Image Segmentation from Consensus Information

Kyujin Cho¹ and Peter Meer

Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08855

Received September 25, 1995; accepted August 19, 1996

A new approach toward image segmentation is proposed. A set of slightly different segmentations is derived from the same input and the final result is based on the consensus among them. The perturbations are introduced by exploiting the probabilistic component of a region adjacency graph (RAG) pyramid-based segmentation. From the set of initial segmentations the cooccurrence probability field is obtained in which global information about the delineated regions becomes locally available. The final segmentation is based on this field and is obtained with the same hierarchical, RAG pyramid technique. No user set parameters or context-dependent thresholds are required. © 1997 Academic Press

1. INTRODUCTION

In image segmentation, given a homogeneity criterion, the image must be partitioned into regions within which the criterion is satisfied. The border between two regions corresponds to a discontinuity, i.e., to an edge. However, not all edges are meaningful. Local variations due to noise often can yield significant discontinuities. A tradeoff exists between oversegmentation, partition into too many regions, and undersegmentation, in which case larger regions are obtained at the expense of possible erroneous fusions. For reviews of different image segmentation techniques see [17], Section 10, and [32].

A segmentation based solely on a simple homogeneity criterion, like constant gray levels, cannot provide the decomposition of an image into regions which correspond to parts of an object in the physical world. To obtain a meaningful segmentation the higher level descriptions of the objects and often also of the relations among them must be taken into account. Image segmentations using simple homogeneity criteria can only provide the partition of the image into "puzzle pieces" from which the objects must be assembled. A complete segmentation system is complex and makes use of many heuristics [4, 30]. To reduce oversegmentation, in the absence of context dependent information, probabilistic models are required to guide the fusion process [21].

The difficulty of segmentation is an aspect of the local/ global duality problem. A region is declared homogeneous by analyzing small local neighborhoods. The larger these neighborhoods, the more reliable are the extracted spatial statistics *given* that the data in the neighborhood is indeed homogeneous. On the other hand, using a larger neighborhood increases the chances of analyzing nonhomogeneous data under the assumption of homogeneity.

To avoid the problem of local/global duality often edge information (local) and homogeneity information (global) are combined during image segmentation. Haddon and Boyce [16] refined a cooccurrence matrix-based segmentation by incorporating boundary information into a relaxation procedure. Wu and Leahy [41] used operations on a weighted region adjacency graph to delineate the boundaries of homogeneous regions. Chu and Aggarwal [12] combined several segmentations of different modalities, intensity, range, thermal, etc., by exploiting both region and edge information. LeMoigne and Tilton [26] used an edgemap to find the optimal stopping condition in a hierarchical segmentation algorithm. The edgemap, however, was defined by the user and contained only the features of interest. Geiger and Yuille [15] used mean field theory to unify several image segmentation techniques in which discontinuities are explicitly taken into account: nonlinear diffusion and minimum description length. Knapman and Dickson [20] have found that edge information is more important for segmentation than region statistics, at least in a Bayesian framework. The influence of the employed edge detection technique was investigated in [22].

In most segmentation methods the edge (discontinuity) information is based on local analysis and its errors affect the fusion of adjacent homogeneous regions. We propose a different technique for image segmentation. Instead of statistics characterizing the spatial structure of the local neighborhood, for every pair of adjacent pixels their *ensemble statistics* under the segmentation task is used for determining local homogeneity.

¹ Current address: Open Solution Center, Samsung Data System, 219-1 Migun-Dong, Seodaemun-Gu, Seoul, Korea.



FIG. 1. The information flow in image segmentation through the consensus paradigm.

In Figure 1 the information flow of the proposed paradigm is shown. Several initial segmentations are derived from the same input image by exploiting the probabilistic component of the hierarchical region adjancency graph (RAG) pyramid-based technique described in [27]. (The RAG pyramid is discussed in the beginning of Section 2.) From the ensemble of initial segmentations, for every adjacent pixel pair a cooccurrence probability, i.e., the probability of belonging to the same delineated region, is derived. The set of cooccurrence probability is different from the probabilities derived from cooccurrence matrices [8, 16]. The former is computed after the initial segmentations, while the latter are obtained from the gray-level differences of the input image. Since the cooccurrence probabilities are derived from the initial image segmentations they capture global information at the local (pixel pair) level. The final segmentation of the input image is obtained by processing the cooccurrence probability field with the same RAG pyramid technique. Only the pixel pairs with high cooccurrence probability are grouped together, i.e., the final segmentation is based on the consensus about local homogeneity. The technique automatically associates a confidence measure to each delineated region.

