Hierarchy in Picture Segmentation: A Stepwise Optimization Approach

JEAN-MARIE BEAULIEU, MEMBER, IEEE, AND MORRIS GOLDBERG, MEMBER, IEEE

最早的区域合并问题->其数学解释是segments的影像拟合
最早介绍区域合并的代价，每次的区域合并都会带来整个拟合误差的增加，这个增加量称之为合并代价。图的每条边都有权重，这个权重就表示这两个区域合并对整个图像拟合误差带来的增加量。

***Abstract*—This paper presents a segmentation algorithm based upon sequential optimization which produces a hierarchical decomposition of the picture. The decomposition is data driven with no restriction on segment shapes. It can be viewed as a tree where the nodes correspond to picture segments and where links between nodes indicate set inclusions. Picture segmentation is first regarded as a problem of piecewise picture approximation, which consists in finding the partition with the minimum approximation error. Then, picture segmentation is presented as a hypothesis testing process which merges only segments that belong to the same region. A hierarchical decomposition constraint is used in both cases, which results into the same stepwise optimization algorithm. At each iteration, the two most similar segments are merged by optimizing a "stepwise criterion." The algorithm is employed to segment a remote sensing picture, and illustrate the hierarchical structure of the picture.**

***Index Terms*—Clustering, hierarchical data structure, hypothesis testing, picture approximation, picture segmentation, sequential optimization.**

I. INTRODUCTION

PICTURE segmentation often constitutes the low-level processing stage of a picture analysis system [3]. An image is thus segmented into regions that roughly correspond to objects, surfaces, or parts of objects of the scene. The high-level stage is then devoted to the interpretation of these regions. However, other low-level processes can also be used such as the edge detection. This assumes that the low-level process has no *a priori* knowledge about the objects in a scene, and would be able to deliver a "plausible" output to the high-level process. It should be noted that the distinction between the two levels of analysis is primarily in terms of the knowledge available to each. Whereas the interpretation system uses domain specific knowledge about the contents of the scene, the low-level segmentation stage employs general purpose models that contain knowledge about images and grouping criteria that are independent of the scene under analysis.

Picture segmentation goals can be easily defined in high-level (interpretation) terms, such as the segmentation of the image into regions that correspond to objects in the scene, or the isolation of a particular object from the background [1], [8]. Goal definitions strictly applicable to the low-level have been rather imprecise. For example, a segmentation can be defined as a partition of the image into regions that are "uniform" and that bear "contrast" to their adjacent neighbors. Uniformity and contrast are measured in terms of a set of low-level features that can be evaluated over the image; for example, the average gray level intensities of the regions [19], [12], [10].

参考书写

区域的平均灰度强度

In order to give picture segmentation a more sound basis, the segmentation goal should be precisely defined; picture segmentation must be presented and analyzed as a mathematical problem. Using computational and mathematical theories, algorithms that solve this problem can then be analyzed and evaluated. An analogous situation is found in pattern recognition, where the clustering approach has been presented as a mathematical problem [11], [34]. Let $\{v_i \mid i = 1 \cdots n\}$ be a set of pattern samples, then pattern clustering can be regarded as finding the partition of the set into m subsets, V_k, $k = 1 \cdots m$, that minimizes the overall intracluster variance:

intra->在这篇文章中就已经提出方差作为块内部均一性的一个度量标准

$$\mathrm{VAR}_{\mathrm{intra}} = \sum_{k=1\cdots m} \sum_{v_i \in V_k} [v_i - \mu_k]^2$$

最大程度减小簇内整体方差，簇即可理解为分割中的segments。更小的方差表示块内更均一。

where μ_k is the mean value of the subset (or cluster) V_k. Other measures can also be used to define well formed clusters. However, the intracluster variance is simple and yet captures the basic aspect of clusters, i.e., compact sets of points with large distances between clusters. A similar approach may be useful in picture segmentation; however, the resulting mathematical problems must take account of the spatial aspects that characterize the picture segmentation task.

Mathematical analysis techniques could, then, be applied to picture segmentation. Hence, it is noted that low-level vision problems are often ill-posed [27], [32]. To make them well posed, generic constraints on the problem must be introduced. These attempt to force the solution to lie in a subspace of the solution space, where it is well posed [33], [27]. Useful constraints have been proposed for edge detection [27], [21], for surface reconstruction [32], [6], and for image restoration [13]. The way to constrain a segmentation algorithm in order to obtain a unique, robust, and reliable solution, is not evident. At first glance, it seems that such constraints can only be introduced at the high-level stage of picture analysis. In this case, a feedback path from the high-level stage to the

Manuscript received June 19, 1986; revised July 20, 1987. Recommended for acceptance by R. Bajcsy. This work was supported by the Canadian Center for Remote Sensing and the Natural Science and Engineering Research Council of Canada.

J.-M. Beaulieu is with the Department of Computer Science, Laval University, Quebec, P.Q. G1K 7P4, Canada.

M. Goldberg is with the Department of Electrical Engineering, Ottawa University, Ottawa, Ont. K1N 6N5, Canada.

IEEE Log Number 8825637.

picture segmentation stage must be introduced. The high-level stage therefore guides the segmentation stage through the set of possible partitions in order to find the unique partition that meets the constraints [22], [24], [18].

One important generic constraint, that can be employed, is the hierarchical structure of the picture [31], [30]. A hierarchical structure means that the picture can be divided into components, corresponding for example to scene objects, which can then be divided into subcomponents corresponding to object subparts. This hierarchical decomposition can be represented by a tree where the nodes correspond to picture regions and where links between nodes indicate set inclusions. The hierarchical levels could be related to the resolution levels: a region, which is higher in the hierarchy than its subparts, is also larger than its subparts. Hence, higher level regions could be discerned at a coarser resolution than their subparts. However, a region hierarchy could be more than just a multiresolution representation of the picture (also named regular hierarchy or pyramid [30]). In a data driven decomposition, a criterion is used to order the regions, i.e., to decide in which order the regions (parts) are merged to produce higher level regions (objects) (see conceptual hierarchies in [2]). Thus, the hierarchical structure is only one aspect of a picture model, the other aspects involve the ordering criterion of the regions. Hence, for example, a textured image could require a complex picture model resulting in a complex criterion for the region ordering.

This paper shows how a hierarchical structure can be effectively exploited to constrain picture segmentation problems. Two picture segmentation problems are considered. First, following a structural approach, the picture segmentation is regarded as a problem of piecewise picture approximation which consists in finding the partition with the minimum approximation error. Then, using a statistical approach, the picture segmentation is presented as a hypothesis testing process which merges only segments that belong to the same region. In each case, the segmentation problem is precisely defined, and an algorithm, that solves this problem while satisfying the hierarchical constraint, is derived.

The analysis of both problems results in a segmentation algorithm based upon sequential optimization and which produces a hierarchical decomposition of the picture. The algorithm starts with an initial picture partition, and at each iteration, merges two segments. An optimization process is used to select the segment pair that minimizes a "stepwise criterion," $C_{i,j}$, corresponding to the cost of merging the segment S_i with the segment S_j. It is shown that the algorithm is a valuable tool, and produces good segmentation results.

This algorithm can be adapted to different picture segmentation applications. However, appropriate picture models should be used in order to obtain good results. Selection of an appropriate model for a given application is a difficult problem, and is not discussed here. This paper shows how the stepwise criterion could be derived from the used picture model. Two classes of picture models are examined: the global optimization and picture approximation approach and the hypothesis testing approach. Simple illustrative examples are given, and indications for the utilization of more complex picture models are provided. Moreover, the relation between the picture approximation and the hypothesis testing approaches is illustrated by using a similar picture model in both cases, which results in a similar stepwise criterion. Following a statistical approach, picture partition can be presented as a best estimate problem, which can be rewritten as a picture approximation problem as shown in the Appendix.

In the next section, the Hierarchical Stepwise Optimization algorithm (HSWO) is described in detail. The third section considers the problem of piecewise picture approximation. The stepwise optimization (HSWO) algorithm is then derived from the global optimization problem by the introduction of a hierarchical structure constraint. In the fourth section, picture segmentation is regarded as a hypothesis testing process which merges two segments only if they belong to the same region. It is shown how the probability of error can be minimized in a stepwise fashion. The last section illustrates the operation of the HSWO algorithm upon remote sensing pictures, and shows how the hierarchical structures of the pictures are extracted.

II. A Hierarchical Picture Segmentation Algorithm

A segment hierarchy can be represented by a tree [25], [36] (see Fig. 1). In a tree, segments at lower levels are joined to form segments at higher levels. The ith node at the τth level of the tree corresponds to the segment S_i^τ. The links between nodes indicate set inclusion. Hence, a link between a segment $S_k^{\tau+1}$ (ancestor or parent) and its disjoint subparts S_i^τ (descendants or sons) indicates that $S_i^\tau \subset S_k^{\tau+1}$. The root of the tree corresponds to I, the whole picture, and the leaves to pixels. A picture partition, P, therefore corresponds to a node set $\{S_1, S_2 \cdots S_n\}$, called a node cutset, which is the minimal set of nodes separating the root from all the leaves [16].

