SELECTION OF SEGMENT SIMILARITY MEASURES FOR HIERARCHICAL PICTURE SEGMENTATION *

Jean-Marie Beaulieu+ and Morris Goldberg++

RÉSUMÉ

Différentes mesures de la similarité entre segments sont étudiées pour la segmentation de l'image. Une segmentation agglomérative hiérarchique compare et fusionne deux segments s'ils sont similaires. L'algorithme utilisé fusionne, à chaque itération, les deux segments les plus similaires. Regardant la segmentation de l'image comme un problème d'approximation, la mesure de similarité est reliée à l'erreur totale d'approximation; l'accroissement de l'erreur produit par la fusion de deux segments est utilisé. Une image de télédétection est segmentée par l'approximation de chaque région soit par une constante ou soit par un plan. Une mesure de similarité sensible à la variance locale de l'image est aussi présentée. Les avantages de la combinaison de différentes mesures sont aussi soulignés. Différentes régions d'une image peuvent nécessiter différentes mesures qui doivent donc être combinées pour obtenir, dans l'ensemble, un bon résultat. De plus, dans une segmentation hiérarchique, une mesure simple peut être utilisée lors des premières fusions, tandis que des mesures plus complexes vont être employées par la suite.

ABSTRACT

The problem of defining appropriate segment similarity measures for picture segmentation is examined. In agglomerative hierarchical segmentation, two segments are compared and merged if found similar. The proposed Hierarchical Step-Wise Optimization (HSWO) algorithm finds and then merges the two most similar segments, on a step-by-step basis. By considering picture segmentation as a piece-wise picture approximation problem, the similarity measure (or the step-wise criterion) is related to the overall approximation error. The measure then corresponds to the increase of the approximation error resulting from merging two segments. Similarity measures derived from constant approximations (zeroth order polynomials) and planar approximations (first order polynomials) are applied to a Landsat picture, and the results are presented. An adaptive measurebased upon local variance is also used. The advantages of combining similarity measures (or cirteria) are also stressed. Different picture areas can require different measures which must therefore be combined in order to obtain good overall results. Moreover, in hierarchical segmentation, simple measures can be used for the first merging steps, while, at a higher level of the segment hierarchy, more complex measures can be employed.

\+ National Research Council of Canada, NAE, Ottawa, Ontario, Canada, K1A 0R8.

++ Electrical Engineering Dep., University of Ottawa, Ottawa, Ont., Canada, K1N 6N5.

* NRCC No. 24388.

KEYWORDS: Hierarchical segmentation, segmentation, similarity measures, clustering.

I - INTRODUCTION

A hierarchy of segments can be represented by a segment tree in which nodes correspond to segments. Each segment S_i^k is linked to segments of the lower level, S_j^{k-1}, which are disjoint sub-sets of S_i^k, and which are called "sons" of S_i^k. A picture partition thus corresponds to a sub-set of these tree nodes. Starting from the bottom of the tree, an agglomerative hierarchical segmentation algorithm climbs up the tree by merging similar segments. Different similarity measures can be used to decide if two adjacent segments must be merged.

Brice and Fennema [3] use two heuristics, based upon information from the segment boundaries, to evaluate the similarity of two segments: the phagocyte and the weakness heuristics. The phagocyte heuristic guides the merging of regions in such a way as to smooth or shorten the resulting boundary. Two regions are merged if their common boundary is weak and if the segment boundary length does not increase too quickly. The weakness heuristic merges two regions if a prescribed portion of their common boundary is weak. The phagocyte heuristic is applied first, followed by the weakness one.

Freuder [5] presents an algorithm where the similarity of two segments is a function of the surrounding segments. For each segment, S_i, the adjacent segment, S_j, which is the most similar to S_i is selected. A directed link is drawn from S_i to S_j. The similarity is related to the difference between segment means values and the segment sizes. Thus, each segment points to one of its neighbours, the one with the closest mean value (weighted to take account of segment sizes). All segments related by a "double link" are then merged. A double link indicates a local minimum of the segment similarity measure, with S_j being minimum among the neighbours of S_i, and S_i among the neighbours of S_j.

