_{j+1} + b_j}{2} \right) \tag{3.1}$$

where the function v is the dishomogeneity function (2.15).

Let g be a piecewise ergodic image: $g = \sum_{i=0}^{N-1} \chi_{[d_i, d_{i+1}]} \psi_i(\omega)$. Then we define the

γ-dilation of g as:

$$g_\gamma = \sum_{i=0}^{N-1} \chi_{[\gamma d_i, \gamma d_{i+1}]} \psi_i(\omega) \tag{3.2}$$

Similarly, for any partition B of g define the γ-dilation of B as:

$$B_\gamma = \{\gamma b_0, \ldots, \gamma b_M\} \tag{3.3}$$

We say that the ergodic random function ψ is *well behaved* for the set of descriptors $\{a_k\}$ if the image descriptions of $\psi(\omega)$ generated by $\{a_k\}$ attain the ergodic limit (2.11) *uniformly* over (k, λ, p)[1] for almost all $\omega \in \Omega$. Also, the piecewise ergodic random function $\phi = \sum_{i=0}^{N-1} \chi_{[d_i, d_{i+1}]} \psi_i$ is said to be well behaved if all the ψ_i are.

Theorem 1 *Let $\{a_k\}$ be a set of image descriptors such that $\sup_{g \in G} a_k(g) < \infty$. Let $\phi = \sum_{i=0}^{N-1} \chi_{[d_i, d_{i+1}]} \psi_i$ be a piecewise ergodic random function with discontinuity set $D = \{d_0, \ldots, d_N\}$ and assume ϕ is well behaved under $\{a_k\}$. Let $g = \phi(\omega)$. Let B be a partition of g. Then it holds for almost all $\omega \in \Omega$:*

1) The limit $\lim_{\gamma \to \infty} v_{g_\gamma}(B_\gamma)$ exists. Define $v_g^\infty(B) = \lim_{\gamma \to \infty} v_{g_\gamma}(B_\gamma)$.

2) If $B \supset D$ then $v_g^\infty(B) = 0$.

3) Assume that for each $i_1 \neq i_2$ there exist i, λ and p, $1 \leq p < \infty$ such that $s_{\psi_{i_1}}(k, \lambda, p) \neq s_{\psi_{i_2}}(k, \lambda, p)$. Then $B \not\supset D \Longrightarrow v_g^\infty(B) > 0$

Comment. The theorem says that in the limit of infinitely large regions, and if the signatures of any two regions are different (i.e. if they differ for at least one value of (k, λ, p)) then the dishomogeneity of a partition is minimized by all the finite partitions which contain the discontinuity points of g.

Proof: 1) From the definition (3.1) of $v_g(B)$ and from the corollary 2 of lemma 1 we have:

$$\lim_{\gamma \to \infty} v_{g_\gamma}(B_\gamma) = \sum_{j=0}^{M-1} \frac{b_{j+1} - b_j}{b_M - b_0} v_{g_\gamma}\left(\gamma \frac{b_{j+1} - b_j}{2}, \gamma \frac{b_{j+1} + b_j}{2}\right) = \tag{3.4}$$

[1]Note that uniformity over λ requires that for $\lambda \to \infty$, $A_g^k(\lambda, x)$ be constant as a function of x. This is obtained if, for instance, $\lim_{\lambda \to \infty} A_g^k(\lambda, x) = 0$.

$$= \sum_{j=0}^{M-1} \frac{b_{j+1} - b_j}{b_M - b_0} v_g^\infty \left(\frac{b_{j+1} - b_j}{2}, \frac{b_{j+1} + b_j}{2} \right) \qquad \text{Q.E.D.} \tag{3.5}$$

If we let $L_j = \frac{b_{j+1} - b_j}{2}$ and $x_j = \frac{b_{j+1} + b_j}{2}$, $0 \leq j \leq M-1$ then this can be rewritten as

$$\lim_{\gamma \to \infty} v_{g_\gamma}(B_\gamma) = v_g^\infty(B) = \sum_{j=0}^{M-1} \frac{b_{j+1} - b_j}{b_M - b_0} v_g^\infty (L_j, x_j) \tag{3.6}$$

2) Since $B \supset D$, for each j, $0 \leq j \leq M-1$, there exists i_j, $0 \leq i_j \leq N-1$, such that $[x_j - L_j, x_j + L_j] = [b_j, b_{j+1}] \subset [d_{i_j}, d_{i_j+1}]$. Then, applying corollary 3 to (3.6) yields $v_g^\infty(B) = 0$. Q.E.D.

3) Since $B \not\supset D$, there exist i,j, $0 \leq i \leq N-1$, $0 \leq j \leq M-1$, such that $b_j < d_i < b_{j+1}$. Then, using the notation of lemma 2, $\Delta([b_j, b_{j+1}], D) > 0$ and from (A.19) it follows that $v_g^\infty(L_j, x_j) > 0$ and from (3.6) we have $v_g^\infty(B) > 0$. Q.E.D.

3.2 Regularization of the Dishomogeneity Function

Theorem 1 shows that dishomogeneity is minimized by an infinite number of partitions obtained by adding an arbitrary set of points to the discontinuity set of g. Formulations of problems which do not yield a unique solution are often referred to as *ill posed* [44, 36]. A way to make the solution unique is to "regularize" the cost function by adding a suitable extra term to it. Usually, this term can be chosen so as to constrain the solution to be "smooth" or "simple" in some sense. For instance, in variational approaches to image segmentation [37, 47] one of the terms in the cost function is proportional to the length of the boundary. As a consequence the amount of boundary produced is kept at a minimum.

Similarly, in our case we can "regularize" the problem by adding a "complexity" term to the dishomogeneity function:

$$c_g(B) = v_g(B) + \alpha S(B) \tag{3.7}$$

If we choose $S(B)$ to be just the number of points in the set B then, under the same conditions of theorem 1 we can prove that the minimizer of the asymptotic cost $c_g^\infty(B) = v_g^\infty(B) + \alpha S(B)$, for α sufficiently small, is unique and equal to the discontinuity set of g.

Theorem 2 *Let the assumptions of theorem 1 hold (including the assumption of part 3). Let* $c_g^\infty(B) = v_g^\infty(B) + \alpha \cdot \mathrm{card}(B)$. *Then there exists* $\alpha_0 > 0$ *such that for* $0 < \alpha < \alpha_0$ *the minimizer of* $c_g^\infty(B)$ *is unique and equal to* D.

Proof: From theorem 1 and the definition of $c_g^\infty(B)$ it is clear that if $\alpha > 0$, $B \neq D$ and $\mathrm{Card}(B) \geq \mathrm{Card}(D)$ then $c_g^\infty(D) < c_g^\infty(B)$. Then we must show that there exists $\alpha_0 > 0$ such that $c_g^\infty(D) < c_g^\infty(B)$ for $0 < \alpha < \alpha_0$ and for all B with $\mathrm{Card}(B) < \mathrm{Card}(D)$. Let $M = \mathrm{Card}(B) - 1$ and $N = \mathrm{Card}(D) - 1$ as before and assume $M < N$. Let $2\delta = \min_{0 \leq i \leq N-1} (d_{i+1} - d_i)$. Then note that since $M < N$ there exists a $d_i \in D$, $i \neq 0$, $i \neq N$ such that $[d_i - \delta, d_i + \delta]$ does not contain any point of B. Then, there exists a b_j, $0 \leq j \leq M - 1$, such that $[d_i - \delta, d_i + \delta] \subset [b_j, b_{j+1}]$. Therefore $\Delta([b_j, b_{j+1}], D) > \delta$, where Δ is as defined in lemma 2. Then by lemma 2

$$v_g^\infty(L_j, x_j) > K_g \frac{\delta}{L_j N} > K_g \frac{2\delta}{(d_N - d_0)N} \tag{3.8}$$

where $L_j = \dfrac{b_{j+1} - b_j}{2}$ and $x_j = \dfrac{b_{j+1} + b_j}{2}$. From (3.6) we have: $v_g^\infty(B) > K_g \frac{2\delta}{(d_N - d_0)N}$. Finally, by choosing $\alpha_0 = K_g \frac{2\delta}{(d_N - d_0)N(N+1)}$ we have for $0 < \alpha < \alpha_0$:

$$c_g^\infty(D) = \alpha(N+1) < K_g \frac{2\delta}{(d_N - d_0)N} < v_g^\infty(B) < c_g^\infty(B) \qquad \text{Q.E.D.} \tag{3.9}$$

In actual implementations, i.e. when the size of the regions is finite, the parameter α must be chosen to yield an acceptable tradeoff between having an overfragmented segmentation (low α) and missing some of the discontinuity points (high α).

An alternative way to measure the complexity is by means of the entropy of the

partition defined as:

$$S(B) = -\sum_{j=0}^{M-1} \frac{b_{j+1} - b_j}{b_M - b_0} \log \frac{b_{j+1} - b_j}{b_M - b_0} \tag{3.10}$$

Note that if the partition were made of M equally large regions, then (3.10) would reduce to $S(B) = \log M = \log(\text{Card}(B) - 1)$. The reason to use this definition of complexity will become clearer later.

3.3 Scale Adaptation within a Homogeneous Region

In the definition of dishomogeneity given in (3.1) the dishomogeneity of each region $R_j = [b_j, b_{j+1}]$ was computed at a statistical scale equal to half the size of the region. Indeed we can rewrite (3.1) as

$$v_g(B) = \sum_{j=0}^{M-1} v_g(R_j) \tag{3.11}$$

where $v_g(R_j) = v_g(L_j, x_j)$ and $L_j = \frac{b_{j+1} - b_j}{2}$, $x_j = \frac{b_{j+1} + b_j}{2}$ ($v_g(L_j, x_j)$ is given by (2.15)).

In real applications this definition is not adequate for two reasons:

1) The dishomogeneity function (2.15) can be easily generalized to the two dimensional case by computing averages, maxima and minima over squares (or circles) of size $2L$. However, the dishomogeneity of a region which is not a square (or a circle) cannot be defined as easily as in the one dimensional case and a more elaborate definition is needed. For instance we can cover the region with square "patches" and consider the values of the function v on the elements of this covering. Then the problem arises of how to choose the size of these patches. In fact, on one hand large patches are preferable in order to approximate the ergodic limit at best but on the

other hand small patches are better at following the contour of the region[2].

2) In real images, slowly varying distorsions are often present which cause otherwise ergodic regions to be non stationary. Two examples may be a tilted textured surface and a surface of a uniform color under inhomogeneous illumination. In these situations, if we let the statistical scale go to infinity, the dishomogeneity also goes to infinity instead of vanishing. This is because very distant locations in the image have very different properties. Therefore, even if the region is arbitrarily large, the "optimal" statistical scale may be finite. Thus, it might be useful to let the scale of each region be a free parameter and optimize over it instead of assuming that the largest scale is always the best.