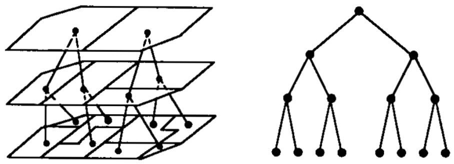
This section first describes a hierarchical segmentation algorithm based upon stepwise optimization, and discusses its efficient implementation. Then, the proposed algorithm is compared with previous hierarchical segmentation algorithms based upon predicate equations.

A. The Hierarchical Stepwise Optimization Algorithm

A hierarchical segmentation algorithm inspired from hierarchical data clustering and based upon stepwise optimization is now presented [35], [29], [28]. In a merging scheme, a hierarchical clustering starts with N clusters corresponding to each of the N data points, and sequentially reduces the number of clusters by merging. At each iteration, the similarity measures $d(C_i, C_j)$, are calculated for all clusters pairs (C_i, C_j), and the clusters of the pair that minimizes the measure are merged. This merging is repeated sequentially until the required number of clusters is obtained.

参考

上下方向，既可以向上合并，也可以继续向下分割。



(a) (b)

Fig. 1. Segment hierarchy and segment tree.

An important limitation of the hierarchical clustering approach is its excessive computing time for large data set. If there are N clusters, then the similarity measure for $N \times (N - 1)$ possible cluster pairs must be calculated. In picture segmentation, however, only adjacent segments can be merged, reducing the number of potential segment pairs per iteration to $N \times M$, where N is the number of segments, and M the average number of neighbors per segment. M is usually small ($4 \leq M \leq 8$) and is quite independent of N. Furthermore, a segment merge affects only the surrounding segments, and only the pairs involving those segments need to be modified or updated. Thus, only a limited number of new segment pairs must be considered at each iteration. Note that this gain of computing efficiency is only obtained for agglomerative and not divisive hierarchical segmentation.

The Hierarchical Stepwise Optimization algorithm (HSWO) which employs a sequence of optimization processes to produce a hierarchical segmentation is now presented. It starts with an initial picture partition, $P^0 = \{S_1, S_2, \cdots S_n\}$, and at each iteration, merges two segments to yield a segment hierarchy. An optimization process is used to select the segment pair that minimizes a "stepwise criterion" $C_{i,j}$ corresponding to the cost of merging S_i with S_j. The variables involved in the algorithm are as follows:

1) B_i, the set of the segments adjacent to S_i, called the neighborhood,

2) D_i, the parameters that describe the segment S_i, e.g., the segment mean and size, and

3) $C_{i,j} = C(D_i, D_j)$, the cost of merging segment S_i with S_j, where S_j is contained in B_i.

The choice for the stepwise criterion and the stopping condition depends upon the particular application; examples are given in the following sections.

The algorithm consists of the following steps:

I—Initialization:

i) $P^0 = \{S_1, S_2, \cdots S_n\}$ (initial partition).
ii) $k = 0$ and $m = n$.
iii) calculate D_i and B_i for $\forall\, S_i \in P^0$.
iv) calculate $CS = \{C_{i,j} \mid S_j \in B_i \text{ and } i > j\}$.

II—Merge the two most similar segments:

i) $k = k + 1$ and $m = m + 1$.
ii) find $C_{u,v} = \underset{C_{i,j} \in CS}{\text{Minimum}}\,(C_{i,j})$.
iii) $P^k = (P^{k-1} \cup \{S_m\}) \cap \overline{\{S_u, S_v\}}$ [1].
iv) calculate D_m from D_u and D_v.
v) $B_m = (B_u \cup B_v) \cap \overline{\{S_u, S_v\}}$ [1].
vi) $\forall\, S_j \in B_m,\ B_j = (B_j \cup \{S_m\}) \cap \overline{\{S_u, S_v\}}$ [1].
vii) $CS = (CS \cup \{C_{m,j} \mid S_j \in B_m\}) \cap \overline{\{C_{i,j} \mid i, j = u \text{ or } v\}}$ [1].

III—Stopping condition:

Stop if no more mergers are required.
Otherwise, go to step II.

An initial picture partition, $P^0 = \{S_1, S_2, \cdots S_n\}$, should be first defined with strictly homogeneous segments; for example, each initial segment could contain only one pixel. The initialization step then calculates the stepwise criteria for each pair of adjacent segments, S_i and S_j. The criterion corresponds to the cost of merging the two segments. For example, the increase of the sum of the squared errors around the segment means could be used.

At each iteration, the segment pair, $\{S_u, S_v\}$, which minimizes the stepwise criterion is found and merged to produce a new segment S_m. The criterion values and neighbor sets are then updated. Steps II and III are repeated until the stopping condition is satisfied.

The algorithm is designed so as to reduce the computing time. In the initialization step, the computing time is a function of the picture size, the number of initial segments, and the number of neighbors per segment. On the other hand, the iterative steps are short; the computing time is mainly a function of the number of neighbors of S_m. The number of iterations depends upon the number of initial and final segments, each iteration reducing by one the number of segments. However, the algorithm requires substantial temporary memory space to store the current descriptive parameters, neighbor sets, and criterion values.

B. Stepwise Optimization Versus Logical Predicates

This section compares the usual approach for hierarchical segmentation, based upon logical predicate equations (LPE), which the HSWO algorithm. In LPE-based algorithms, logical predicate equations are used to define the desired picture partition $P = \{S_1, S_2 \cdots S_n\}$ [36]:

$$\text{Prd}_1: Q(S_i) = \text{true} \qquad \text{for all } i$$

$$\text{Prd}_2: Q(S_i \cup S_j) = \text{false}$$
$$\text{for all } i \neq j \text{ and } S_i \text{ adjacent to } S_j \qquad (2.2\text{-}1)$$

where S_i represents a segment or a region. The logical predicate $Q(\cdot)$ is used to express the requirements that all segments S_i of a partition P must satisfy.

The predicate equations Prd_1 and Prd_2 can therefore be regarded as the definition of a node cutset. A merging scheme starts with small segments S_i (or pixels) which satisfy Prd_1, and proceeds to satisfy Prd_2 by region merg-

[1]Taking the intersection with the complement corresponds to removing those elements from the set.

ing. It starts from the leaves of the tree, and climbs up the tree until it meets nodes $S_k^{\tau+1}(=S_i^\tau \cup S_j^\tau)$ for which the predicate values are false, $Q(S_i^\tau \cup S_j^\tau)$ = false. Thus, S_i^τ and S_j^τ are in the node cutset, and Prd_2 is used as the stopping criterion.

参加：块之间的相似性度量加入块与块的边缘考虑，还真有人做。

The logical predicate, $Q(S_i^\tau \cup S_j^\tau)$, can be considered as an evaluation of the similarity of S_i^τ and S_j^τ, meaning that segment merging stops when there are no more similar segments. Thus, Brice and Fennema [7] use two heuristics, based upon information from the segment boundaries, to evaluate the similarity of two segments. The phagocyte heuristic guides the merging of regions in such a way as to smooth or shorten the resulting boundary. The weakness heuristic merges two regions if a prescribed portion of their common boundary is weak.

参考：基于金字塔的合并

Horowitz and Pavlidis [16] propose a split-and-merge segmentation approach based upon the pyramidal data structure. The logical predicate, $Q(S_i^\tau)$, is regarded as an evaluation of segment homogeneity: $Q(S_i^\tau)$ is true if the segment approximation error is smaller than a threshold value. The process begins at an intermediate level τ of the tree $\{S_i^\tau\}$. This node cutset will be moved upward by segment merging or downward by splitting, until the segments of the node cutset satisfy Prd_1 and Prd_2.

LPE->HSWO

The HSWO algorithm operates in a more global and gradual manner than LPE-based algorithms. The global aspect of the HSWO algorithm results from the examination of all segment pairs, (S_i, S_j), in order to find the minimum $C_{i,j}$; while the LPE algorithms consider only two segments at a time. The stepwise optimization rule implies that the HSWO algorithm considers the whole picture context before merging two segments.

The stepwise optimization rule also implies that the most similar segments are merged first. The HSWO algorithm gradually merges the segments, starting with the ones having the smallest $C_{i,j}$ values. This gradual aspect is not possible in the LPE algorithms where only two states are considered: the true state for similar segments and the false state when segments are not similar.

The gradual aspect also means that the HSWO algorithm can be used to produce not just one partition, but a sequence of partitions from one picture. Moreover, it will be seen in Section V that the sequence of partitions reflect the hierarchical structure of the picture: in the initial partitions small details and objects in the picture are preserved, while only the most important components remain in the latter partitions. This sequence of partitions carries useful information that a high level process (interpreter) may exploit in order to select the most appropriate partition from the sequence.

In the next section, the piecewise picture approximation approach to segmentation [25] is considered, and the HSWO algorithm is derived from this optimization problem by imposing a hierarchical structure constraint. In Section IV, the picture segmentation is regarded as a hypothesis testing process, and it is shown that the stepwise optimization rule reduces the probability of error in hierarchical segmentation.