Horowitz and Pavlidis [6] propose a split-and-merge approach using a pyramidal data structure. The data structure defines the way in which segments can be merged or split. A pyramid is a stack of regular picture blocks of decreasing sizes. The picture blocks (or segments) of one level are split into four regular sub-parts to form the next lower level. A pyramid can be regarded as a segment tree where each node corresponds to a block of $2^k \times 2^k$ pixels. A segment is considered as homogeneous if the segment approximation error is smaller than a predefined threshold. The algorithm consists of 1) merging the homogeneous segments, if the resulting segments are also homogeneous, or 2) splitting the segments that are not homogeneous into their four sub-parts.

Chen and Pavlidis [4] employ a statistical decision process in the preceding split-and-merge approach. The segments of the initial partition are first tested for uniformity, and if not uniform, they are divided into smaller segments. The uniform segments are then subjected to a cluster analysis to identify similar types which are then merged.

The Hierarchical Step-Wise Optimization (HSWO) algorithm is first presented in the next section. Different segment similarity measures derived from a picture approximation model are defined and employed for the segmentation of a remote sensing picture. In section IV, a measure that takes into account the local variance is examined. The advantages of combining similarity measures are studied in section V. The different proposed measures are used to segment a remote sensing picture, and experimental results are presented.

II - THE HIERARCHICAL STEP-WISE OPTIMIZATION ALGORITHM

A hierarchical segmentation algorithm based upon step-wise optimization is used in this paper [1], [2]. A segment similarity measure, $C_{i,j}$, is defined as the step-wise criterion to optimize. At each iteration, the algorithm employs an optimization process to find the two most similar segments, which are then merged.

The Hierarchical Step-Wise Optimization (HSWO) algorithm can be defined as follows:

i) Define an initial picture partition.
ii) For each adjacent segment pair, (S_i, S_j), calculate the step-wise criterion, $C_{i,j}$; then find and merge the segments with the minimum criterion value.
iii) Stop, if no more merges are needed; otherwise, go to ii).

Different segment similarity measures (step-wise criteria) can be employed, each one corresponding to different definitions of the picture segmentation task. Different measures are examined in the following sections.

III - PICTURE APPROXIMATION

Let $f_i(x,y)$ designate the pixel values for the segment S_i, ($f_i(x,y)=f(x,y)$ for $(x,y) \in S_i$). Piece-wise picture approximation, therefore, consists in approximating each segment by a polynomial function, $r_i(x,y)$. The approximation error for each segment is defined as the sum of the squared deviations:

$$H(S_i) = \sum_{(x,y)\in S_i} [f_i(x,y) - r_i(x,y)]^2$$

The goal of picture approximation is then to find the partition, $\{S_i\}$, that minimizes the overall approximation error, $\sum H(S_i)$.

The segment similarity measure, thus, can be related to the increase of the approximation error produced by the merging of two segments, S_i and S_j:

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

The utilization of $C_{i,j}$ in the HSWO algorithm ensures that each iteration does it best to minimize the overall approximation error. Segmentation results produced by constant value and planar approximation are now examined.

CONSTANT VALUE APPROXIMATION

A portion of a Landsat picture (64x64 pixels) is presented in Figure 1, together with an enlargement of the lower left area. The first model assumes that constant value regions constitues a representation of this picture. The segment approximation function is therefore

$$r_i(x,y) = u_i$$

where u_i is the segment mean value. This approximation function is employed for the calculation of $C_{i,j}$. The segmentation results for the constant approximation are shown in Figure 2. The picture is divided into 100 segments, and an approximation picture is produced by replacing each segment by its mean value.

PLANAR APPROXIMATION

The constant value approximation is seen to be generally appropriate but its limitations are clear. For example, in Figure 3, a 1-dimensional case is shown where a constant value region is appropriate for regions 1 and 3, while it is inappropriate for region 2. A planar approximation is more suitable:

$$r_i(x,y) = a_i + b_i\,(x) + c_i\,(y)$$

This approximation function is now used in the calculation of $C_{i,j}$ for the segmention of the Landsat picture of Figure 1 and the results are shown in Figure 4. Region 4 is an example where a planar approximation is needed.

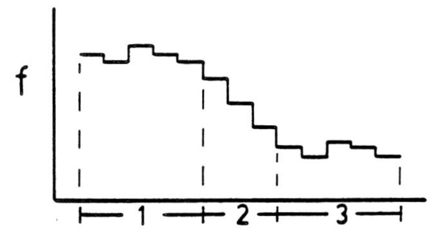


Figure 3 : An example with constant value and inclined line regions.