Hierarchical methods were often employed for image segmentation (see [17], Sections 10.5 and 10.6, for a review). Traditional hierarchical methods, the split-and-merge techniques, are based on a rigidly structured hierarchy, most often square neighborhoods. This yields strong artifacts in segmentation: the region borders have ragged appearance; long, thin, waving objects are difficult to delineate. To improve the quality of segmentation, the structure of the hierarchy is often dynamically restructured [1, 37]. Pavlidis and Liow [34] used edge information to refine a rigid, quadtree-based segmentation. Ng *et al.* [31] employed the coarse segmentation (higher levels of the image pyramid) to guide boundary refinement by local operator at the lower levels.

In order to eliminate the influence of rigid hierarchies, the underlying structure should be based on the *region adjacency graph* of the image. Most of the proposed techniques employing region adjacency graphs, however, are not hierarchical and at each contraction of the RAG only one fusion is allowed through exhaustive search over all the possibilities [28, 40]. User-specified parameters are also often required [40, 41]. Nacken [29] proposed a pyramid relinking process guided by RAG for hierarchical image segmentation.

Our computational tool, the RAG pyramid automatically adapts its structure to the image while still maintaining the $O[log(image_size)]$ processing time property of the hierarchical techniques. The artifacts of image pyramids [5] are eliminated. Construction of the RAG pyramid has a probabilistic component based on which several slightly different outputs can be obtained from the same input (see Fig. 1).

In Section 2, the RAG pyramid is reviewed and the method of generating an initial segmentation of the input image is described. In Section 3, several initial segmentations are combined to obtain the cooccurrence probability field and the method of extracting the final delineation is presented together with experimental results. In Section 4 a faster, but less accurate method for generating the cooccurrence probability field is proposed. In Section 5 an example of top-down processing using the delineated homogeneous regions as input is shown, and the implications of the employed paradigm are discussed.

2. IMAGE SEGMENTATION WITH RAG PYRAMID

In this section the technique of segmentation with the RAG pyramid is reviewed. For more details see [27]. The RAG pyramid is built by recursive graph contractions. The RAG of the input image is given by the 8-connected graph of the underlying mesh, i.e., every pixel is taken as a homogeneous region. In the reduced resolution representations every vertex corresponds to a compact region in the input image, its *support*. To generate the next level of the hierarchy only a subset of the vertices is retained. The spatial relations among the retained vertices, *survivors*, and the *nonsurvivors* must satisfy two properties in order for the graph contraction to be optimal:

• No two survivor vertices should be neighbors.

• Any nonsurvivor vertex should have a survivor neighbor.

These two conditions are equivalent to the vertices retained for the RAG of the next level being a maximal independent set [11] of the RAG at the current level.

The vertices are selected with a parallel, probabilistic symmetric breaking algorithm [23]. Every vertex in the graph is allocated a random number drawn from the [0, 1] uniform distribution. A vertex becomes a survivor if its outcome is a local maximum. Its neighbors are removed from subsequent iterations. After less than five iterations the algorithm converges and the maximal independent set of the graph is extracted. The adjacency relations for the reduced resolution representation of the next level, i.e., the edges of that level's RAG, can now be obtained by using the paths (of length at most three) between the survivors of the current level's RAG.

The RAG-based representations can only be useful for segmentation if the supports of the vertices correspond at any level of the hierarchy to homogeneous patches in the input image. To satisfy this condition, the RAG must be decomposed into *similarity subgraphs* before graph contraction.