III. Optimization and Segment Hierarchy

A piecewise polynomial approximation is often used to represent a picture. The approximation error can then be employed as a gobal criterion $G(P)$, and an optimization process can be used to find the partition that minimizes this criterion. It is shown that the hierarchical stepwise optimization (HSWO) algorithm constitutes an interesting suboptimal approach to the global optimization problem. The segment hierarchy assumption reduces the search space, while the stepwise optimization assures that each iteration optimizes the global criterion. A detailed description of the algorithm for the constant piecewise approximation case is given, and its operation is illustrated by a simple example.

A. Piecewise Picture Approximation

A picture can be regarded as a two dimensional function $f(x, y)$, where $(x, y) \in I$, I being the picture plane. A picture partition P divides the picture plane I into n regions, $S_1, S_2, \cdots S_n$. Let $f_i(x, y)$ designate the pixel values for the region S_i, $f_i(x, y) = f(x, y)$ for $(x, y) \in S_i$. Then, each region S_i can be approximated by a polynomial function, $r_i(x, y)$, [15], [16], [25],

$$f_i(x, y) \approx r_i(x, y) = \sum_{(p,q)\in T} a_{p,q}^i (x)^p (y)^q \quad (3.1\text{-}1)$$

where T is the set of (p, q) pairs employed to define the terms of the polynomial function, $a_{p,q}(x)^p(y)^q$. The approximation error for each segment can then be calculated by the sum of the squared deviations:

$$H(S_i) = \sum_{(x,y)\in S_i} \left(f(x, y) - r_i(x, y)\right)^2. \quad (3.1\text{-}2)$$

Once the segment S_i is given, the coefficients $a_{p,q}^i$ that minimize $H(S_i)$ can be calculated and must yield the best polynomial approximation for S_i. The minimization of $H(S_i)$ implies:

$$\frac{\partial H}{\partial a_{p,q}} = 0, \quad \text{for } (p, q) \in T. \quad (3.1\text{-}3)$$

This can be rewritten as follows:

$$\sum_{p',q'} a_{p',q'} \sum_{(x,y)} (x)^{p+p'}(y)^{q+q'} = \sum_{(x,y)} f(x, y)(x)^p(y)^q \quad \text{for } (p, q) \in T. \quad (3.1\text{-}4)$$

This is a linear system with m equations and m unknowns, m being the number of allowed pairs (p, q); e.g., if $T = \{(0, 0), (0, 1), \text{and } (1, 0)\}$ then $m = 3$. The polynomial coefficients $a_{p,q}^i$ that minimize $H(S_i)$ can be obtained by solving this linear system. However, a unique solution may not result; for example, when the number of pixels in S_i is smaller than m, the number of coefficients.

Having defined the *segment* approximation problem, the problem of *picture* approximation is now considered. Once a picture is divided into segments $S_1, S_2, \cdots S_n$, each can be approximated, and a picture approximation, $r(x, y)$, results from the concatenation of the piecewise approximations $r_i(x, y)$:

$$r(x, y) = \begin{cases} r_1(x, y), & \text{if } (x, y) \in S_1 \\ \quad '' & \quad '' \\ \quad '' & \quad '' \\ r_n(x, y), & \text{if } (x, y) \in S_n. \end{cases} \tag{3.1-5}$$

The approximation error for the whole picture is, consequently:

$$G(P) = \sum_{S_i \in P} H(S_i) \tag{3.1-6}$$

where $P = \{S_1, S_2, \cdots S_n\}$. The minimum value for G results necessarily from the sum of the minimum values for $H(S_i)$, where $H(S_i) = H_{\min}(S_i)$. The picture approximation problem consists then in finding the partition P that minimizes the global criterion G. Other forms of global criterion can also be used; for example, Horowitz and Pavlidis [16] use the Chebyshev (min–max) norm.

The importance of the number of segments n in the minimization of $G(P)$ must be stressed. The minimum value $G_{\min}(P)$ is monotonically nonincreasing with increasing number of segments for P. Therefore, a picture approximation problem consists in finding the partition P_n such that

$$G(P_n) = \min_{P_n} \{G(P_n')\} \tag{3.1-7}$$

where P_n and P_n' are picture partitions with n segments.

B. Stepwise Optimization for Picture Segmentation

The identification of the partition $P_{\min}$ that minimizes a global criterion or cost function G requires a search over the entire space of all possible picture partitions $\{P\}$. The implementation of an exhaustive search is prohibitive because of the large size of the $\{P\}$ space. One possible approach is to constrain the search to a subset of $\{P\}$, $U \subset \{P\}$, which only guarantees a suboptimum solution. Examples of this approach are "gradient descent" techniques which consist in moving pixels from one segment to another if such moves improve the global criterion or cost function G [23], [26].

A hierarchical data structure can also be employed to define a useful subset of picture partitions [16]. A hierarchy of segments can be represented by a segment tree in which nodes correspond to segments. Each segment S_i^{τ} is linked to the segments of the lower level $S_{i,1}^{\tau-1}$, $S_{i,2}^{\tau-1}$, $\cdots$ which are disjoint subsets of S_i^{τ}, and which are called "sons" of S_i^{τ}. Therefore, a picture partition corresponds to a subset of these tree nodes.

A picture segmentation algorithm which involves the construction of a segment tree as the result of a sequence of stepwise optimizations is now introduced. The presentation is similar to the one proposed by Ward [35] for hierarchical clustering. It requires a global criterion or cost function $G(P)$ which reflects the cost or loss of information resulting by representing the picture with the partition P.

An initial picture partition $P^0 = \{S_1^0, S_2^0, \cdots S_n^0\}$ with n segments is first defined. At the kth iteration, the algorithm merges two segments from the P^{k-1} partition to produce a new partition $P^k = \{S_1^k, S_2^k \cdots S_{n-k}^k\}$. As the number of segments is decreased by one at each iteration, P^k must contain $n - k$ segments. $G(P^k)$ tends generally to increase from step to step and can be written as:

$$G(P^k) = G(P^0) + \sum_{\tau=1}^{k} G(P^{\tau}) - G(P^{\tau-1}). \tag{3.2-1}$$

The minimization of $G(P^k)$ is therefore associated with the minimization of each term of the summation which corresponds to the increase of G at each iteration. Thus, the global optimization problem is reduced to a sequence of stepwise optimizations. However, the minimization of each term, $G(P^{\tau}) - G(P^{\tau-1})$, yields the global optimum for $G(P^k)$ only if the terms are independent, which is not necessarily the case. Nevertheless, it can constitute an interesting suboptimal approach.

The goal of the stepwise optimization is therefore to find the two segments whose merger produces the smallest increase of G. For the picture approximation problem, G increases monotonically with the number of iterations, k,

$$G(P^0) \leq G(P^1) \cdots \leq G(P^{\tau}) \cdots \leq G(P^k).$$

This increase results from the merging of two segments S_i and S_j, and can easily be calculated from (3.1-6):

$$C_{i,j} = H(S_i \cup S_j) - H(S_i) - H(S_j). \tag{3.2-3}$$

The only terms of $G(P)$ that are affected by the merging are $H(S_i)$ and $H(S_j)$ which are replaced by $H(S_i \cup S_j)$. Thus, $C_{i,j}$ is the stepwise criterion to be optimized. So, each iteration involves:

1) the identification of all pairs of connected segments (S_i, S_j),
2) the calculation of $C_{i,j}$,
3) the selection of the lowest $C_{i,j}$, and
4) the merging of the two corresponding segments.

It must be noted that the algorithm does not guarantee that P^k will optimize $G(P)$ amongst all the partitions with $n - k$ segments. Nevertheless, it yields good results as will be seen, and the implied hierarchical data structure can constitute an advantage for many applications.

C. Picture Approximation by Constant Value Regions

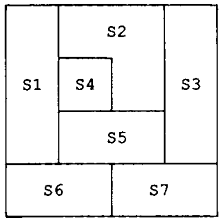
The stepwise optimization algorithm (HSWO) described in Section II-A can be adapted to the global optimization case by an appropriate definition of the stepwise criterion. As a particular example, the concepts are illustrated by the piecewise approximation of a picture, $f(x, y)$, by constant value regions.

In constant piecewise approximation (or zero order approximation), a picture is divided into segments S_i which are approximated by their mean values, μ_i:

$$r_i(x, y) = \mu_i. \tag{3.3-1}$$

1	2	2	13
1	10	2	13
1	3	3	13
6	6	10	10

(a)



(b)

Fig. 2. A small picture with its initial partition. (a) Gray level values. (b) Initial partition.