Going back to one dimensional images, we can define the dishomogeneity of region $R_j = [b_j, b_{j+1}]$ at statistical scale $L_j \leq \frac{b_{j+1}-b_j}{2}$ as:

$$v_g(L_j, R_j) = \max_{b_j + L_j \leq x' < b_{j+1} - L_j} v(L_j, x') \tag{3.12}$$

Note that maximization is carried over all x' such that $[x' - L_j, x' + L_j] \subset R_j$ and therefore $v_g(L_j, R_j)$ depends only on $g_{|R_j}$.

Then the dishomogeneity of a partition $B = \{b_0, \ldots, b_M\}$ at scales $\mathcal{L} = \{L_0, \ldots, L_{M-1}\}$, $L_j \leq \frac{b_{j+1}-b_j}{2}$ is given by:

$$v_g(B, \mathcal{L}) = \sum_{j=0}^{M-1} v_g(L_j, R_j) \tag{3.13}$$

Now, optimization of the cost function must be carried out over both the boundaries and the set of statistical scales $\mathcal{L}$. If we want to take into account the fact that a region which is *globally* ergodic is preferable to a region which is only *locally* ergodic then we can introduce a term $T(B, \mathcal{L})$ in the cost function which is monotonically

[2]This problem is an instance of the uncertainty principle. In fact, "class" localization acheived by approaching the ergodic limit must be traded off against boundary resolution. (A consequence of this is that large macrotexture, which exhibit ergodic properties only at large statistical scales can only have coarse boundaries.)

decreasing with the L_j's, for instance:

$$T(B,\mathcal{L}) = -\sum_{j=0}^{M-1} \frac{b_{j+1}-b_j}{b_M-b_0} \log \frac{2L_j}{b_{j+1}-b_j} \tag{3.14}$$

Then, by taking $S(B)$ to be the entropy of B the cost becomes:

$$\begin{aligned} c_g(B,\mathcal{L}) &= v_g(B,\mathcal{L}) + \alpha S(B) + \beta T(B,\mathcal{L}) \\ &= \sum_{j=0}^{M-1} \frac{b_{j+1}-b_j}{b_M-b_0} \left(v_g(L_j,R_j) - \alpha \log \frac{b_{j+1}-b_j}{b_M-b_0} - \beta \log \frac{2L_j}{b_{j+1}-b_j} \right) \end{aligned}$$

By letting for simplicity $\beta = \alpha$ we have:

$$c_g(B,\mathcal{L}) = \sum_{j=0}^{M-1} \frac{b_{j+1}-b_j}{b_M-b_0} \left(v_g(L_j,R_j) - \alpha \log \frac{2L_j}{b_M-b_0} \right) \tag{3.15}$$

3.4 A Local Formulation of the Problem

Before discussing an algorithm based on the above optimization criterium we want to rewrite the cost function as an integral of a cost density. In this way, the problem reduces to minimizing the cost density locally and becomes more tractable from a computational point of view.

Let the labeling function $j(x)$ be defined by $x \in R_{j(x)} = [b_{j(x)}, b_{j(x)+1}]$. Then (3.15) becomes:

$$c_g(B,\mathcal{L}) = \frac{1}{b_M-b_0} \int_{b_0}^{b_M} \left(v_g(L_{j(x)},R_{j(x)}) - \alpha \log \frac{2L_{j(x)}}{b_M-b_0} \right) \mathrm{d}x = \frac{1}{b_M-b_0} \int_{b_0}^{b_M} H_g(B,\mathcal{L},x)\mathrm{d}x \tag{3.16}$$

where

$$H_g(B,\mathcal{L},x) = v_g(L_{j(x)},R_{j(x)}) - \alpha \log \frac{2L_{j(x)}}{b_M-b_0} \tag{3.17}$$

In the above formulas, $\mathcal{L}$ represents the choice of the statistical scales for each region. So far, for the sake of simplicity, we have assumed that the statistical scale is constant within each region. However, this is not necessary and, indeed, our algorithm allows for the scale to vary also within each region.

Chapter 4

An Algorithm for Image Segmentation

4.1 Cost Minimization in an Overlapped Pyramid

Our algorithm can be divided in 5 steps:

- Computation of the scale-space description $A_g(\lambda, x)$
- Computation of the scale-scale-space description $\mathbf{A}_g(L, \lambda, x)$
- Computation of the dishomogeneity function and of the cost function (4.1)
- Optimization of (4.1)
- Connected component labeling

This algorithm has been implemented in the overlapped pyramid shown in figure 4-1. In the one dimensional case, each node of the pyramid represents an interval of the real line and the size of these intervals doubles when moving up in the pyramid. The two children nodes of a parent node represent nonoverlapping intervals having a boundary point in common. Adjacent nodes in a horizontal level (other than the bottom level) overlap. The fractional amount of overlap doubles at each level up to

Figure 4-1: A one dimensional overlapped pyramid with 32 pixels. Each node in the l-level of the pyramid represents a window of diameter 2^l lying on the image. This kind of pyramid is characterized by an integer parameter, Z (Z = 2 in this figure) which is the highest level of the pyramid sampled at 1-pixel intervals. That is, adjacent windows in the levels $0, \cdots, \mathrm{Z}$ are separated by a distance equal to the size of 1 pixel. From the Z-th level this distance doubles when one moves up by one level in the pyramid so that the fractional overlap between adjacent windows is always equal to $1 - 2^{-\mathrm{Z}}$ for the higher levels. Each node "stores": 1) a description of the image contained in that window; 2) a dishomogeneity value computed by using the image descriptions stored in the nodes at the next lower level; 3) a dishomogeneity cost.

the Z-th level of the pyramid. From the Z-th level up it is constant. (Z = 3 in our implementation).

The multiscale descriptions $A_g(\lambda, x)$ and $\mathbf{A}_g(L, \lambda, x)$ are computed for $L \geq \lambda$, L, λ equal to the sizes of nodes and at all positions corresponding to the centers of the nodes.

A cost is also associated with each node n. If we let $\mathcal{N}$ be the set of all nodes then the cost function $h_g : \mathcal{N} \to \mathbf{R}$ is given by (compare with (3.17)):

$$h_g(n) = v_g(L(n), x(n)) - \alpha \log \frac{L(n)}{X} \tag{4.1}$$

where X is the size of the image and $L(n)$ and $x(n)$ are defined by: $L(n) = x_2 - x_1$, $x(n) = \frac{x_1+x_2}{2}$ where $[x_1, x_2]$ is the interval corresponding to the node n ($n = [x_1, x_2]$ if we identify nodes with intervals).

The cost function (4.1) can be minimized at each location of the image by letting each node of the bottom level of the pyramid (i.e. each pixel) select the lowest cost node containing it.

That is, for each pixel ν_i we define n_i^* by:

$$h(n_i^*) = \min_{n \supset \nu_i} h(n) \tag{4.2}$$

The minimization in (4.2) realizes a multiscale local optimization because nodes of all sizes, and therefore descriptions at all scales, are compared independently at each pixel ν_i. In a sense, it is made of two "orthogonal" simultaneous optimizations, one "vertical" — which selects the most appropriate scale and maximizes class localization — and one "horizontal" — which favours nodes that do not contain boundaries, thus maximizing space localization.

The computation of n_i^* can be done very efficiently with a single top-down scan of the overlapped pyramid by using a very simple routing algorithm in which every node of the pyramid sends a "packet" to every pixel included in it.

Usually, unless slowly varying distorsions are present, the minimization (4.2) selects one of the nodes at the largest scale which does not contain a boundary. Oth-

erwise, smaller nodes can be activated (see figures T2b and T6).

At this point the segmentation is implicitely defined. To make it explicit it is necessary to label each pixel according to the region to which it belongs. This can be done in two steps.

First, the set of "active" nodes $\mathcal{N}^* = \{n_i^*\}$ is enlarged to the set $\overline{\mathcal{N}^*}$ which also contains nodes whose cost is not too much greater than some nearby node in $\mathcal{N}^*$. Therefore, $\overline{\mathcal{N}^*}$ should contain all the nodes of optimal size included in a uniform region.

Second, the set $\overline{\mathcal{N}^*}$ whose elements make up the optimal partitioning graph is split into connected components[1], which represent the regions of the segmentation. Each component is assigned a unique label, which is then propagated to all the linked pixels.

4.2 Implementation Details and Results of Computer Experiments.

In this section we describe the results of some computer experiments carried out with 1 dimensional syntethic images. The main purpose of these experiments is to show that a wide variety of images can be segmented correctly by the *same* implementation of the scheme. All the 37 trials shown in figures T1-T10 have been obtained by the same implementation, that is, no parameter has been adjusted from one trial to another.

In this implementation only one image descriptor has been used given by:

$$a(g) = (F_1 * g)(0) = -\int_{-\frac{1}{2}}^{0} g(x)\mathrm{d}x + \int_{0}^{\frac{1}{2}} g(x)\mathrm{d}x \tag{4.3}$$

[1] A topology based on overlap is used to define connectivity. Two nodes of the graph, are said to be connected if their overlap is more than a given threshold — say 50%.

Figure 4-2: The filter F_λ used in our implementation.

The filter F_λ is shown in figure 4-2. The image description $A_g(\lambda, x)$ is then:

$$A_g(\lambda, x) = \frac{1}{\lambda}(F_\lambda * g)(x) = \frac{1}{\lambda}\left[-\int_{x-\frac{\lambda}{2}}^{x} g(x)\mathrm{d}x + \int_{x}^{x+\frac{\lambda}{2}} g(x)\mathrm{d}x\right] \tag{4.4}$$

The description in scale-scale-space has been computed for two values of p: $p = 2$ and $p = \infty$. In our computer experiments the term with $p = \infty$ has proved to be very helpful in improving the spacial resolution of boundaries, especially at brightness discontinuities.

The coefficient α in the cost expression (4.1) has been chosen to be proportional to the "maximum contrast" in the image defined as: $C_g = \max_{\lambda,x} |A_g(\lambda, x)|$. (In this implementation $\alpha = C_g$). In this way, the two terms in (4.1) have the same "physical" dimensions. Moreover, it is easy to verify that $h_{c_1 g + c_2} = c_1 h_g$ and therefore the minimization of the cost function is not affected by homogeneous illumination changes of the form: $g \to c_1 g + c_2$.

The results of the experiments are shown in figures T1-T10. All images have 1024 pixels apart from T5c,d and T8c which have 2048.

For each trial the input function g is shown. In T1-T10 the axis are not shown for clarity. The ticks below the graph of g correspond to the boundaries of the regions

of the output segmentation. The boundaries of the segmentation are also marked on the graph of g with diamonds.

For some trials, the nodes of the optimal partitioning graph (that is the elements of $\overline{\mathcal{N}^*}$) are shown above the image. Each horizontal segment corresponds to such a node. The vertical displacement between segments has been introduced to make them distinguishable and has no other meaning. Each node which also belongs to $\mathcal{N}^*$ is marked with a heavier line in corrispondence of the pixels which have classified themselves into it.

The image in T1a contains 9 "white" regions of different sizes lying on a "black" background. Note that all regions, even the smallest ones, are detected and their boundaries are located with 1 pixel resolution. Also note that small nodes are activated near the boundaries while large nodes are selected in the middle of each region.