The similarity subgraphs are derived from the RAG by local decisions. Let g(v) be the value associated with vertex v of an RAG and $g(v_i)$, $i = 1, ..., n_v$, the values associated with its neighbors. In the sequel, we will use gray-level constancy as homogeneity criterion, i.e., the image is to be segmented according to the constant facet model. In this case g(v) is an approximation of the average gray level of the support. The technique, however, remains the same for more complex homogeneity criteria. A local threshold, LT(v), is defined for vertex v from the ordered sequence of differences

$$\delta_i(v) = |g(v_i) - g(v)| \quad i = 1, \dots, n_v.$$
(1)

The most significant jump in the sequence yields LT(v). The jumps are estimated from the left and right averages at every location in the ordered sequence of differences. Thus the local threshold is established by a cumulative sum type, jump detection algorithm [3]. Since using only local decisions can yield undersegmentation, a global threshold *GT*, defining the maximum allowable tolerance for the homogeneity measure, must also be provided. (The issue of global threshold selection is discussed in Section 2.1.)

The connectivity threshold CT(v) based on which the vertex v selects the neighbors in the same homogeneity class is then

$$CT(v) = \min[LT(v), GT].$$
 (2)

The edges in the RAG connecting two neighbors whose $\delta_i(v) > CT(v)$ are temporarily removed. Each of the resulting connected subgraph is a similarity subgraph which corresponds to a homogeneous region in the input. The similarity subgraphs are contracted independently by the probabilistic procedure described above. The next level's RAG is then constructed taking into account the previously removed edges. The recursive contraction of the RAG continues till no connected similarity subgraph can be derived, i.e., all the vertices in the RAG are *roots* of homogeneous regions in the input image. The RAG pyra-



FIG. 2. An example of building the next level of the RAG pyramid. (a) The RAG of level l. (b) Similarity subgraphs, with the survivor vertices filled in. (c) Allocation of nonsurvivors to the most similar survivor. (d) The RAG of level l + 1.

mid is constructed in $O[log(image_size)]$ steps and can be implemented on parallel hardware [43].

In Fig. 2 an example of the contraction of RAG is shown. Figure 2a shows the RAG at level *l*. In the RAG each vertex v represents a currently delineated homogeneous region (solid line boundaries), while the edges (dotted lines) connect spatially adjacent regions. Let the decomposition of RAG into similarity subgraphs be as shown in Fig. 2b with the arrows. Note that the similarity subgraphs are directional subgraphs. Assume that the random numbers associated with each vertex are as shown in Fig. 2b. Then the vertices retained for the next level of hierarchy (survivors) are those marked as black. The nonsurvivors are allocated to the most similar surviving neighbor in the similarity subgraph as shown in Fig. 2c. The next level's RAG is completed by connecting two survivors with an edge if there exists in the current RAG a path of maximum length three between them (Fig. 2d). Theoretical issues of the RAG hierarchies were investigated in [38, 39].

2.1. Automatic Selection of the Global Threshold

The global threshold was a user set parameter in [27], and its value had a significant influence on the output. In the absence of any a priori information the range of possible global threshold values is large. However, a reasonable good estimate can be obtained from the input. Let for the moment GT = 256. Since this value exceeds the largest possible gray level difference (1), the connectivity threshold is solely given by the local threshold (2). The first level of the RAG pyramid is then generated from the input image. In the support of a surviving vertex v the largest gray level difference is

$$\varepsilon_m(v) = \max_{i,j} |g(v_i) - g(v_j)| \le 2 * LT(v), \qquad (3)$$

where v_i , v_j are any two vertices (pixels) belonging to this support.



FIG. 3. Segmentation example. (a) Input, aerial image. (b) Image segmentation with the RAG pyramid.

The value of $\varepsilon_m(v)$ for all the vertices of the first level's RAG are sorted, and the global threshold is defined as the point of significant increase in the list. This is defined (for $\varepsilon_m(v) \neq 0$) as the smaller value between the one corresponding to the "corner" of the graph and the one corresponding to the 80th percentile. Note that for a noisy image, the value of global threshold is automatically increased to compensate for the noise induced local variations. The graph of sorted $\varepsilon_m(v)$ for the *aerial* image (Fig. 3a) is shown in Fig. 4, and the extracted global threshold was 22. The shape of the graph is typical for several classes of images. The 80th percentile condition is used only to avoid expensive (and often unreliable) searches for the corner of the graph.