The segment approximation error or the segment cost is then the sum of the squared differences between the pixel values and the segment mean μ_i:

$$H(S_i) = \sum_{(x,y) \in S_i} \left(f(x, y) - \mu_i\right)^2. \quad (3.3\text{-}2)$$

In order to use the HSWO algorithm, the stepwise criterion and the segment descriptive parameters D_i must be defined. The stepwise criterion is as given before:

$$C_{i,j} = H(S_i \cup S_j) - H(S_i) - H(S_j) \quad (3.3\text{-}3)$$

which can now be rewritten as:

$$C_{i,j} = \frac{N_i \cdot N_j}{N_i + N_j} (\mu_i - \mu_j)^2 \quad (3.3\text{-}4)$$

where N_i and N_j are the number of pixels in S_i and S_j. The stepwise minimization of $C_{i,j}$ therefore results in the merger that minimizes the increase in the overall pixel variance around the segment means. The segment descriptive parameters D_i needed to calculate the criterion are the segment size N_i and mean μ_i. The definition of a stopping condition is discussed in more detail in Section V.

Approximations with higher order polynomials can be developed in a similar manner. For example, the case of planar approximation has been analyzed [4], [5]. A segment S_i is approximated by a plane:

$$r_i(x, y) = a^i_{0,0} + a^i_{1,0}(x) + a^i_{0,1}(y) \quad (3.3\text{-}5)$$

and the sum of the squared errors is used as the segment cost $H(S_i)$.

D. An Illustrative Example

The operation of the algorithm is now illustrated by a simple example. Fig. 2 shows a 4×4 pixel picture with 7 initial constant level segments. The algorithm starts with an initial partition P^0 of 7 segments, $S_1, S_2, \cdots S_7$. Fig. 3 shows a segment tree which represents the sequence of segment mergers; the algorithm sequentially creates the segments S_8–S_{13}. Table I shows the sets of criterion values, $\{C_{i,j}\}$, used at each iteration, with the minimum enclosed by a rectangle.

To further illustrate the operation of the HSWO algorithm, consider the first iteration shown in Table I. The minimum criterion value corresponds to $C_{2,5} = 1.2$; thus segments S_2 and S_5 are merged to form the segment S_8; the neighborhoods and the criterion values are then updated. The process is repeated and yields a hierarchy of partitions, with a reasonable stopping point being S_{10} and S_{12}.

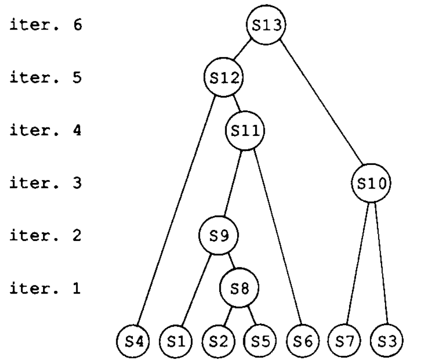


Fig. 3. Sequence of segment mergers.

TABLE I
SETS OF CRITERION VALUES $C_{i,j}$

i, j	$C_{i,j}$	Sets of criteria for each iteration it.1	it.2	it.3	it.4	it.5	it.6
1, 2	1.5	1.5					
1, 4	60.7	60.7	60.7				
1, 5	4.8	4.8					
1, 6	30.0	30.0	30.0				
2, 3	181.5	181.5					
2, 4	48.0	48.0					
2, 5	1.2	$\boxed{1.2}$					
3, 5	120.0	120.0					
3, 7	10.8	10.8	10.8	$\boxed{10.8}$			
4, 5	32.7	32.7					
5, 6	9.0	9.0					
5, 7	49.0	49.0					
6, 7	16.0	16.0	16.0	16.0			
8, 1	3.7		$\boxed{3.7}$				
8, 3	210.7		210.7				
8, 4	48.1		48.1				
8, 6	18.5		18.5				
8, 7	82.5		82.5				
9, 3	270.0			270.0			
9, 4	58.7			58.7	58.7		
9, 6	27.2			27.2	$\boxed{27.2}$		
9, 7	105.6			105.6			
10, 6	48.1				48.1		
10, 9	303.1				303.1		
11, 4	48.4					$\boxed{48.4}$	
11,10	277.0					277.0	
12,10	244.6						$\boxed{244.6}$

IV. Probability of Error in Hierarchical Segmentation

Pattern recognition and picture analysis are often regarded as statistical decision processes. In this section, it is first shown that statistical hypothesis testing can be employed for picture segmentation. A picture is regarded as composed of regions with different gray level probability density functions. A picture segmentation can be produced by testing and merging two segments if they belong to the same region. There is two types of error that can occur: type I error occurs when two similar segments are kept disjoint, and type II occurs when dissimilar segments are merged. In general, decision processes can be evaluated by cost functions which assign different weights to the different decision outcomes. This cost function must also take into account the interdependencies between each decision, which is very difficult for applications such as picture segmentation problems. Also, in the hypothesis testing approach, the cost function has been reduced to the probability of error. In previous works [17], [14], [9], tests are designed such that the probability of type I error does not exceed threshold values, the evaluation of the probability of type II error being usually impossible in real applications.

This section will examine the sequential testing aspect of hierarchical segmentation. It is shown that type II errors are the most important and it is therefore advantageous to minimize its probability. This is achieved by a stepwise optimization process which finds and merges the most similar segment pair. This process corresponds to the HSWO algorithm presented previously. Moreover, an example illustrates the relation between the picture approximation and the hypothesis testing approach: using the same picture model, both approaches result in the same stepwise criterion. Following a statistical approach, picture segmentation can also be regarded as a best estimate problem, and in the Appendix, it is shown how the best estimate of a picture partition can be rewritten as a picture approximation problem.

A. Hypothesis Testing

The statistical picture model employed for the picture segmentation is now presented. It is assumed that a picture $f(x, y)$ is composed of distinct regions $\{R_k\}$, where each region is viewed as a statistical population and is defined by its probability density function, PDF_k:

$$f(x, y) \approx \mathrm{PDF}_k, \qquad \text{for } (x, y) \in R_k. \quad (4.1\text{-}1)$$

The goal of a picture segmentation process is then to find the true picture partition $\{R_k\}$. Let S_i designate any arbitrary subpart of a true region R_k, $S_i \subset R_k$. Therefore, a hypothesis testing approach can be used to decide if the observed values of the segment S_i, $f(x, y)$, for $(x, y) \in S_i$, belong to PDF_k.

The merging of segments can also be based upon hypothesis testing. For two arbitrary adjacent segments, S_i and S_j, the usual statistical test would determine if they belong to the same true region R_k: $S_i \subset R_k$ and $S_j \subset R_k$. However, as the characteristics of R_k are unknown, the statistical decision must instead consider whether the pixel values of the two segments come from the same probability density function.

A test is usually described in terms of some statistic d which is a reduction of the observed data [20]. Hence, let d be a measure of the similarity of the estimated probability density functions of segments S_i and S_j. A statistical decision process can then be used to determine which one of the following two hypotheses is true.

$$H_0 : d = 0,$$
$$H_1 : d = d_{\text{true}}, \quad (4.1\text{-}2)$$

where the hypothesis H_0 indicates that the two segments belong to the same true region, while the hypothesis H_1 defines segments belonging to different regions.

The statistical decision therefore consists of accepting H_0 if d is small, more precisely, if $d \le t$, where t is a selected threshold (see Fig. 4). The performance of a test is judged according to its tendency to lead to wrong decisions. Two types of error are considered:

Type I: rejecting H_0 when H_0 is true

Type II: accepting H_0 when H_1 is true. (4.1-3)

The probability of these two types of error are represented, respectively, by α and β. They must both be low for a good decision process. The threshold value t can be modified such as to reduce either α or β, but not both simultaneously. Therefore, some compromise must be reached; for example, to select a t such that $\alpha = \beta$.

An example is now considered where it is assumed that an ideal picture is composed of constant value regions corrupted by a uniform Gaussian white noise. Each pixel value $f(x, y)$ inside a region R_k, $(x, y) \in R_k$, is regarded as a random variable, with Gaussian distribution of mean m_k and variance σ^2, $N(m_i, \sigma^2)$. Therefore, the difference of the segment means may be used to decide if two segments belong to the same region;

$$d_{i,j} = |\mu_i - \mu_j|, \quad (4.1\text{-}5)$$

where μ_i and μ_j are the mean values of segments S_i and S_j.

This statistic is employed to test the two hypotheses, H_0 and H_1, and the probabilities of types I and II error are calculated. If H_0 is true, $\mu_i - \mu_j$ has a Gaussian distribution with a zero mean and a variance of σ_d^2:

$$\sigma_d^2 = (1/N_i + 1/N_j)\,\sigma^2 \quad (4.1\text{-}6)$$

where N_i and N_j are, respectively, the sizes of segments S_i and S_j. The H_0 hypothesis is accepted if $d \le t$, and, therefore, the probability of type I error is:

$$\alpha = 1 - \int_{-t}^{t} \frac{1}{\sqrt{2\pi}\,\sigma_d} \exp\left(\frac{-x^2}{2\sigma_d^2}\right) dx. \quad (4.1\text{-}7)$$

If H_1 is true, $\mu_i - \mu_j$ has also a Gaussian distribution with the same variance, σ_d^2, but with mean d_{true}. The probability of type II error is then:

$$\beta = \int_{-t}^{t} \frac{1}{\sqrt{2\pi}\,\sigma_d} \exp\left(\frac{-(x - d_{\text{true}})^2}{2\sigma_d^2}\right) dx. \quad (4.1\text{-}8)$$

Low values for α and β can be achieved, simultaneously, only if d_{true}/σ_d is large. As σ_d decreases with the segment sizes, the probabilities of errors are smaller for decisions involving larger segments. This is illustrated in Table II which gives the α and β values for different segment sizes, with $d_{\text{true}} = 3\sigma$ and $t = 1.5\,\sigma$.