T1b shows the segmentation of an image containing 10 regions of different brightness. Only regions delimited by high contrast boundaries are detected. The boundary resolution is again of the order of 1 pixel.

T2a shows an image containing 5 regions having blurred boundaries. The tiny region delimited by two empty diamonds is only one pixel large. Apart from this mistake, all regions are detected and their boundaries are placed near the center of the corresponding discontinuity.

T2b illustrates the detection of abrupt intensity changes in presence of very strong gradients. Note that in this image (and also in T6) each region is only *locally* homogeneous and it is necessary to activate many small nodes in order to cover the whole region.

Figure T3a shows a quite complex image containing both "white" and "black" regions over a "grey" background. Regions differ in brightness and size and some have blurred boundaries. In T3b-e different kinds of perturbations are superimposed to the original image. In T3b some noise with uniform distribution has been added independently to every pixel (we say then that the correlation length of the noise is $\lambda_c = 1$). T3c contains an uniform illumination gradient. In T3d spatially extended perturbations are superimposed to the uniform regions. T3e contains an illumination

gradient and 4 perturbations (noise) with uniform distribution and correlation lengths of 1,10,37,121. The only mistake occurs in T3e in corrispondence of the second discontinuity, which is not discovered. This is due to the fact that the contrast of this discontinuity is of intermediate strength and therefore this boundary is unstable. Also, sometimes, the boundaries of blurred discontinuities are placed some distance away from its center. This is mainly due to the discreteness of the sampling in the scale-space.

Figures T4a-c show the detection of a bright region in highly perturbed conditions. Since here the ergodic limit is acheived only at a large scale, the region must be large in order to be detected. In T4a,b the boundary resolution is about 1 pixel while in T4c is of the order of 1 wavelength. In T4d the intensity of the noise varies across the image.

T5a,b show the segmentation of a high-frequency texture on a low frequency background. In T5c (T5d) the frequency (contrast) has been randomly modulated around the mean. A boundary is placed only at the abrupt frequency change.

In T6 the frequency changes linearly (within each region) simulating a perspective distorsion. Note that the large range of feature values in each region does not prevent it to be interpreted as a single region.

T7 and T8 show some examples of 1-dimensional texture segmentation. Both deterministic and stochastic textures are present.

Images in T9 contain both textured and non textured regions. In T9b,d some noise and an illumination gradient have been added. Note that sometimes a textured region is slightly extended onto the background.

Finally, T10 illustrates the scale invariance of the segmentation process.

Appendix A

Lemmas

Lemma 1 *Assume the same conditions of theorem 1 hold. Let x,L be such that* $[x_1, x_2] \subset \Sigma$ *where* $x_1 = x - \frac{L}{2}$, $x_2 = x + \frac{L}{2}$, *and* Σ *is the support of g. Let* i_1 *and* i_2 *be defined by:* $d_{i_1} = \max\{d_i \in D \ : \ d_i \leq x_1\}$; $d_{i_2} = \min\{d_i \in D \ : \ d_i \geq x_2\}$. *Let* $\hat{d}_{i_1} = x_1$, $\hat{d}_{i_2} = x_2$, *and* $\hat{d}_i = d_i$ *for* $i_1 < i < i_2$. *Then*

$$\lim_{\gamma\to\infty} \mathbf{A}^{k,p}_{g_\gamma}(\gamma L, \lambda, \gamma x) = \left[\sum_{i=i_1}^{i_2-1} \frac{\hat{d}_{i+1} - \hat{d}_i}{L} \left[s_{\psi_i}(k, \lambda, p)\right]^p\right]^{\frac{1}{p}} \equiv \mathbf{A}^{k,p}_{\infty,g}(L, \lambda, x) \tag{A.1}$$

and the limit is acheived uniformly over k, λ, p, x.

Proof: Using the definition (2.9) we have:

$$\left[\mathbf{A}^{k,p}_{g_\gamma}(\gamma L, \lambda, \gamma x)\right]^p = \frac{1}{\gamma L - \lambda} \int_{\gamma\hat{d}_{i_1}+\frac{\lambda}{2}}^{\gamma\hat{d}_{i_2}-\frac{\lambda}{2}} \left[A^k_{g_\gamma}(\lambda, \gamma x')\right]^p \mathrm{d}x' = \tag{A.2}$$

$$= \sum_{i=i_1}^{i_2-1} \frac{1}{\gamma L - \lambda} \int_{\gamma\hat{d}_i+\frac{\lambda}{2}}^{\gamma\hat{d}_{i+1}-\frac{\lambda}{2}} \left[A^k_{g_\gamma}(\lambda, \gamma x')\right]^p \mathrm{d}x' + \sum_{i=i_1+1}^{i_2-1} \frac{1}{\gamma L - \lambda} \int_{\gamma\hat{d}_i+\frac{\lambda}{2}}^{\gamma\hat{d}_i-\frac{\lambda}{2}} \left[A^k_{g_\gamma}(\lambda, \gamma x')\right]^p \mathrm{d}x' \tag{A.3}$$

The last equality is obtained by splitting the domain of integration into the ergodic components of g_γ (recall that $g_\gamma = \sum_{i=0}^{N-1} \chi_{[\gamma\hat{d}_i, \gamma\hat{d}_{i+1}]}\psi_i(\omega)$). The second summation includes that part of the domain of integration where the *supporting window* of $A^k_{g_\gamma}(\lambda, \gamma x')$, which is the interval $[\gamma x' - \frac{\lambda}{2}, \gamma x' + \frac{\lambda}{2}]$, contains a discontinuity point.

From (2.7) we have

$$A^k_{g_\gamma}(\lambda, \gamma x') = a_k(T_{-\lambda\gamma x'} D_{\lambda^{-1}} g_\gamma) \tag{A.4}$$

Since we have assumed $\sup_{g\in G} a_k(g) < \infty$, (A.4) is bounded as a function of x' and therefore the second summation in (A.3) is of order $O(\gamma^{-1})$. By using (A.4) and the expansion of g_γ we get:

$$\left[\mathbf{A}^{k,p}_{g_\gamma}(\gamma L, \lambda, \gamma x)\right]^p = \sum_{i=i_1}^{i_2-1} \frac{1}{\gamma L} \int_{\gamma\hat{d}_i+\frac{\lambda}{2}}^{\gamma\hat{d}_{i+1}-\frac{\lambda}{2}} [a_k(T_{-\lambda\gamma x'} D_{\lambda^{-1}} \psi_i(\omega))]^p \, \mathrm{d}x' + O(\gamma^{-1}) = \tag{A.5}$$

$$= \sum_{i=i_1}^{i_2-1} \frac{1}{\gamma L} \int_{\gamma\hat{d}_i}^{\gamma\hat{d}_{i+1}} [a_k(T_{-\lambda\gamma x'} D_{\lambda^{-1}} \psi_i(\omega))]^p \, \mathrm{d}x' + O(\gamma^{-1}) \tag{A.6}$$

Then, by taking the limit and using ergodicity we get:

$$\lim_{\gamma\to\infty} \left[\mathbf{A}^{k,p}_{g_\gamma}(\gamma L, \lambda, \gamma x)\right]^p = \sum_{i=i_1}^{i_2-1} \lim_{\gamma\to\infty} \frac{1}{\gamma L} \int_{\gamma\hat{d}_i}^{\gamma\hat{d}_{i+1}} [a_k(T_{-\lambda\gamma x'} D_{\lambda^{-1}} \psi_i(\omega))]^p \, \mathrm{d}x' = \tag{A.7}$$

$$= \sum_{i=i_1}^{i_2-1} \frac{\hat{d}_{i+1} - \hat{d}_i}{L} \mathrm{E}\left[a_k(D_{\lambda^{-1}} \psi_i(\omega))\right]^p = \sum_{i=i_1}^{i_2-1} \frac{\hat{d}_{i+1} - \hat{d}_i}{L} \left[s_{\psi_i}(k, \lambda, p)\right]^p \tag{A.8}$$

Uniformity of (A.1) is guaranteed by the fact the g is well behaved. Q.E.D.

Corollary 1 *Let the assumptions of theorem 1 hold. Then, if the restriction of ϕ to $[x-\frac{L}{2}, x+\frac{L}{2}]$ contains a single ergodic random function, i.e. if $[x-\frac{L}{2}, x+\frac{L}{2}] \subset [d_i, d_{i+1}]$ for some i, it follows:*

$$\lim_{\gamma\to\infty} \mathbf{A}^{k,p}_{g_\gamma}(\gamma L, \lambda, \gamma x) = \mathbf{A}^{k,p}_{\infty,g}(L, \lambda, x) = s_{\psi_i}(k, \lambda, p) \tag{A.9}$$

Proof: Since, for the i given above, $d_i \le x - \frac{L}{2} = x_1$, $d_{i+1} \ge x + \frac{L}{2} = x_2$, we have $i_1 = i$, $i_2 = i+1$, $\hat{d}_{i_1} = \hat{d}_i = x_1$, $\hat{d}_{i_2} = \hat{d}_i + 1 = x_2$, the summation in (A.1) reduces to a single term yielding the result. Q.E.D.

Corollary 2 *Let the conditions of theorem 1 hold. Then:*

$$\lim_{\gamma\to\infty} v_{g_\gamma}(\gamma L,\gamma x) = \max_{(k,\lambda,p)}\left[\max_{x'} \mathbf{A}^{k,p}_{\infty,g}(L,\lambda,x') - \min_{x'} \mathbf{A}^{k,p}_{\infty,g}(L,\lambda,x')\right] \equiv v^{\infty}_{g}(L,x) \tag{A.10}$$

where $\mathbf{A}^{k,p}_{\infty,g}(L,\lambda,x)$ *is as given in lemma 1 and the extrema inside the square brackets are performed in the usual way, i.e. for* $x-\frac{L}{2} \le x' \le x+\frac{L}{2}$ *(compare with (2.12) and (2.13)).*

Proof: From the definitions of section 2.4:

$$\lim_{\gamma\to\infty} v_{g_\gamma}(\gamma L,\gamma x) = \lim_{\gamma\to\infty}\max_{(k,\lambda,p)} v^{k,p}_{g_\gamma}(\gamma L,\lambda,\gamma x) = \tag{A.11}$$

$$= \lim_{\gamma\to\infty}\max_{(k,\lambda,p)}\left[\max_{x'} \mathbf{A}^{k,p}_{g_\gamma}(\gamma L,\lambda,\gamma x') - \min_{x'} \mathbf{A}^{k,p}_{g_\gamma}(\gamma L,\lambda,\gamma x')\right] = \tag{A.12}$$

$$= \max_{(k,\lambda,p)}\left[\max_{x'}\lim_{\gamma\to\infty} \mathbf{A}^{k,p}_{g_\gamma}(\gamma L,\lambda,\gamma x') - \min_{x'}\lim_{\gamma\to\infty} \mathbf{A}^{k,p}_{g_\gamma}(\gamma L,\lambda,\gamma x')\right] = \tag{A.13}$$

$$\max_{(k,\lambda,p)}\left[\max_{x'} \mathbf{A}^{k,p}_{\infty,g}(L,\lambda,x') - \min_{x'} \mathbf{A}^{k,p}_{\infty,g}(L,\lambda,x')\right] \tag{A.14}$$

The equality before the last follows from the uniformity of the limit in lemma 1. Q.E.D.