The RAG pyramid segmentation of the *aerial* image is shown in Fig. 3b. Since only the bottom-up information flow is used in building the hierarchy, the output of the RAG pyramid is an oversegmented version of the input. The number of regions tessellating the input can be reduced by an order of magnitude through postprocessing (to be discussed in Section 3.3).

3. CONSENSUS SEGMENTATION

A simple homogeneity criterion, like constant gray levels, cannot account for a complex image structure. It is therefore more secure to oversegment the image and pro-



FIG. 4. Aerial image. The sorted maximum local differences for the supports of level 1 of the RAG pyramid.



FIG. 5. The influence of the probabilistic component on the structure of RAG pyramid. (a) The RAG of level l, outcomes of random variables, similarity subgraphs, survivors (filled in). (b) The RAG of level l + 1. Compare with Fig. 2d.

vide a measure of confidence for each delineated region. The confidences can be assessed through the consensus technique (Fig. 1). The construction of the RAG pyramid has a probabilistic component, selection of the vertices for the next level's region adjacency graph. The structure of the hierarchy thus may differ locally based on the outcome of the random variables. Changes in the set of nonsurvivors allocated to a survivor changes the value g(v) associated with the survivor. This in turn results in slightly different similarity subgraphs at subsequent levels of the hierarchy. The cumulative effect of the changes can be significant at blurred transitions between regions where the structure of the similarity subgraph is less robust. At a steep discontinuity, on the other hand, the same similarity subgraph is derived in spite of small variations in the values associated with the vertices.

Given the input image, therefore, many slightly different region delineations can be obtained by repeating the RAG pyramid segmentation process. An example of the variation in the structure of hierarchy caused by the probabilistic component is shown in Fig. 5. Different random numbers were assigned to the vertices of the RAG in Fig. 2a, and a different set of survivors (black nodes) was obtained. Notice that the similarity subgraphs in Fig. 5a are the same as in Fig. 2b. The difference in the RAG of level l + 1 has an influence on the output (segmented image) only if these regions are not fused into one at higher levels.

The individual segmentations obtained with the RAG pyramids will be called *initial* segmentations. The result shown in Fig. 3b is such a segmentation. Let N slightly different initial segmentations be derived from the input image. The differences are mostly in the small regions and along the blurred edges. These segmented images are registered pixelwise, i.e., on the 8-connected mesh of the input. Thus N values are associated with every pixel. Assume that the pixel v and its neighbor v_i belong to the

same region N_{v,v_i} times. The empirical *cooccurrence probability* of this pixel pair is then

$$p(v, v_i) = \frac{N_{v, v_i}}{N}$$
 $i = 0, \dots, 7.$ (4)

The set of all cooccurrence probabilities defines the cooccurrence probability field of the input image analyzed under the given homogeneity criterion. Pixels on the two sides of a strong edge will have low cooccurrence probabilities since their gray-level difference is large enough to always exceed the connectivity threshold. However, low probabilities are also obtained for pixel pairs in regions whose delineation changes significantly when the structure of the RAG pyramid varies (e.g., blurred edges). In Fig. 6b, the cooccurrence probability field associated with the gray-level image in Fig. 6a is shown. The gray level value of a pixel is given inside the corresponding node. Note that the cooccurrence probabilities being based on the outputs of RAG pyramids (segmented images) incorporate information about pixels not shown in Fig. 6a.

In the cooccurrence probability field global information becomes locally accessible. The segmented region to which a pixel belongs (and which can be very large) is a global feature extracted from the input image. The cooccurrence probabilities associated with this pixel are based on the initial segmentations. Thus, in spite of being computed locally, these probabilities carry information about the relation between the pixel and other pixels which were grouped into the same region. This information is global. For example, when the employed homogeneity criterion is not adequate for the analysis of the pixel's neighborhood, the delineated boundary varies strongly with the structure of the RAG pyramid and low cooccurrence probabilities result.



FIG. 6. An example of cooccurrence probability field obtained from 20 RAG pyramids. (a) Part of a gray-level image. (b) Cooccurrence probabilities.