B. Sequential Testing in Hierarchical Segmentation

Hierarchical segmentation begins with many small segments which are sequentially merged to produce larger

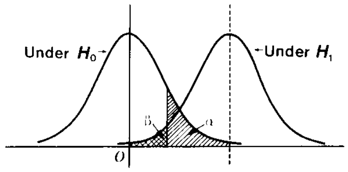


Fig. 4. Probabilities of error in hypothesis testings.

TABLE II
PROBABILITIES OF ERRORS FOR DIFFERENT SEGMENT SIZES WITH $d_{true} = 3\sigma$ AND $t = 1.5\sigma$

Segment sizes	α	β
1 pixel	.289	.144
2 pixels	.134	.067
4 pixels	.034	.017

ones. Statistical decision can be employed to determine whether, or not, two adjacent segments must be merged. However, the sequential aspect of hierarchical segmentation must be considered in the design of the decision process. It can be noted that type II error results from merging of two different segments, and therefore, cannot be recovered by an agglomerative process. Whereas, type I error keeps separated two similar segments which can be corrected in a following step. Therefore, it seems preferable to keep β at a low level to avoid type II errors. To develop this point, the hierarchical segmentation is now regarded as a sequential testing process.

In sequential testing, it is the error probability of the final result that must be considered, not the error probabilities of each individual test. However, the relation between the error probability of the final result and those of each test is often difficult to derive. Sequential testing is involved in hierarchical segmentation. With the utilization of the appropriate assumptions, some relations between the error probabilities of each test and of the final result can be demonstrated. Hence, it will be shown that the probability of type I error, α, for the final result is equal or lower than those of the individual tests, while the probability of type II error, β, for the final result is equal or higher than those of the individual tests. This means that in hierarchical segmentation, it is advantageous to keep β at an appropriately low level for each test, even if large α values must be used in the first tests.

A two stage test for merging is first examined. In stage 1 (see Fig. 5), the segments S_i^1 and S_j^1 are compared by a first test, test #1. If the segments are not merged after this test, they will, sooner or later, be involved in a second test. Before this second comparison, S_i^1 and/or S_j^1 are merged with some adjacent segments belonging to the

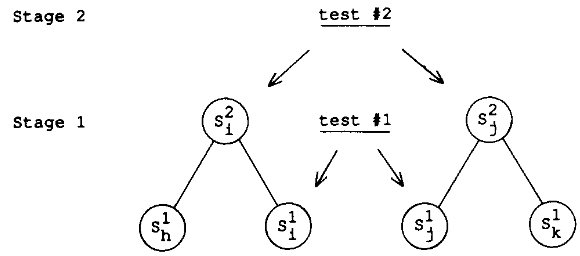


Fig. 5. Sequence of segment testings in a hierarchy.

same regions in order to produce S_i^2 and S_j^2, $S_i^1 \subset S_i^2 \subset R_i$ and $S_j^1 \subset S_j^2 \subset R_j$. The second stage test, test #2, considers therefore the segments S_i^2 and S_j^2 in which S_i^1 and S_j^1 are still disjoint. The same hypotheses, H_0 versus H_1, are employed in both stages. Let α_1, β_1 be the probabilities of errors for test #1, and α_2, β_2 for test #2. The probabilities of errors for the combined test are designated by α_{1+2} and β_{1+2}. If H_0 is accepted in test #1, then as the segments are merged, test #2 is not needed.

If H_0 is true, we have:

$$\text{Prob (accept } H_0 \text{ at test \#1)} = 1 - \alpha_1$$

$$\text{Prob (accept } H_0 \text{ at test \#2)} = \alpha_1(1 - \alpha_2^*) \quad (4.2\text{-}1)$$

where α_2^* is the probability that test #2 rejects H_0 when H_0 has been rejected by test #1, $\alpha_2 \le \alpha_2^* \le 1$. Then, we obtain:

$$\alpha_{1+2} = 1 - \text{Prob (accept } H_0 \text{ at test \#1)} - \text{Prob (accept } H_0 \text{ at test \#2)}$$

$$\alpha_{1+2} = \alpha_1\alpha_2^*. \quad (4.2\text{-}2)$$

If H_1 is true, we have

$$\text{Prob (accept } H_0 \text{ at test \#1)} = \beta_1$$

$$\text{Prob (accept } H_0 \text{ at test \#2)} = (1 - \beta_1)\beta_2^* \quad (4.2\text{-}3)$$

where β_2^* is the probability that test #2 accepts H_0 when H_0 has been rejected by test #1, $0 \le \beta_2^* \le \beta_2$. Thus, we obtain:

$$\beta_{1+2} = \text{Prob (accept } H_0 \text{ at test \#1)} + \text{Prob (accept } H_0 \text{ at test \#2)}$$

$$\beta_{1+2} = \beta_1 + (1 - \beta_1)\beta_2^* \approx \beta_1 + \beta_2^* \quad (4.2\text{-}4)$$

where the term $\beta_1\beta_2^*$ is usually small and may be ignored.

It can be noted that if the two tests are identical (i.e., if they always give the same results), then $\alpha_2^* = 1$ and $\beta_2^* = 0$. If the tests are independent (i.e., if the results of tests #1 do not affect the results of test #2) then $\alpha_2^* = \alpha_2$ and $\beta_2^* = \beta_2$.

If a third step is then added to the process, we obtain:

$$\alpha_{1+2+3} = (\alpha_{1+2})\alpha_3^* = \alpha_1\alpha_2^*\alpha_3^*$$

$$\beta_{1+2+3} \approx (\beta_{1+2}) + \beta_3^* \approx \beta_1 + \beta_2^* + \beta_3^* \quad (4.2\text{-}5)$$

and, more generally, for an m step process, we have:

$$\alpha_{1+\cdots+m} = \alpha_1 \alpha_2^* \cdots \alpha_m^*$$

$$\beta_{1+\cdots+m} \approx \beta_1 + \beta_2^* + \cdots + \beta_m^*. \quad (4.2\text{-}6)$$

The probability of type I error is therefore reduced from one stage to the next. Hence, a high value can be assigned to α_1, as the following tests will reduce the overall α. It is easily shown that:

$$\alpha_{1\cdots+m} \leq \text{minimum}\ (\alpha_1, \alpha_2, \cdots, \alpha_m). \quad (4.2\text{-}7)$$

On the other hand, the probability of type II error increases from stages to stages. A large β at the first stage cannot be subsequently reduced.

$$\beta_{1\cdots+m} \geq \text{maximum}\ (\beta_1, \beta_2, \cdots, \beta_m). \quad (4.2\text{-}8)$$

An upper bound for $\beta_{1\cdots+m}$ is given by:

$$\beta_{1\cdots+m} \leq \beta_1 + \beta_2 + \cdots + \beta_m. \quad (4.2\text{-}9)$$

As an example, Table III shows the probabilities of errors for a three stage process using the same threshold value, $t = 1.5\ \sigma$. The segment sizes are, respectively, 1, 2, and 4 pixels for stages 1, 2, and 3. As noted before, the probabilities of errors decrease with the segment sizes. The progression of the lower and upper bounds of β_{1+2+3}, is also reported. The bound for β_{1+2+3} will usually be determined by test #1 where the value of β_1 (= 0.144) is high, $0.144 \leq \beta_{1+2+3} \leq 0.228$. The upper bound for α_{1+2+3}, which is the minimum of α_k values, will instead be determined by test #3, $\alpha_{1+2+3} \leq \alpha_3 = 0.034$.

By reducing the threshold values of tests #1 and #2, the β_{1+2+3} bounds can be reduced without changing the upper bound of α_{1+2+3}. In Table IV, the threshold values are chosen such that β_k values are small and almost equal for the three stages. This results in smaller bounds for β_{1+2+3}, $0.017 \leq \beta_{1+2+3} \leq 0.044$. On the other hand, the corresponding increases in α_1 and α_2 have not changed the upper bound of α_{1+2+3}, which is still determined by test #3, $\alpha_{1+2+3} \leq \alpha_3 = 0.034$.

In hierarchical segmentation, it is therefore advantageous to start with large α values in the first stages and then subsequently reduce the α's, in order to keep β at an appropriately low level for each stage. This concept is exploited in the next section.