Corollary 3 *Let the conditions of corallary 3 hold. Then*

$$\lim_{\gamma\to\infty} v_{g_\gamma}(\gamma L,\gamma x) = v^{\infty}_{g}(L,x) = 0 \tag{A.15}$$

Proof: Since $[x-L, x+L] \subset [d_i, d_{i+1}]$, then $[x'-\frac{L}{2}, x'+\frac{L}{2}] \subset [d_i, d_{i+1}]$ for $x'-\frac{L}{2} \le x'+\frac{L}{2}$. Therefore, from corollary 2 and corollary 1 we have:

$$\lim_{\gamma\to\infty} v_{g_\gamma}(\gamma L,\gamma x) = \max_{(k,\lambda,p)}\left[\max_{x'} \mathbf{A}^{k,p}_{\infty,g}(L,\lambda,x') - \min_{x'} \mathbf{A}^{k,p}_{\infty,g}(L,\lambda,x')\right] = \tag{A.16}$$

$$= \max_{(k,\lambda,p)}\left[\max_{x'} s_{\psi_i}(k,\lambda,p) - \min_{x'} s_{\psi_i}(k,\lambda,p)\right] = 0 \qquad \text{Q.E.D.} \tag{A.17}$$

Lemma 2 *Let the assumptions of theorem 1 hold. Let* $R = [x_1, x_2] \subset \Sigma$ *where* Σ *is the support of gand let* $x = \frac{x_1+x_2}{2}$, $L = \frac{x_2-x_1}{2}$. *Assume that for all* $i_1 \neq i_2$ *there*

exist i, λ *and* p, $1 \leq p < \infty$ *such that* $s_{\psi_{i_1}}(k,\lambda,p) \neq s_{\psi_{i_2}}(k,\lambda,p)$. *Suppose* $Q = \mathrm{Card}(D \cap]x_1, x_2[) \geq 1$ *and define* $\Delta = \Delta(R,D) = \max_{d_i \in R \cap D} \min\left((d_i - x_1), (x_2 - d_i)\right)$. *Then*

$$v_g^{\infty}(R) \equiv v_g^{\infty}(L,x) > \tag{A.18}$$

$$> \min_{i_1 \neq i_2} \sup_{(k,\lambda,p)} \left\{ \left[\left| [s_{\psi_{i_1}}(k,\lambda,p)]^p - [s_{\psi_{i_2}}(k,\lambda,p)]^p \right| \frac{\Delta}{LQ} \right]^{\frac{1}{p}} 2^{1-\frac{1}{p}} \right\} > K_g \frac{\Delta}{LQ} \geq K_g \frac{\Delta}{LN} > 0 \tag{A.19}$$

where $N = \mathrm{Card}(D) - 1$ *and the costant* K_g *does not depend on* x_1, x_2.

Bibliography

[1] L. Ambrosio and V. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via γ-convergence. *Comm. on Pure and applied Math.*, ? to appear.

[2] R. Bajcsy and L. Liberman. Texture gradients as a depth cue. *Comput. Graphics Image Process.*, 5:52–67, Mar. 1976.

[3] D. Blostein and N. Ahuja. Shape from texture: Integrating texture-element extraction and surface estimation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(12):1233–1250, Dec. 1989.

[4] A.C. Bovik, M. Clark, and W.S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(1):55–73, 1990.

[5] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 8(6):679–698, 1986.

[6] P.C. Chen and T. Pavlidis. Segmentation by texture using a co-occurence matrix and a split-and-merge algorithm. *Comput. Graphics Image Process.*, 10:172–182, 1979.

[7] P.C. Chen and T. Pavlidis. Segmentation by texture using correlation. *IEEE Trans. Pattern Anal. Machine Intell.*, 5(1):64–69, Jan. 1983.

[8] M. Clark and A.C. Bovik. Texture segmentation using gabor modulation/demodulation. *Pattern Recogn. Lett.*, 6:261–267, 1987.

[9] G.R. Cross and A.K. Jain. Markov random fields texture models. *IEEE Trans. Pattern Anal. Machine Intell.*, 5(1):25–39, Jan. 1983.

[10] J.L. Crowley and A.C. Parker. A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(2):156–170, Mar. 1984.

[11] L.S. Davis, S.A. Johns, and J.K. Aggarwal. Texture analysis using generalized co-occurence matrices. *IEEE Trans. Pattern Anal. Machine Intell.*, 1:251–259, 1979.

[12] H. Derin and W.S. Cole. Segmentation of textured images using gibbs random fields. *Comput. Graphics Image Process.*, 35:72–98, 1986.

[13] H. Derrin and H. Elliott. Modeling and segmentation of noisy and textured images using gibbs random fields. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(1):39–55, 1987.

[14] C.R. Dyer. Multiscale image understanding. In L. Uhr, editor, *Parallel Computer Vision*, pages 171–213. Academic Press, Orlando, Flo., 1987.

[15] M. Galloway. Texture analysis using gray-level run length. *Comput. Graphics Image Process.*, 4:172–199, 1974.

[16] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constraint optimization. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(7):609, 1990.

[17] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(6):721–741, Nov. 1984.

[18] A. Hanson and E. Riseman. The visions image-understanding system. In C. Brown, editor, *Advances in Computer Vision*, volume 1, chapter 1, pages 1–114. Lawrence Erlbaum Associates, Hillsdale, N.J., 1988.

[19] R.M. Haralick. Statistical and structural approaches to texture. *Proc. IEEE*, 67(5):786–804, 1979.

[20] R.M. Haralick. Edge and region analysis for digital image data. *Comput. Graphics Image Process.*, 12:60–73, 1980.

[21] T.H. Hong and A. Rosenfeld. Compact region extraction using weighted pixel linking in a pyramid. *IEEE Trans. Pattern Anal. Machine Intell.*, 6:222–229, 1984.

[22] S.L. Horowitz and T. Pavlidis. Picture segmentation by a tree traversal algorithm. *J. ACM*, 23:368–388, Apr. 1976.

[23] J.Y. Hsiao and A.A. Sawchuk. Supervised textured image segmentation using feature smoothing and probabilistic relaxation techniques. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(12):1279–1292, Dec 1989.

[24] J.Y. Hsiao and A.A. Sawchuk. Unsupervised textured image segmentation using feature smoothing and probabilistic relaxation techniques. *Comput. Graphics Image Process.*, 48(1), 1989.

[25] D. Jeong and P.M. Lapsa. Unified approach for early-phase image understanding using a general decision criterion. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(4):357–371, Apr. 1989.

[26] B. Julesz. Textons, the elements of texture perception. *Nature*, 290:91–97, 1981.

[27] B. Julesz and J.R. Bergen. Textons, the fundamental elements in preattentive vision and perception of textures. *Bell Syst. Tech. J.*, 62:1619–1645, 1983.

[28] R.L. Kashyap and A. Khotanzad. A model-based method for rotation invariant texture classification. *IEEE Trans. Pattern Anal. Machine Intell.*, 8(4):472–481, 1986.

[29] A. Khotanzad and J. Chen. Unsupervised segmentation of texture images by edge detection in multidimensional features. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(4), Apr. 1989.

[30] B.P. Kjell and C.R. Dyer. Segmentation of textured images using pyramid linking. In V. Cantoni and S. Levialdi, editors, *Pyramidal Systems for Image Processing and Computer Vision*. Springer-Verlag, Berlin, 1987.

[31] M. Kocher and R. Leonardi. Adaptive region growing technique using polynomial functions for image approximation. *Signal Processing*, 11:47–60, 1986.

[32] K.I. Laws. Rapid texture identification. *Proc. SPIE*, 238:376–380, 1980.

[33] Jitendra Malik and Pietro Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America – A*, 7(5):923–932, 1990.

[34] D. Marr. *Vision*. Freeman, New York, 1982.

[35] D. Marr and E. Hildreth. Theory of edge detection. *Proc. Roy. Soc. Lon. Ser. B*, 207:187–217, 1980.

[36] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *J. Am. Stat. Ass.*, 82(397):76–89, Mar. 1987.

[37] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *Proc. IEEE CVPR*, pages 22–26, San Francisco, CA, 1985.

[38] R. Ohlander, K. Price, and R. Reddy. Picture segmentation using a recursive region splitting method. *Comput. Graphics Image Process.*, 8:313–333, 1978.

[39] T. Pavlidis and Y.L. Liow. Integrating region growing and edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(3):225–233, Mar. 1990.

[40] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[41] Pietro Perona and Jitendra Malik. Detecting and localizing edges composed of steps, peaks and roofs. In *Third International Conference on Computer Vision*, pages 52–57. IEEE Computer Society, Osaka, 1990.

[42] M. Pietikainen and A. Rosenfeld. Image segmentation by texture using pyramid node linking. *IEEE Trans. Syst., Man, Cybern.*, 11:822–825, 1981.

[43] M. Pietikäinen, A. Rosenfeld, and L.S. Davis. Experiments with texture classification using averages o local pattern matches. *IEEE Trans. Syst., Man, Cybern.*, 13:421–426, 1983.

[44] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–139, Sep. 1985.

[45] T.R. Reed and H. Wechsler. Tracking for non-stationarities for textured fields. *Signal Processing*, 14(1):95–102, Jan. 1988.

[46] T.R. Reed and H. Wechsler. Segmentation of textured images and gestalt organization using spatial/spatial-frequency representations. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(1):1–12, Jan. 1990.

[47] T.J. Richardson. Scale independent piecewise smooth segmentation of images via variational methods. Technical Report LIDS-TH-1940, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, February 1990.

[48] De Souza. Texture recognition via autoregression. *Pattern Recogn.*, 15:471–475, 1982.

[49] M. Spann and R. Wilson. A quad-tree approach to image segmetation which combines statistical and spatial information. *Pattern Recogn.*, 18:257–269, 1985.

[50] C.W. Therrien. An estimation-theoretic approach to terrain image segmentation. *Comput. Graphics Image Process.*, 22:313–326, 1983.

[51] M. Tüceryan and A.K. Jain. Texture segmentation using voronoi polygons. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(2):211–216, 1990.

[52] M. Unser. Local linear transforms for texture measurments. *Signal Processing*, 11:61–79, 1986.

[53] F. Vilnrotter, R. Nevatia, and K. Price. Structural analysis of natural texture. *IEEE Trans. Pattern Anal. Machine Intell.*, 8(1):76–89, 1986.

[54] H. Voorhees and T. Poggio. Computing texture boundaries from images. *Nature*, 333:364–367, May 1988.

[55] J.S. Weszka, C.R. Dyer, and A. Rosenfeld. A comparison study of texture measures for terrain classification. *IEEE Trans. Syst., Man, Cybern.*, 6(4):269–286, 1976.