Visualization of the probability field involving an eightdimensional vector for every pixel is difficult. A coarse measure is obtained by defining the scalar quantity

$$c(v) = \frac{255}{8} \cdot \sum_{i=0}^{7} p(v, v_i),$$
(5)

with values between 0 and 255. The gray-level image of c(v) is the *consensus image* in which a lighter value indicates a better change for a pixel to belong to a homogeneous region. That is, there exists a consensus in the set of initial segmentations about the assignment of the pixel. In Fig. 7 two consensus images obtained with different number of initial segmentations of the *aerial* image are shown. Strong edges (dark pixels) are immediately extracted (Fig. 7a). The assessment of the homogeneity of regions is refined as more initial segmentations are considered. The

consensus image is a low-dimensional projection of the cooccurrence probability field and thus it discards the directionality of the latter. We have found that N = 20 suffices for a cooccurrence probability field to provide a reliable final segmentation.

3.1. Weighted Region Adjacency Graph

The technique developed for building the RAG pyramid (and providing the initial segmentations) can be used to extract the high confidence homogeneous regions from the cooccurrence probability field. Each edge of the base level's RAG graph (the 8-connected mesh) has now a cooccurrence probability associated with it. The new graph is called the *weighted region adjacency graph*. To perform the graph contractions an additional procedure for computing the weights at the subsequent levels of the hierarchy is required.



FIG. 7. Consensus images for the *aerial* image. Number of initial segmentations N combined together: (a) N = 5. (b) N = 20.



FIG. 8. An example of updating the weighted region adjacency graph. See text.

Let the vertex v in the weighted RAG of level l be selected as a survivor. The vertex is connected by m_{v,v_i} path lengths of two or three to another survivor v_i , a neighbor in the weighted RAG of level l + 1. For every path we define $q_j(v, v_i)$ as the smallest weight along that path. Thus the path is characterized by the weakest link between the two survivors. Several methods can be employed to compute the weight of the edge between v and v_i in the l + 1 level's weighted RAG, $w(v, v_i)$. In a very conservative approach the minimum over all the $q_j(i, j), j = 1, ..., m_{v,v_i}$ values is taken, while in a less demanding application the maximum of these values can be used. We employed as compromise the average

$$w(v, v_i) = \frac{1}{m_{v, v_i}} \sum_{j=1}^{m_{v, v_i}} q_j(v, v_i).$$
(6)

In Fig. 8 an example is shown. At level *l*, the weighted RAG has six vertices corresponding to regions whose aver-

age gray level is shown inside the nodes (Fig. 8a). In a weighted RAG-based hierarchy, the undirected similarity subgraphs are obtained simply by setting a *context-independent* threshold $0 < T_{co} \le 1$. In Fig. 8a the threshold $T_{co} = 0.8$ was used to obtain the similarity subgraph (solid lines). Assume that vertices 2 and 5 are selected as survivors. The allocation of nonsurvivors is shown by the arrows. There are four path lengths of at the most three between the vertices 2 and 5 in the RAG of level *l*. These paths and the smallest weight along them are shown in Fig. 8b. The weight associated with the edge between vertices 2 and 5 in the RAG of level *l* = 1 can take different values depending on the employed updating strategy: 0.8 for maximum, 0.625 for average, and 0.4 for minimum.

3.2. Analysis of the Cooccurrence Probability Field

The weighted region adjacency graph can serve as the input of a relaxation process ([13, 16, 17]; Section 17.4 of



FIG. 9. Consensus segmentation of the *aerial* image. (a) $T_{co} = 0.8$. (b) After postprocessing with $T_{gf} = 18$.

b а 180 70 160 60 140 50 120 gray level difference 40 100 8(30 6(20 10 20 0.1 0.2 0.3 0.4 0.5 0.6 0.8 0.9 2 10 12 14 16 18 20 0.7 0 6 8 confidence gray level difference

FIG. 10. Selection of the gray-level fusion threshold, T_{gf} . (a) The gray-level differences between adjacent regions in Fig. 9a function of the corresponding weights in the weighted RAG. The average of the global thresholds (value 22) is shown with the dotted horizontal line. (b) The histogram of the gray-level differences considered for T_{gf} .

[19]) in which adjacent pixels having high cooccurrence probabilities are recursively joined together into homogeneous regions. To extract the homogeneous regions, using the technique developed for building the RAG pyramid, however, has several advantages:

• It converges in $O[\log(\text{image}_\text{size})]$ by reaching the root level of the hierarchy.