C. Stepwise Optimization

In hierarchical segmentation, it is preferable for each stage to keep β_k, the probability of type II error, as low as possible. But usually β_k cannot be evaluated because d_{true} is unknown. Also, the probability of type I error, α_k, must be employed instead to select the appropriate threshold value t. Thus, the evaluation of the maximum value allowed for α_k at stage k is now examined, the maximization of α_k being associated with the minimization of β_k.

At each stage or segment level, there are many possible segment mergers, which can be represented by segment pairs (S_i, S_j). The segment similarity statistic, $d_{i,j}$, can be calculated for each pair. A statistical decision process

TABLE III
PROBABILITIES OF ERROR FOR SEQUENTIAL TESTING WITH THE SAME THRESHOLD ($d_{\text{true}} = 3\ \sigma$)

Test	Threshold	α_k	β_k	$\beta_{1\ldots+k}$ bound lower	upper
# 1	1.5 σ	.289	.144	.144	.144
# 2	1.5 σ	.134	.067	.144	.211
# 3	1.5 σ	.034	.017	.144	.228

TABLE IV
PROBABILITIES OF ERROR FOR SEQUENTIAL TESTING WITH DIFFERENT THRESHOLDS ($d_{\text{true}} = 3\ \sigma$)

Test	Threshold	α_k	β_k	$\beta_{1\ldots+k}$ bound lower	upper
# 1	.25 σ	.859	.015	.015	.015
# 2	.75 σ	.453	.012	.015	.027
# 3	1.5 σ	.034	.017	.017	.044

accepts the hypothesis H_0 and merges segments only if:

$$d_{i,j} \leq t \quad (4.3\text{-}1)$$

which can be rewritten as:

$$v_{i,j} \leq 1 - \alpha \quad \text{or} \quad \alpha \leq 1 - v_{i,j} \quad (4.3\text{-}2)$$

where $v_{i,j}$ is the confidence level associated with the interval $(0, d_{i,j})$, i.e., $v_{i,j}$ is the probability of obtaining a value d such that $d \leq d_{i,j}$:

$$v_{i,j} = \text{prob}\ (d \leq d_{i,j};\ H_0). \quad (4.3\text{-}3)$$

Defining $v_{\min}$ as the minimum over $v_{i,j}$, $v_{\min} = \min(v_{i,j})$, then the utilization of an α greater than $1 - v_{\min}$ implies that no segments are merged, which renders the stage redundant. Therefore, the maximum allowed value for α is $\alpha_{\max} = 1 - v_{\min}$, which results at least in one merger. Hence, a hierarchical segmentation algorithm can employ a stepwise process which finds the segment pair with the minimum confidence level $v_{i,j}$ and merges the corresponding segments. This is equivalent to using the maximum allowed α value for each stage.

Step-wise optimization, by maximizing α_k, assures that, at each step, the probability of type II error β_k is kept to it lowest value. This should also keep $\beta_{1+\cdots+m}$ at a low value. The stepwise optimization algorithm (HSWO) described in Section II-A can therefore be employed here with the appropriate stepwise criterion.

The stepwise criterion for the previously discussed example (Section IV-A) is now derived. This example considered a picture composed of constant value regions and corrupted by a uniform Gaussian white noise. A statistic

$d_{i,j}$ related to the difference of segment means was then introduced. The confidence level associated with the interval $(0, d_{i,j})$ under the H_0 hypothesis can be derived:

$$v_{i,j} = \text{prob}\,(d \leq d_{i,j};\, H_0)$$

$$v_{i,j} = \int_{-d_{i,j}}^{d_{i,j}} \frac{1}{\sqrt{2\pi}\,\sigma_d} \exp\left(\frac{-x^2}{2\sigma_d^2}\right) dx$$

$$v_{i,j} = 2\,\text{erf}\,(d_{i,j}/\sigma_d) \tag{4.3-4}$$

where

$$\sigma_d^2 = (1/N_i + 1/N_j)\sigma^2$$

$$\text{erf}\,(y) = \int_0^y \frac{1}{\sqrt{2\pi}} \exp\,(-x^2/2)\, dx.$$

Note that the segment pair, S_i and S_j, that minimizes $d_{i,j}/\sigma_d$ will also minimize $v_{i,j}$; therefore $d_{i,j}/\sigma_d$ may be used as a stepwise criterion:

$$\frac{d_{i,j}}{\sigma_d} = \sqrt{\frac{N_i N_j}{N_i + N_j}}\,\frac{|\mu_i - \mu_j|}{\sigma} \tag{4.3-5}$$

where μ_i, μ_j and N_i, N_j are, respectively, the means and the sizes of segment S_i and S_j, and σ^2 is the variance of noise.

This criterion is similar to the criterion derived for picture approximation with constant value regions (Section III-C);

$$\frac{d_{i,j}}{\sigma_d} = \frac{\sqrt{C_{i,j}}}{\sigma} \tag{4.3-6}$$

and will produce the same results. This is achieved because constant value regions are used in both cases, and the mean squared error norm is employed with the uniform Gaussian white noise. This illustrates the relation between the picture approximation and the hypothesis testing approaches. Moreover, following a statistical approach, picture segmentation can also be regarded as a best estimate problem, and in the Appendix, it is shown how the best estimate of a picture partition can be rewritten as a picture approximation problem.

The stepwise optimization criterion for picture segmentation by hypothesis testing can also be derived for other picture models. In each case, the statistic d and its distribution under H_0 (which is required to calculate the confidence level) must be derived. For example, in many cases, the likelihood ratio statistic could be considered [20], [17].

V. Experimental Results

A hierarchical segmentation algorithm based upon stepwise optimization has been described and analyzed in the preceding sections. Picture segmentation is first presented as a picture approximation problem. Segmentation is then regarded as an hypothesis testing process, and the probability of error in hierarchical segmentation is analyzed. In both cases, constant value regions are used to model the pictures, and to derive the segmentation criterion. However, it is always considered that the pictures possess a hierarchical structure, which provides a more realistic modeling. The hierarchical stepwise optimization algorithm is now employed to segment two pictures. This will illustrate the usefulness of the model and the operation of the algorithm. A more complete evaluation is presented in [5] with the utilization of multispectral LANDSAT and SAR images and other segment models.

A. A Checkerboard Example

The hierarchical stepwise optimization algorithm (HSWO) is first applied to a checkerboard example, where the two tones are designated by m_1 $(= -2.0)$ and m_2 $(= +2.0)$. The checkerboard is corrupted by a Gaussian white noise with zero mean and variance equal to one (see Fig. 6). The picture size is of 64 by 64 pixels, and, in the initial partition, each pixel corresponds to one segment. After 4095 mergers (iterations), the number of segments is reduced to one, the entire picture. Let $C_{\min,k}$ designate the minimum criterion value at iteration k. The first steps of the algorithm yield $C_{\min,k}$ values close to zero, which increase in the following steps as larger segments are merged. The resultant minimum criterion values, $C_{\min,k}$, for the last 150 mergers are shown in Fig. 7. The abscissa corresponds to the number of segments in the partition (increasing from left to right), or to the iteration number (increasing from right to left). There is an important jump in the $C_{\min,k}$ values, where the merging of dissimilar segments (squares) begins. The arrow indicates an appropriate stopping point just before this occurs, and Fig. 8 presents the resultant picture segmentation. The situation in real images is more complex.

B. Segmentation of a Remote Sensing Picture

Fig. 9 shows a 32 × 32 Landsat satellite picture (0.8–1.1 μm band) of an agricultural area near Melfort in Saskatchewan, Canada. The picture is initially divided into 1024 regions of one pixel each, and is segmented by the HSWO algorithm using the increase of the constant approximation error as the stepwise criterion. This picture possesses a rather complex structure with regions having varying sizes and mean value differences. In Fig. 10, the sequence of $C_{\min,k}$ values are drawn as a function of the number of segments contained in the picture partition at step k. This sequence does not indicate a unique stopping point as in the checkerboard sample.

In Fig. 11, four partitions are presented containing, respectively, (a) 18 segments, (b) 36 segments, (c) 118 segments, and (d) 212 segments. Partition (a) divides the picture into what seems to be its most basic parts. These basic regions can moreover be considered as composed of finer elements, which can be obtained by reducing the number of segment mergers as shown by partition (b). Consider, for example, segment 1 in Fig. 11(a) which is divided into segments 2 and 3 in Fig. 11(b). There is a small gray level

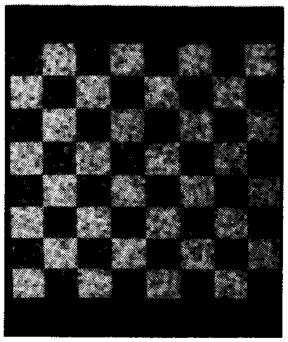


Fig. 6. Checkerboard picture.