The utilization of a higher degree polynomial however can cause spurious results to arise, as illustrated in Figure 5. The original signal (a) is composed of two constant value regions which are then corrupted by noise (b). In this case, the true regions would be correctly detected by a constant approximation, (c), while, a first order approximation leads to misleading results, (d). Higher degree polynomials, having more degrees of freedom, attempt to match more closely the noise deformation, thus yielding spurious results. A similar situation is observed by comparing regions 1, 2 and 3 of Figure 2-c, and regions 5, 6 and 7 of Figure 4-c. Region 2 resulting from a constant approximation, is divided between regions 5 and 6 in the case of planar approximation. The inclined plane of region 6 represents both the light values of region 2 and the darker values of region 3. The pixels of region 7 do not fit this plane and therefore form a distinct region. The constant approximation results are, therefore, more appropriate for these regions.

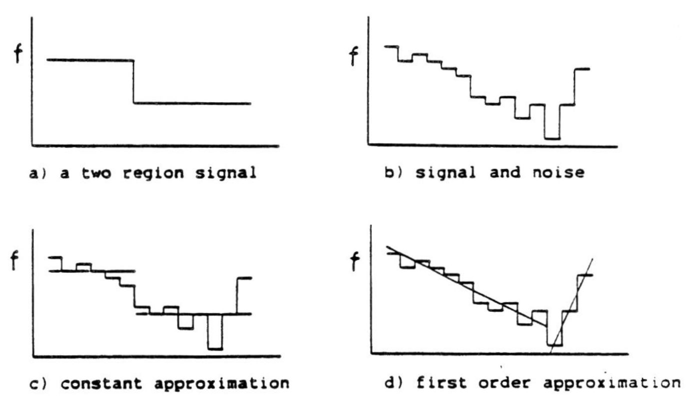


Figure 5 : A 1-D function composed of two regions (a), with added noise (b), and approximated by constant values (c) and inclined lines (d).

IV - LOCAL VARIANCE

The HSWO algorithms based upon constant or planar approximation attempts to minimize the approximation error. The evaluation of the error for a given pixel does not consider the importance of the gray level variance in the surrounding area. Thus, in Figure 6, both examples, a and b, have the same criterion value with respect to the regions 1 and 2. It can be advantageous to make the criterion value depend upon the segment variance, and define a new criterion such as:

$$C^*_{i,j} = C_{i,j} \; / \; (1 + \sigma_{i,j})$$

where $\sigma^2_{i,j}$ is the mean value of the squared approximation error :

$$\sigma^2_{i,j} = (H(S_i) + H(S_j)) \; / \; (N_i + N_j)$$

Here, $H(S_i)$ is, as previously defined, the sum of the squared approximation error for segment S_i, and N_i is its size. For constant approximation, $\sigma^2_{i,j}$ corresponds to the combined variance of both segments. Thus $C^*_{i,j}$ is equal to $C_{i,j}$ when $\sigma_{i,j}$ is zero, and decreases for large values of the variance. The results given by this new criterion are shown in Figure 7. The regions marked by "X" correspond to a zone of large gray level variation, and thus the utilization of $C^*_{i,j}$ produces smaller criterion values and forces more segment merging in this area. This new criterion seems preferable, as it adjusts itself to local picture variations.

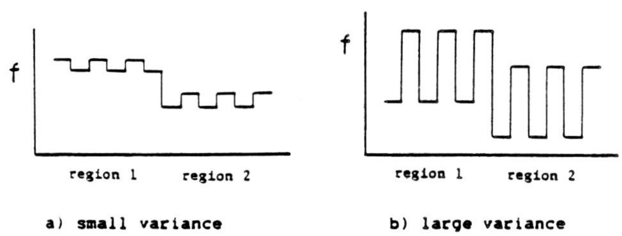


Figure 6 : Examples of regions with the same criterion values but different variances.

V - SIMILARITY MEASURE COMBINATION

The step-wise optimization algorithm can employ different similarity measures, which correspond to different segment description models. Those previously introduced involve very simple models. However, more complex models can be required for segmentation tasks. Complex measures can be obtained from combinations of simpler ones.

Zobrist and Thompson [8] point out that human vision employs many cues such as brightness, contour, color, texture and stereopsis to perform perceptual grouping. They stress the limitations of using only one cue at a time for computer grouping, and show the importance of studying mechanisms that combine many cues. For computer simulation of human perception, they derive from each cue a distance function that measures the similarity of two scene parts. Then, they perform a weighted sum of these distances to obtain a global perceptual distance.