[56] R. Wilson and G.H. Granlund. The uncertainty principle in image processing. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(6):758–767, Nov. 1984.

[57] R. Wilson and M. Spann. Finite prolate spheroidal sequences and their application i: generation and properties. *IEEE Trans. Pattern Anal. Machine Intell.*, pages 787–795, Nov. 1987.

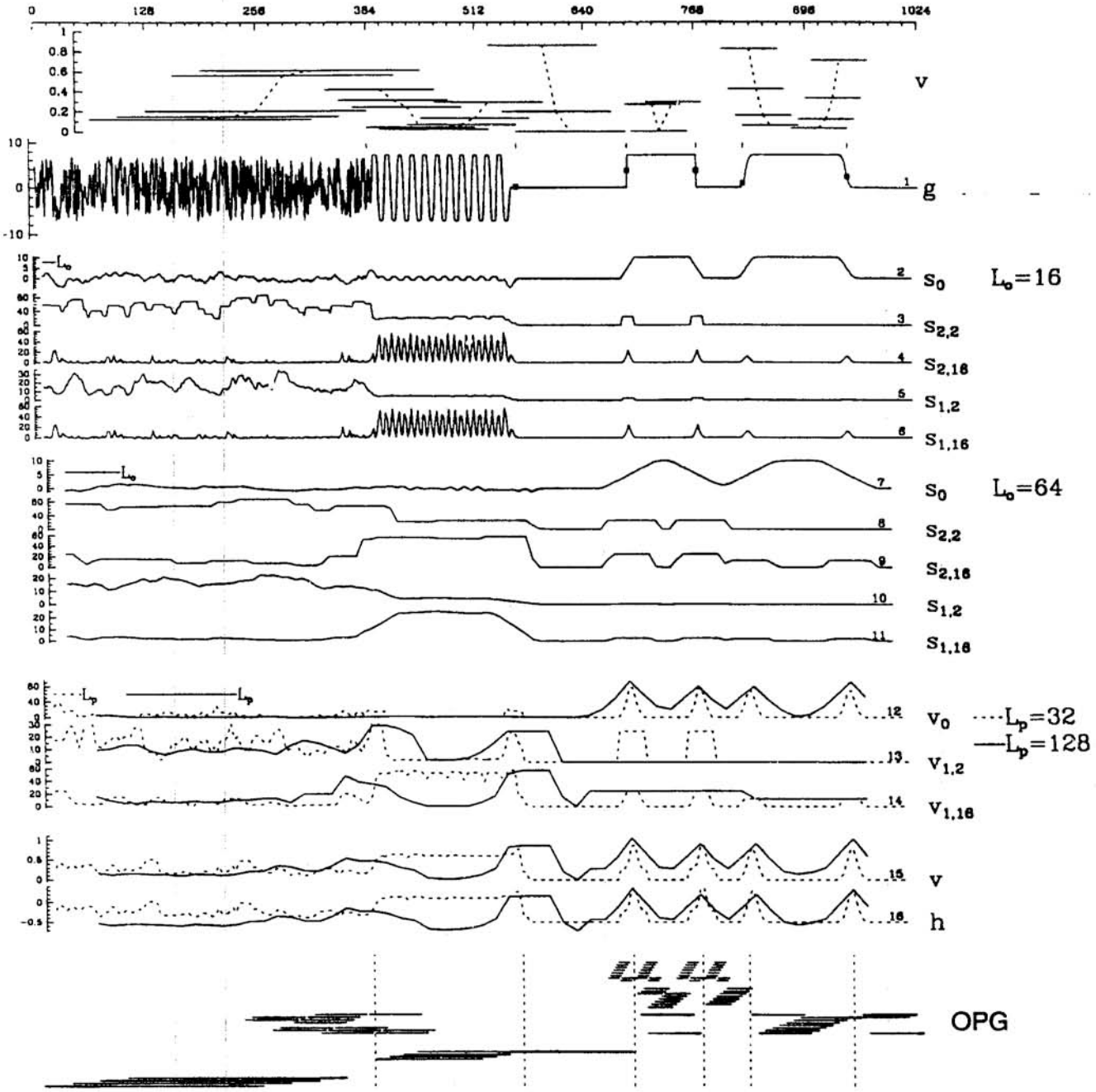
[58] R. Wilson and M. Spann. Finite prolate spheroidal sequences and their application ii: Image feature description and segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 10:193–203, Mar. 1988.

[59] A.P. Witkin. Scale-space filtering. In *International Joint Conference on Artificial Intelligence*, pages 1019–1021, 1983. Karlsruhe.

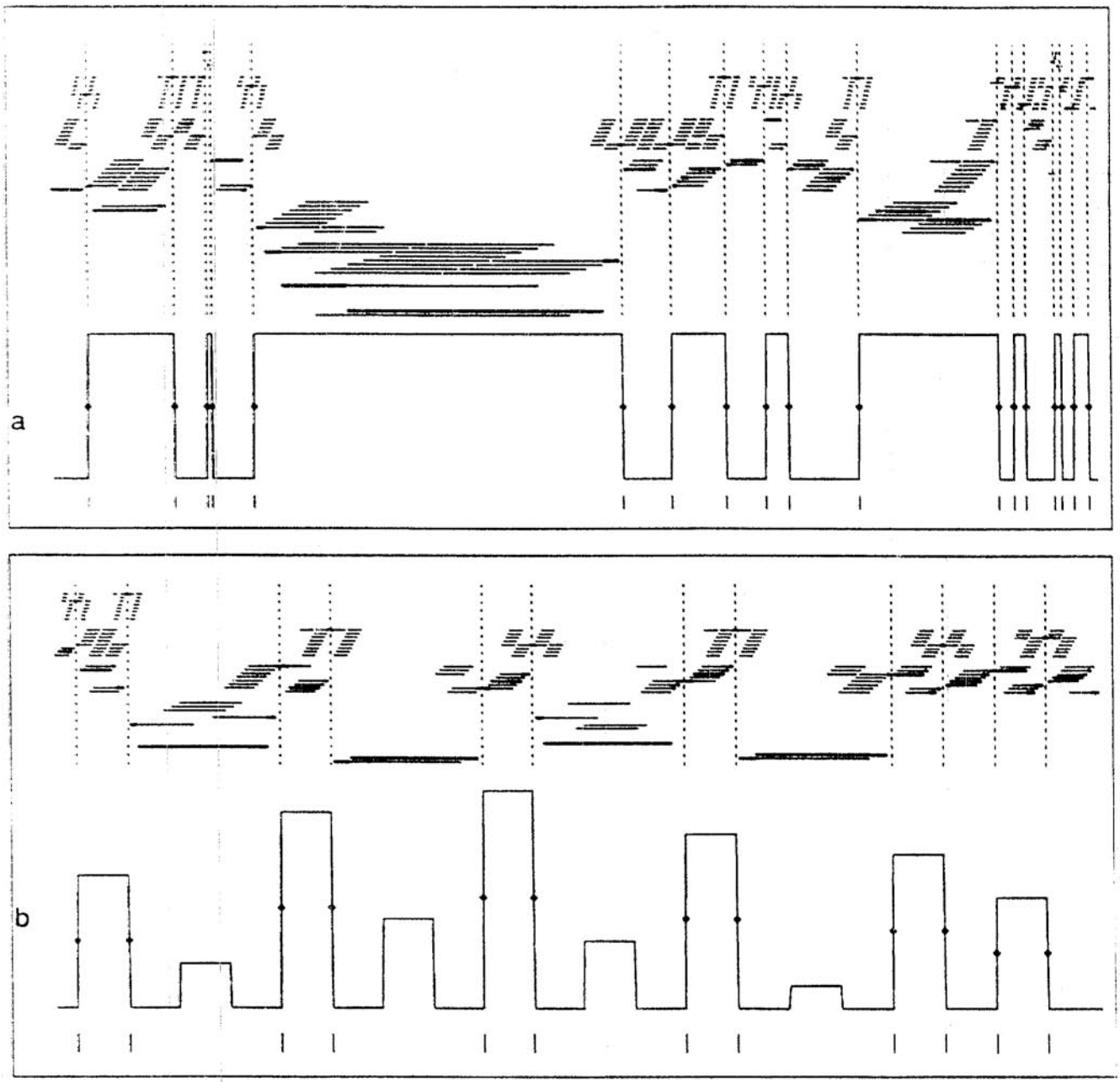
[60] A. Yuille and T. Poggio. Scaling theorems for zero crossings. *IEEE Transactions on Pattern Analysis, and Machine Intelligence*, 8, January 1986.

[61] Sifen Zhong and Stephane Mallat. Compact image representation from multiscale edges. In *Third International Conference on Computer Vision*. IEEE Computer Society, Osaka, 1990.

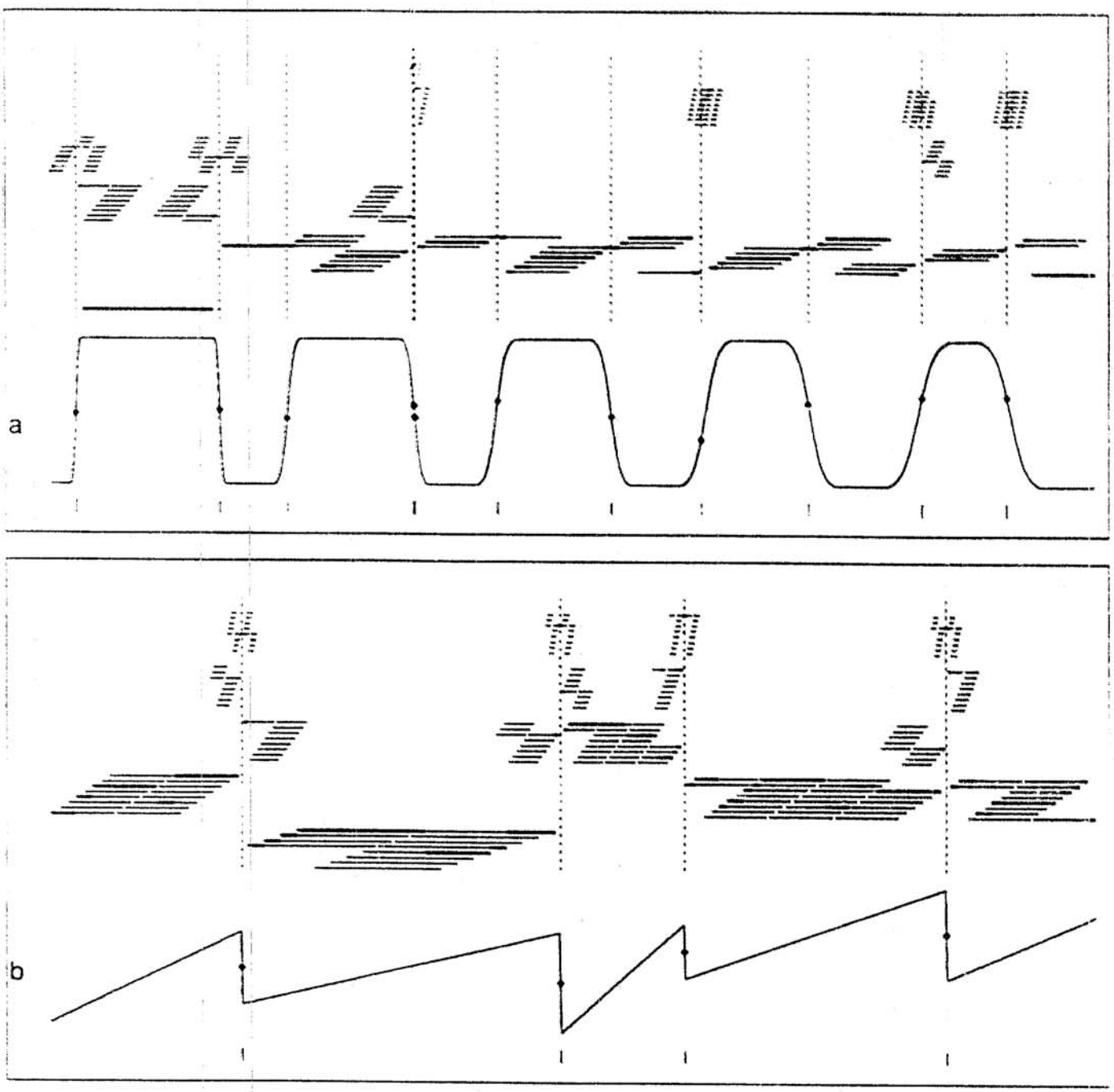
[62] S.W. Zucker. Region growing: childhood and adolescence. *Comput. Graphics Image Process.*, 5:382–399, 1976.