• The hierarchical structure combines nonlocal information more efficiently than the propagation in most relaxation methods.

• The same method is used for both the initial and the final segmentations providing a modular system.

The quality of segmentation is controlled by the threshold T_{co} defining the similarity subgraphs in the weighted RAG. This threshold is repeatedly decreased during the graph contractions. First $T_{co} = 1$ and thus the regions with the highest confidence are fused first. The segmentation process converges at the hierarchy level at which all the edges in the weighted RAG graph have weights less than T_{co} . That is, the confidence in the homogeneity of the delineated regions (the supports of the vertices) is at least T_{co} . The building of the RAG pyramid is then continued with T_{co} decreased to 0.9. After the second convergence, T_{co} is again decreased to 0.8. For a given T_{co} the influence of the probabilistic component in the graph contractions is reduced since the segmentation of the cooccurrence probability field is similar to the segmentation of a labeled image which was shown to have a deterministic output [27].

The segmentation derived from the cooccurrence probability field is a *consensus segmentation*. The cooccurrence probabilities are global measures for the reliability of local connections. When the value of a probability is high, there is consensus among the different initial segmentations that the two adjacent supports can be merged. In Fig. 9a the consensus segmentation of the *aerial* image obtained from the cooccurrence probability field in Fig. 7b is shown.

3.3. Postprocessing Techniques

Optimal postprocessing of oversegmented images should be goal oriented, i.e., based on a priori information about the sought features. In this section, only contextindependent techniques are discussed with the understanding that these techniques should be integrated into topdown, knowledge-driven procedures.



FIG. 11. Two segmentations of the *aerial* image (Fig. 3a). (a, b) Consensus images based on N = 20 initial segmentations. (c, d) Segmented images after postprocessing. (e, f) Boundaries of the delineated regions.

After the cooccurrence probability field was segmented, the confidence in the delineated regions was at least 0.8 (all adjacent pixels have a cooccurrence probability larger than 0.8). Since the consensus segmentation is obtained by analyzing only the cooccurrence probabilities, the segmented image may have adjacent regions whose gray-level difference is very small. In Fig. 10a the gray-level differences between all adjacent region pairs in Fig. 9a are plotted function of the corresponding weights in the weighted RAG.



FIG. 12. Outliers (white pixels) in regions larger than 25 pixels in the segmented image in Fig. 11c.

For fusion an appropriate gray-level fusion threshold, T_{gf} , must be chosen automatically. Only those gray-level differences of two adjacent regions are taken into account in the computation of T_{gf} which are less than the average of the global thresholds used in the initial segmentations and have the corresponding weight in the weighted RAG larger than or equal to 0.5. The histogram of those satisfying the above two conditions is in Fig. 10b. From this histogram, the most probable gray-level difference value, i.e., the mode of the histogram, is chosen as T_{gf} . In the example $T_{gf} = 18$ and the segmented *aerial* image after postprocessing is shown in Fig. 9b. There are no visible differences between the two segmentations, while the 1124 regions in Fig. 9a were reduced to 909 in Fig. 9b.

Due to noise and the employed simple homogeneity criterion the remaining delineated regions can have small sizes. The segmentation in Fig. 9b, for example, has 818 regions containing 9 or fewer pixels with a total area of only 1788 pixels. The average size of a small region thus is close to 2 pixels showing the sensitivity of the RAG pyramid-based segmentation technique. There are 91 regions larger than 9 pixels, and their total area of 14596 pixels is about 89% of the 128×128 image.

Should the features of interest have a few pixels area, they can now be extracted by a top-down process employing context specific information. In the absence of such information several techniques can be used to fuse the small regions together and/or into their larger neighbors. The simplest one is to recursively fuse the supports with area less than T_A (nine pixels in our example) into their most similar neighbor, i.e., having the least gray-level difference. At the end of postprocessing all small regions are eliminated. When the image in Fig. 9b is postprocessed with $T_A = 9$ the segmentation in Fig. 11c is obtained.