Applying this approach to picture segmentation, it can be noted that different picture areas can require different segment models (cues) and that these models must be combined in order to obtain good overall results. Hence, the constant approximation can be appropriate for some parts of a picture while the planar approximation can be preferable for some other parts. Thus, it can be advantageous to combine the similarity measures associated with both models. For example, a composite measure can be obtained as follows:

$$C_{(composite)} = C^*_{(constant)} * C^*_{(planar)}$$

This corresponds to using the geometric mean of the two measures to form the composite one. $C^*_{(.)}$ indicates a local variance adaptive measure as defined in the preceding section.

In picture segmentation, an ordering of segment descriptions can also be considered [7]. For example, the pixel gray level can be employed to form small homogeneous regions, then more complex descriptors, such as segment contour shape, can be considered for forming larger regions. Many segment descriptors, such as contour shape, or higher order approximation coefficients, are meaningless for small regions and only become useful at a latter stage. In the hierarchical segmentation scheme, this corresponds to using a simple measure for the first merging steps, then, as we get to a higher level in the segment hierarchy, more complex measures, involving more complex segment descriptors, are introduced.

The ordering of segment descriptions and composite measures are now employed to segment the Landsat picture. The constant approximation measure, $C_{i,j}$, is first used to obtain a partition with 1000 segments. Then the previously defined composite measure is employed to continue the segment merging. The results which combine the characteristics of the preceding measures are shown in Figure 8. For example, in Figure 8-c, region 8 is represented by an inclined plane as is shown in Figure 4-c for the planar approximation. While regions 9, 10 and 11 correspond to those obtained by constant value approximation in Figure 2. Thus, the advantages of planar approximation are exploited, while the previously noted artefacts are avoided. The constant value approximation is still predominant for large constant areas.

REFERENCES

[1] J.M. Beaulieu and M. Goldberg, "Hierarchical Picture Segmentation by Approximation," Canadian Communications and Energy Conference, Montreal (Canada), 1982, 393-396.

[2] J.M. Beaulieu and M. Goldberg, "Step-Wise Optimization for Hierarchical Picture Segmentation," Conf. on Computer Vision and Pattern Recognition, Washington, D.C., 1983, 59-64.

[3] C. Brice and C. Fennema, "Scene Analysis Using Regions," Artificial Intelligence, vol. 1, 1970, 205-226.

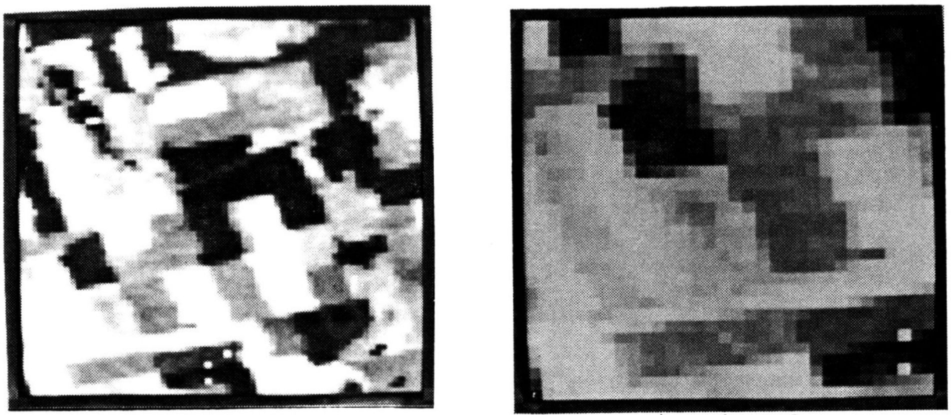
[4] P.C. Chen and T. Pavlidis, "Image Segmentation as an Estimation Problem," Comp. Grap. & Image Proc., Vol. 12, 1980, 153-172.

[5] E.C. Freuder, "Affinity: A Relative Approach to Region Finding," Comp. Graph. & Image Proc., vol. 5, 1976, 254-264.

[6] S.L. Horowitz and T. Pavlidis, "Picture Segmentation by a Tree Traversal Algorithm," Journal of the Ass. for Computing Machinery, vol. 23, 1976, 368-388.