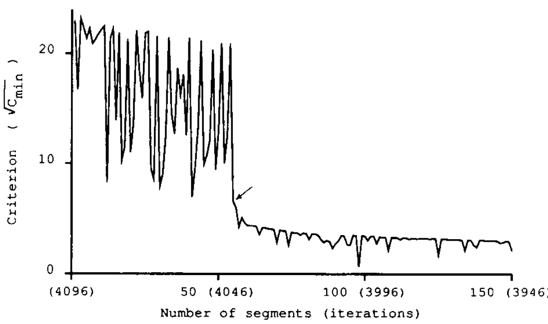
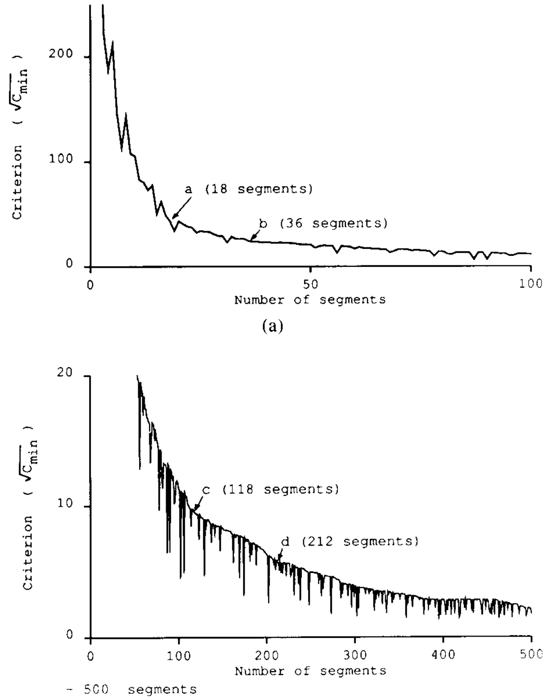


Fig. 7. Minimum criterion value curve. The abscissa corresponds to the number of segments in the partition (increasing from left to right), or to the iteration number (increasing from right to left).



(b)

Fig. 10. Minimum stepwise criterion curve for different axis scales. The abscissa corresponds to the number of segments in the partition which is related to the iteration number (increasing from right to left). (a) 0–100 segments. (b) 0–500 segments.

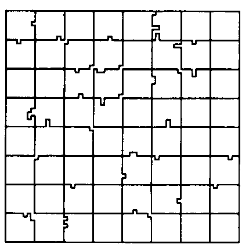
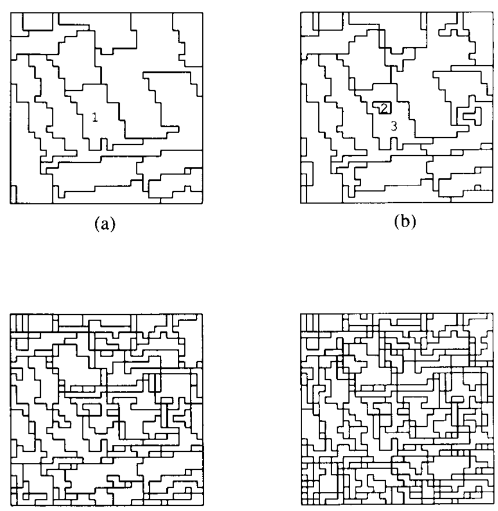


Fig. 8. Segmentation of the checkerboard picture.



(c) (d)

Fig. 11. Segmentations of the Landsat picture. (a) 18 segments. (b) 36 segments. (c) 118 segments. (d) 212 segments.

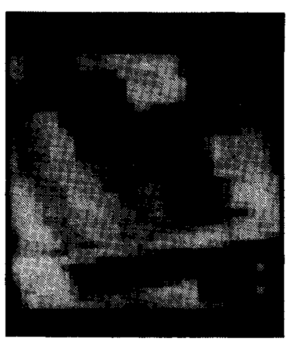


Fig. 9. A Landsat satellite picture (32 × 32 pixels, 0.8–1.1 μm band).

difference between 2 and 3, thus, region 2 represents a finer picture component than region 1. Partitions (c) and (d) correspond to still finer picture segmentations.

The segmentation algorithm performs a hierarchical decomposition of the picture where the hierarchical levels could be related to the resolution levels. Hence, a region, which is higher in the hierarchy than its subparts, is also larger than its subparts. In the remote sensing picture, the partition with 18 segments can be regarded as the highest

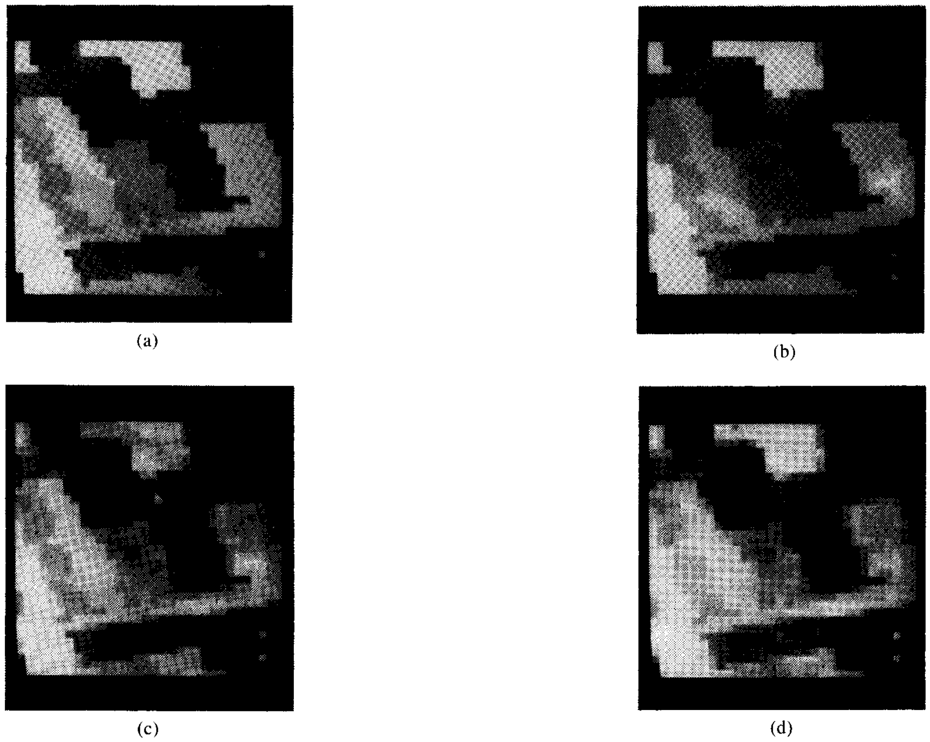


Fig. 12. Approximations of the Landsat picture. (a) 18 segments. (b) 36 segments. (c) 118 segments. (d) 212 segments.

level where only the most important components of the picture are preserved. This is illustrated by Fig. 12(a) where each picture segment has been replaced by its mean value. These segments encode the gross information of the picture; they indicate the most prominent areas. The other partitions of Fig. 11 correspond to splitting these segments into subunits. The corresponding approximation pictures are shown in Fig. 12 and indicate that finer picture components are retained. These picture partitions can be regarded as different levels of the hierarchy which correspond to different picture component resolutions.

One important consequence of this hierarchical structure for the stepwise optimization algorithm is that the user must specify at which level to stop the segment merging. The segment level can be defined by the approximation error, by the $C_{\min}$ value, or by the number of segments in the partition, each of these parameters being interrelated. The examination of the approximation error and $C_{\min}$ curves can then complement the context knowledge in order to select a stopping point.

The hierarchical decomposition of the picture contains more information than one particular partition. This information could be exploited advantageously by a high level process. For example, the $C_{\min}$ curve can be examined in order to know when to stop the segment merging and obtain an appropriate partition. The hierarchical decomposition could also be used to characterize picture structure. Different types of texture, for example, could result into different types of hierarchical decomposition. More work is required in order to understand what kind of information is present in the decomposition, and how a high level process could take advantage of it.

VI. Conclusion

The importance of providing a precise definition of the picture segmentation task in low level terms has been stressed. This has revealed the need for generic constraints that could be effectively introduced into the segmentation stage. The hierarchical structure of a picture is one such important generic constraint, and it has been shown to efficiently exploit this structure that the utilization of a stepwise optimization rule is required. A new hierarchical segmentation algorithm based upon stepwise optimization (HSWO) has been introduced. It starts with an initial picture partition, and merges at each iteration the two most similar segments found by the optimization of a "stepwise criterion." In contradistinction to previous hierarchical segmentation algorithms based upon logical predicate equations, the HSWO algorithm employs a more global and gradual strategy.

Regarding picture segmentation as an optimization problem is useful in providing a precise definition of a segmentation task. In this paper, picture segmentation is regarded as the piecewise polynomial approximation of a picture. The approximation error is then employed as a global criterion, and the picture segmentation consists therefore in finding the partition that minimizes this criterion. It is shown that the addition of a hierarchical structure constraint reduces the search space and the global optimization problem is replaced by a sequence of stepwise optimizations, where the stepwise criterion is derived from the global criterion.