0
128
256
384
512
640
768
896
1024
v
g
L_o
s_0
L_o=16
$s_{2,2}$
$s_{2,16}$
$s_{1,2}$
$s_{1,16}$
L_o=64
L_p
L_p
v_0
L_p=32
L_p=128
$v_{1,2}$
$v_{1,16}$
v
h
OPG

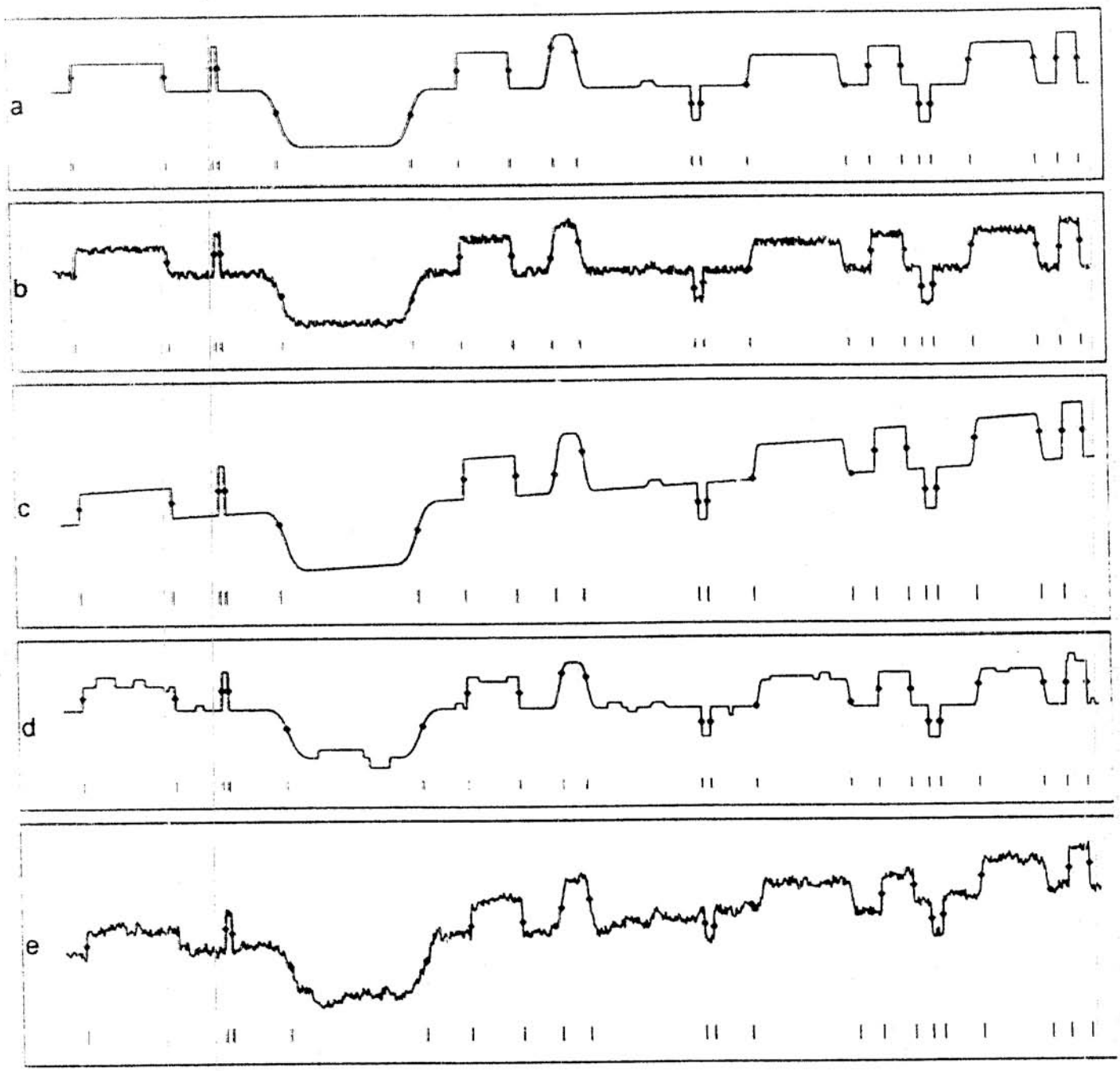