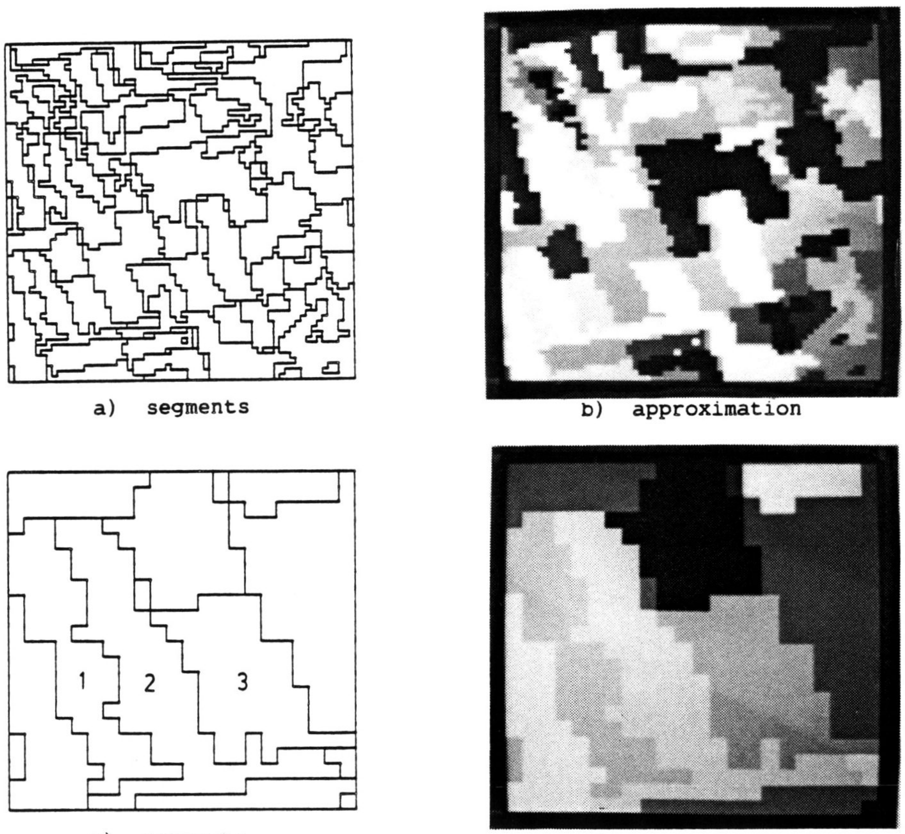
[7] T. Pavlidis, "Hierarchies in Structural Pattern Recognition," Proceeding of IEEE, vol. 67 (5), 1979, 737-744.

[8] A.L. Zobrist and W.B. Thompson, "Building a Distance Function for Gestalt Grouping," IEEE Trans. on Computer, Vol. C-4 (7), 1975, 718-728.



a) the entire picture　　　b) sub-area

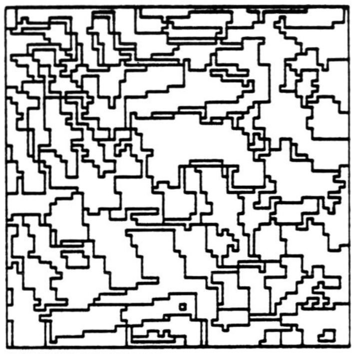
Figure 1 : Landsat picture



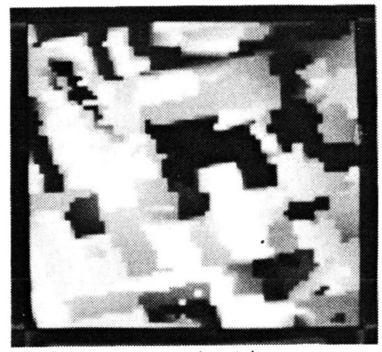
a) segments　　　b) approximation

c) segments　　　d) approximation

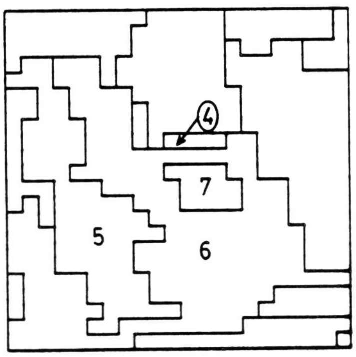
Figure 2 : Segmentation results for constant approximation.