Pattern recognition and picture analysis are often defined in probabilistic terms. Hence, picture segmentation can be regarded as an hypothesis testing process which merges only segments that belong to the same region. Two types of error may occur: type I error occurs when two similar segments are kept disjoint, and type II error occurs when dissimilar segments are merged. It is stressed that, in hierarchical segmentation, type II error is the most important and it is therefore advantageous to minimize the probability of this error. This is achieved by a stepwise optimization process which finds and merges the most similar segment pair.

It has also been shown experimentally that the HSWO algorithm can correctly detect and express the hierarchical structure of a picture. The sequence of the criterion values selected at each iteration provides useful information about the picture structure, and can indicate convenient stopping points for segment merging.

Appendix
The Best Estimate of a Picture Partition

The best estimate of the true picture partition is derived and shown to correspond to the picture partition that minimizes the approximation error. Let $R = \{R_i\}$ be the true picture partition, m_i be the constant value for region R_i, and $f(x, y)$ be the observed picture value

$$f(x, y) = m_i + e(x, y); \qquad \text{for } (x, y) \in R_i \quad \text{(A-1)}$$

where $e(x, y)$ are Gaussian independent random variables with zero mean. Then, the best estimate $\hat{R} = \{\hat{R}_i\}$ of the picture partition maximizes the likelihood function

$$\begin{aligned} L(\hat{R}; f) &= \text{prob}\,(f; \hat{R}) \\ &= \prod_{(x,y)} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\left(f(x, y) - \hat{m}_{(x,y)}\right)^2/2\sigma^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\sum_{(x,y)} \left(f(x, y) - \hat{m}_{(x,y)}\right)^2/2\sigma^2\right\} \end{aligned} \quad \text{(A-2)}$$

where $\hat{m}_{(x,y)} = \hat{m}_i$ for $(x, y) \in \hat{R}_i$, $\hat{m}_i$ being the constant value for region $\hat{R}_i$, and where n is the number of pixels in the picture. Thus, maximizing $L(\hat{R}; f)$ corresponds to minimizing

$$\sum_{(x,y)} \left(f(x, y) - \hat{m}_{(x,y)}\right)^2 \quad \text{(A-3)}$$

which can be rewritten as

$$\sum_{\hat{R}_i} \sum_{(x,y) \in \hat{R}_i} \left(f(x, y) - \hat{m}_i\right)^2. \quad \text{(A-4)}$$

The best estimate thus corresponds to the partition with the lowest approximation error, and it can be shown that the best value for $\hat{m}_i$ is the mean value of the region $\hat{R}_i$.

References

[1] J. W. Bacus, W. A. Yasnoff, and M. G. Belanger, "Computer techniques for cell analysis in hematology," in *Proc. First Annu. Symp. Computer Applications in Medical Care*, Oct. 1977, pp. 24-35.

[2] R. Bajcsy and D. A. Rosenthal, "Visual and conceptual focus of attention," in *Structured Computer Vision*, S. Tanimoto and A. Klinger, Eds. New York: Academic, 1980, pp. 133-149.

[3] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[4] J. M. Beaulieu and M. Goldberg, "Hierarchical picture segmentation by approximation," in *Proc. Canadian Communications and Energy Conf.*, Montreal, P.Q., Canada, 1982, pp. 393-396.

[5] J. M. Beaulieu, "Hierarchical picture segmentation by step-wise optimization," PH.D. dissertation, Dep. Elec. Eng., Univ. Ottawa, Ottawa, Ont., Canada, 1985.

[6] T. E. Boult and J. R. Kender, "On surface reconstruction using sparse depth data," in *Proc. Image Understanding Workshop*, Miami Beach, FL, 1985, pp. 197-208.

[7] C. Brice and C. Fennema, "Scene analysis using regions," *Artificial Intell.*, vol. 1, pp. 205-226, 1970. 参考

[8] S. G. Carlton and O. R. Mitchell, "Object/background segmentation in FLIR imagery," in *Proc. Conf. Pattern Recognition and Image Processing*, 1978, pp. 360-362.

[9] P. C. Chen and T. Pavlidis, "Image segmentation as an estimation problem," *Comput. Graphics Image Processing*, vol. 12, pp. 153-172, 1980.

[10] L. S. Davis, "A survey of edge detection techniques," *Comput. Graphics Image Processing*, vol. 4, pp. 248-270, 1975.

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

[12] E. C. Freuder, "Affinity: A relative approach to region finding," *Comput. Graphics Image Processing*, vol. 5, pp. 254-264, 1976.

[13] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian Restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 6, pp. 721-741, 1984.

[14] J. N. Gupta and P. A. Wintz, "A boundary finding algorithm and its applications," *IEEE Trans. Circuits Syst.*, vol. CAS-22, no. 4, pp. 351-362, 1975.

[15] R. M. Haralick, "Edge and region analysis for digital image data," *Comput. Graphics Image Processing*, vol. 12, pp. 60-73, 1980.

[16] S. L. Horowitz and T. Pavlidis, "Picture segmentation by a tree traversal algorithm," *J. ACM*, vol. 23, pp. 368-388, 1976.

[17] R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Trans. Geosci. Electron.*, vol. GE-14, no. 1, pp. 19-26, 1976.

[18] R. Kohler, "A segmentation system based on thresholding," *Comput. Graphics Image Processing*, vol. 15, pp. 319-338, 1981.

[19] M. D. Levine and A. M. Nazif, "Dynamic measurement of computer generated image segmentations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 2, p. 155, 1985.

[20] B. W. Lindgren, *Statistical Theory*, 3rd ed. New York: Macmillan, 1976.

[21] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *Proc. Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, 1985, pp. 22-26.

[22] M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*. New York: Plenum, 1980.

[23] K. A. Narayanan, D. P. O'Leary, and A. Rosenfeld, "Image smoothing and segmentation by cost minimization," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-12, no. 1, pp. 91-96, 1982.

[24] C. C. Parma, A. R. Hanson, and E. M. Riseman, "Experiments in schema driven interpretation of a natural scene," "Traitement et analyse numerique des images," OTAN, Institut d'Etudes Avancees, 1980.

[25] T. Pavlidis, "Segmentation of pictures and maps through functional approximation," *Comput. Graphics Image Processing*, vol. 1, no. 4, pp. 360-372, 1972.

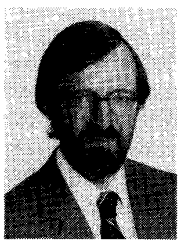
[26] S. Peleg, "Classification by discrete optimization," *Comput. Vision, Graphics, Image Processing*, vol. 25, no. 1, pp. 122-130, 1984.

[27] T. Poggio, "Early vision: From computational structure to algorithms and parallel hardware," *Comput. Vision, Graphics, Image Processing*, vol. 31, pp. 139-155, 1985.

[28] J. C. Simon, "Some current topics in clustering in relation with pattern recognition," in *Proc. 4th Int. Joint Conf. Pattern Recognition*, 1978, pp. 19-29.

[29] P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*. San Francisco, CA: Freeman, 1973.

[30] S. L. Tanimoto, "Regular hierarchical image and processing structures in machine vision," in *Computer Vision System.* New York: Academic, 1978, pp. 165–174.
[31] S. Tanimoto and A. Klinger, *Structured Computer Vision.* New York: Academic, 1980.
[32] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 413–424, 1986.
[33] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems.* Washington, DC:, Winston, 1977.
[34] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles.* Reading, MA: Addison-Wesley, 1974.
[35] J. H. Ward, "Hierarchical grouping to optimize an objective function," *J. Amer. Stat. Ass.*, vol. 58, pp. 236–245, 1963.
[36] S. W. Zucker, "Region growing: Childhood and adolescence," *Comput. Graphics Image Processing*, vol. 5, pp. 382–399, 1976.



Jean-Marie Beaulieu (S'83–M'84) received the Bachelor's and Master's degrees in electrical engineering from Sherbrooke University, Sherbrooke, Quebec, Canada, in 1974 and 1976, respectively, and the Ph.D. degree in electrical engineering from Ottawa University, Ottawa, Ontario, Canada, in 1985.

He is currently an Assistant Professor in the Department of Computer Science at Laval University, Quebec, Quebec, Canada. From 1985 to 1987, he was working at the National Research Council of Canada for the Public Safety Project Office. His areas of research interest include image processing and analysis, pattern recognition and artificial intelligence, remote sensing, software development, and user interfaces.

Morris Goldberg (S'68–M'70) received his undergraduate training at the University of McGill and the Ph.D. degree in electrical engineering from Imperial College in 1972.

Since 1974 he has been working in the general area of image processing, first for the Canada Centre for Remote Sensing and since 1976 at the University of Ottawa, where he is at present an Associate Professor in the Department of Electrical Engineering. In 1982–1983 he was a Visiting Professor at the ENST in Paris, and in 1986–1987 a consultant for Bell Northern Research. His current research interests are in the areas of multimedia communications for medical applications, expert systems for remote sensing, and image compression.

Dr. Goldberg is Treasurer of the Canadian Image Processing and Pattern Recognition Society, and Chairman of the IAPR Technical Committee on Remote Sensing.