T0



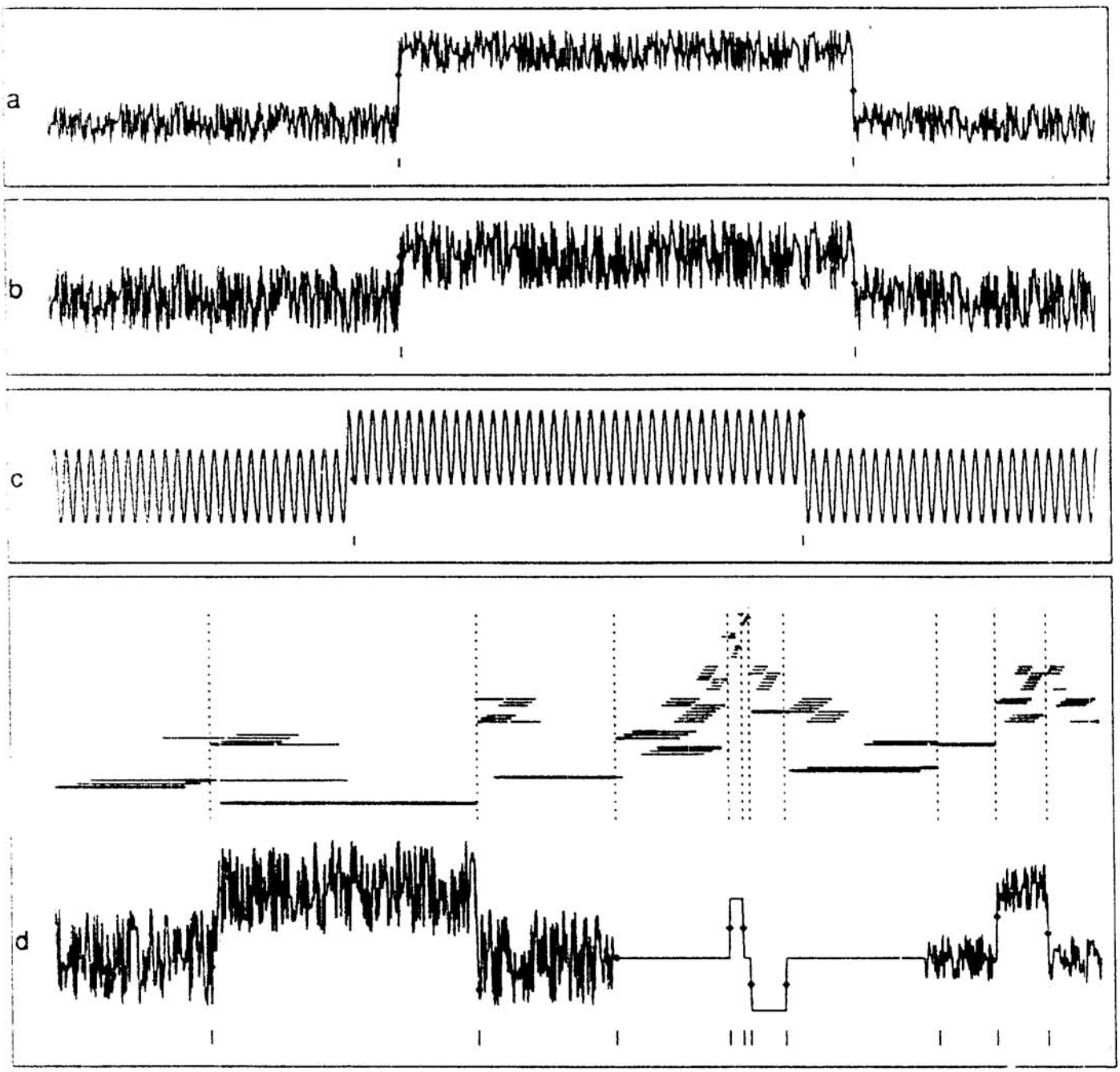
T1



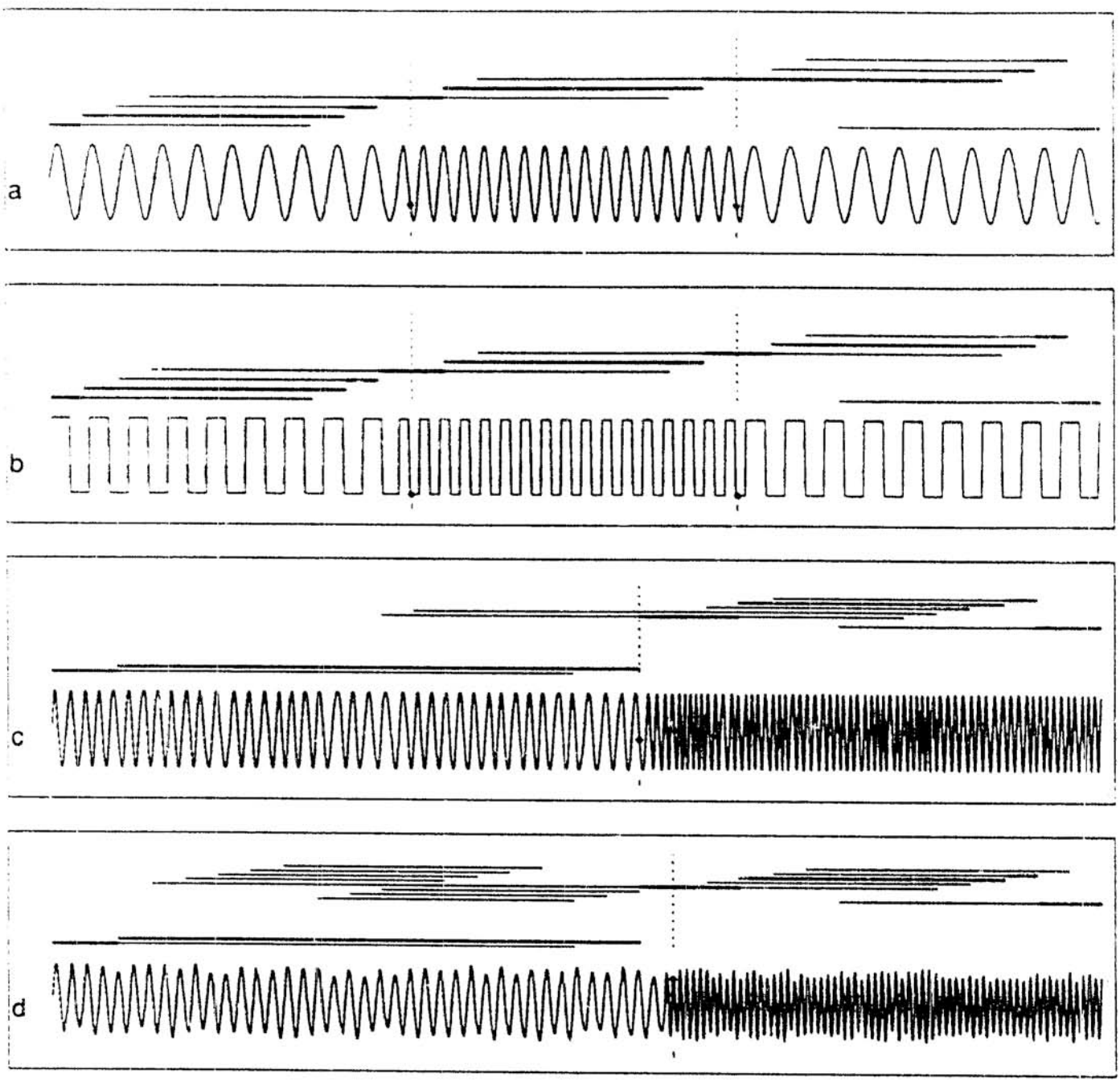
T2



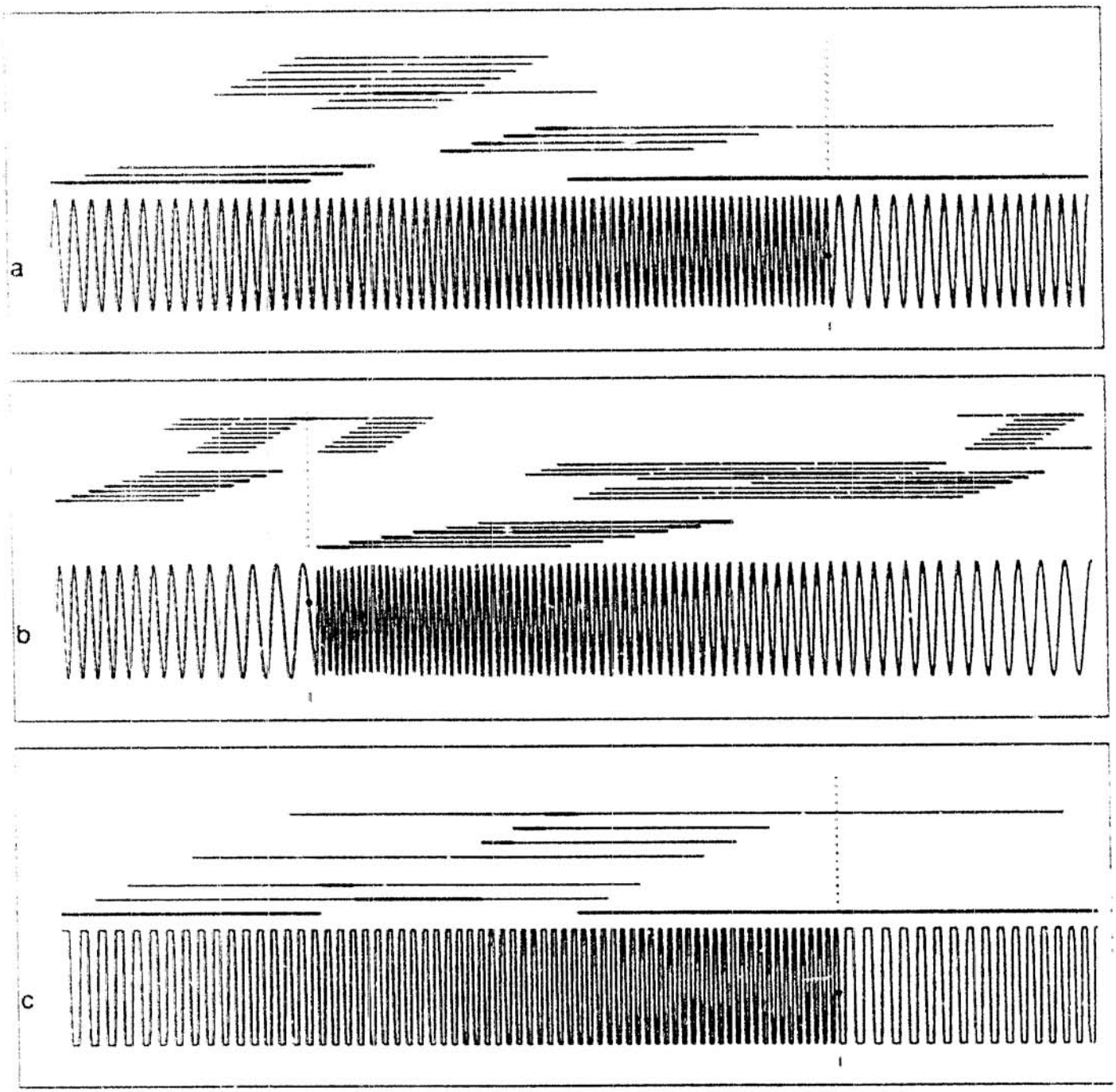
T3



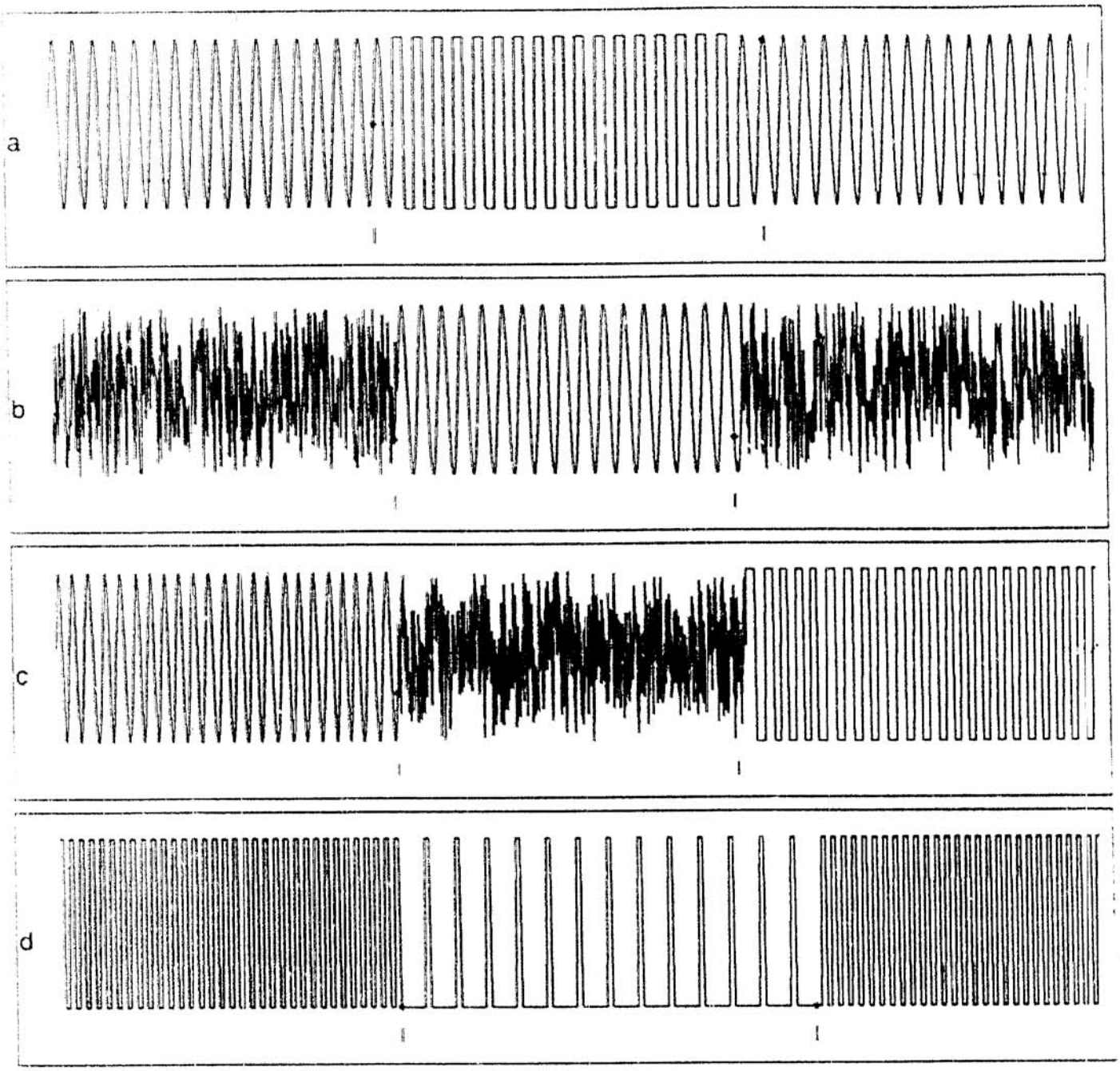