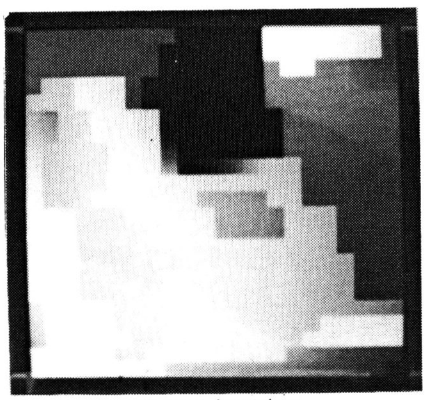
a) segments



b) approximation

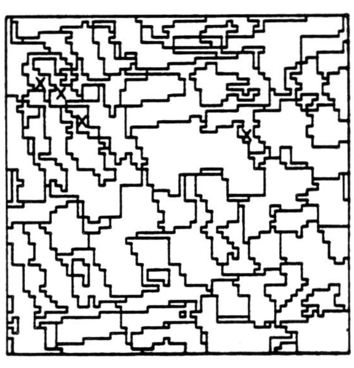


c) segments

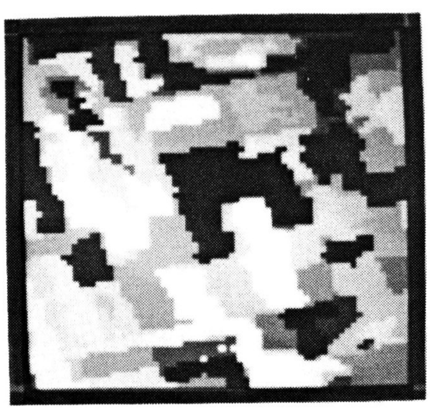


d) approximation

Figure 4 : Segmentation results for planar approximation.



a) segments



b) approximation

Figure 7 : Segmentation results for the local variance adaptable approximation.

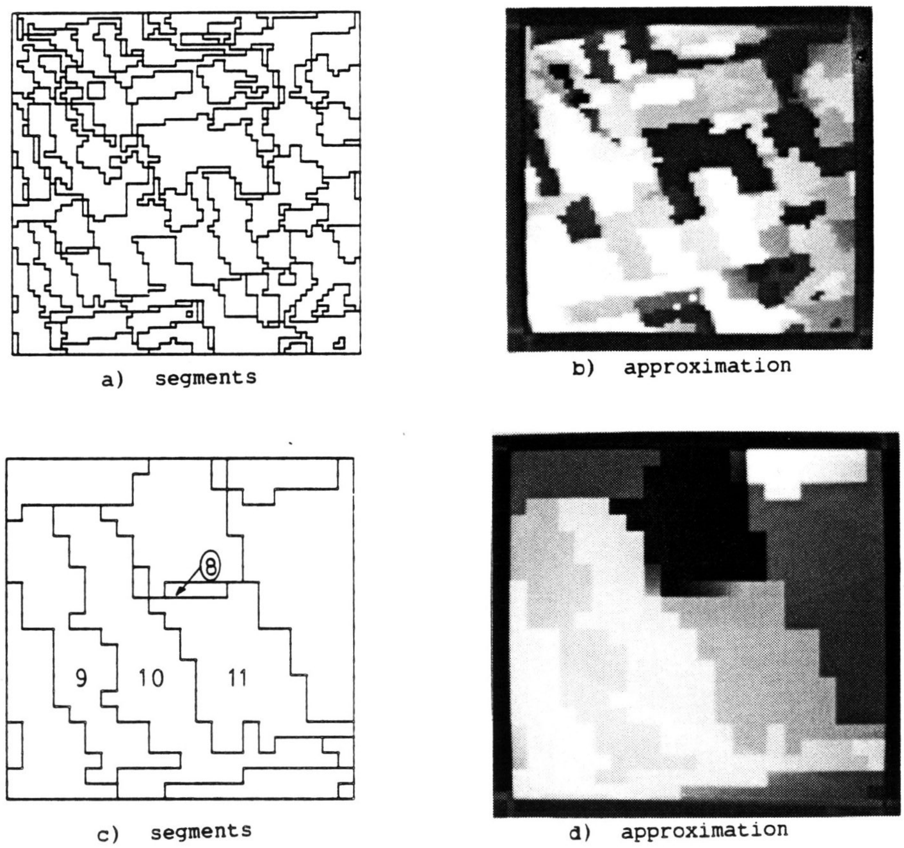


Figure 8 : Segmentation results from criterion combination.