T4

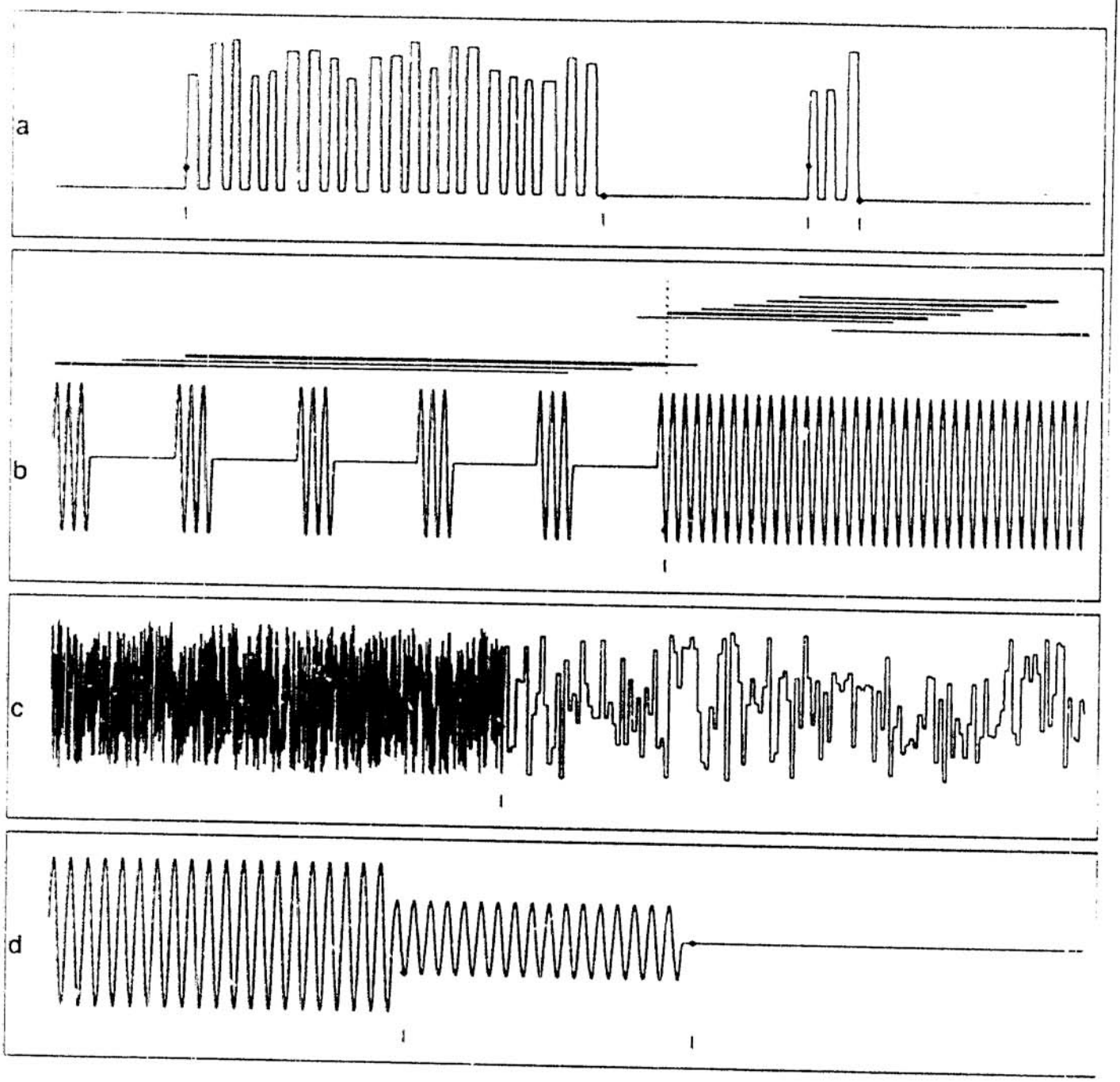


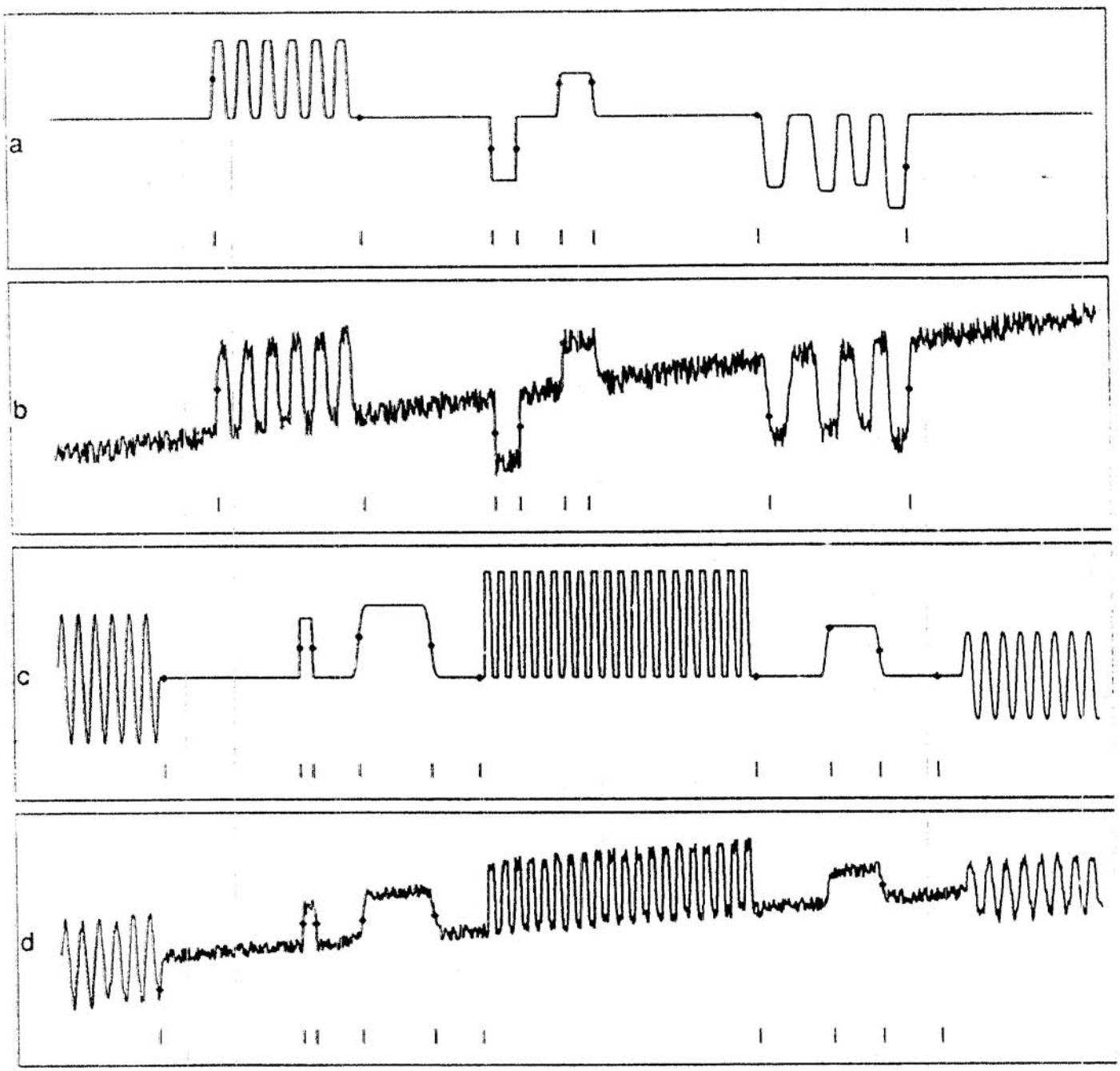
T5

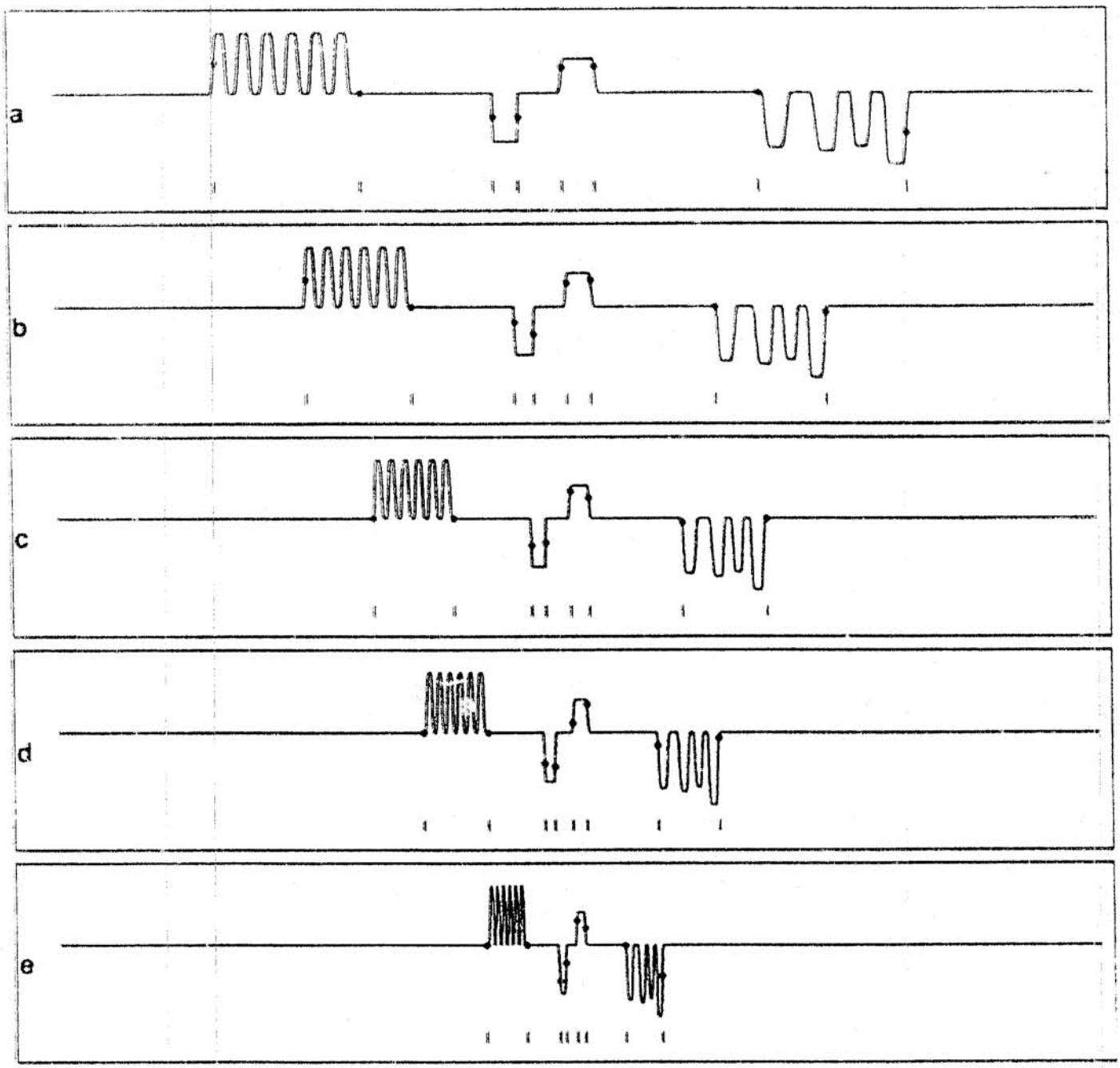


T6


a
b
c
d


a
b
c
d


a
b
c
d


a
b
c
d
e