3.4. Experimental Results

In Figure 11 two segmentations of the *aerial* image are shown after postprocessing. The differences in the under-

lying cooccurrence probability fields (built from 20 initial segmentations) can be observed from the consensus images (Figs. 11a and 11b). The differences in the consensus segmentations followed by postprocessing (Figs. 11c and 11d) are not very significant as the boundaries of the delineated regions (Figs. 11e and 11f) also illustrate. This differences are mostly for the small regions and in the textured areas where the homogeneity criterion is inadequate. There are 114 and 122 regions in the two segmentations.

Forcing small regions to fuse into larger areas incorporates nonhomogeneities into regions initially delineated as homogeneous. These nonhomogeneities, however, can be recovered by robust analysis [24]. In every delineated region detecting the outliers relative to the mode of the graylevel distribution immediately highlights the incorporated nonhomogeneities. In Fig. 12 the outliers in regions larger than 25 pixels are shown for the segmentation in Fig. 11c. As expected most outliers are located near the discontinuities in the image.

The region adjacency graph of the final segmentation provides the spatial relationship between delineated regions and can be used for interactive analysis of the image. The hierarchical structure makes possible fast access to the characteristics of a delineated region chosen by a cursor location on the screen. See [10] for details of the implementation of such system.

The defined segmentation procedure is completely unsupervised. All the parameters are context-independent and/or are automatically derived from the input: the connectivity threshold CT(v), the smallest acceptable cooccurrence probability threshold T_{co} and the thresholds for post-processing T_{gf} and T_A .

To illustrate the unsupervised nature of the computations, in Fig. 13 two segmentations of the 200×150 roads image are shown. The original image (Fig. 13a) was segmented into 180 regions after postprocessing with $T_{gf} =$ 19 and $T_A = 9$. The average of the global thresholds was GT = 23. The image was then corrupted with additive Gaussian white noise with $\sigma = 20$ (Fig. 13b). The average global threshold was increased by the system to GT = 32, and the segmentation contained 161 regions after postprocessing with $T_{gf} = 27$ and $T_A = 9$. The noise occludes some of the weaker edges and less details are delineated. Note the delineations of several long features only a few pixels wide. Both segmentations were obtained from cooccurrence probability fields built from 20 initial segmentations and $T_{co} = 0.8$.

4. SPEEDUP OF THE COOCCURRENCE FIELD GENERATION

The cooccurrence probability field was generated from the set of initial segmentations. The initial segmentations can be obtained in parallel; however, most often such implementation is not available. On a Sparc 20 workstation, an average complexity 256×256 image takes about 20 min to be processed. In this section we discuss a faster method using only the local image structure around each

pixel for generating the cooccurrence probabilities. The tradeoff of accuracy for speed does not degrade significantly the performance as will be shown with experimental results.



FIG. 13. Another segmentation example. (a) Original image. (b) Noisy image corrupted with additive Gaussian noise, $\sigma = 20$. (c, d) Consensus segmentation and postprocessing. (e, f) Region boundaries.



FIG. 14. The allocation of pixels in the *aerial* according to the size of the associated window. The white pixels are those selected at that resolution. (a) 7×7 . (b) 5×5 . (c) 3×3 . (d) Adjacent neighbors only.

To determine the cooccurrence probability of a pixel pair, a $(2p + 1) \times (2p + 1)$ window is centered on every pixel. The optimum size of the window is set adaptively following the technique introduced in [25], and for each window a connectivity threshold is determined. A given pixel pair is contained in several windows. Note that because of the adaptive window size the number of windows is not always the same. The cooccurrence probability is then approximated by dividing the number of times the gray-level difference of the pixel pair was less than the local connectivity threshold to the number of times the pixel pair was examined.

4.1. Window Size Selection

The local component of the connectivity threshold is determined based on the gray-level differences in the window. For more reliable thresholds it is desirable to use larger windows for pixels in homogeneous regions. On the other hand, when a pixel is located close to an edge, smaller windows are needed to least corrupt the computations. To select the proper window size for each pixel, we employed the technique proposed in [25] for image smoothing. For every pixel the local gray-level variance in a $(2p + 1) \times (2p + 1), p = p_{\min}, \dots, p_{\max}$ window centered on the pixel is computed. For each window size, these local variances are accumulated and the mode of the global distribution, Γ_p^2 , is found. The value of Γ_p^2 is an approximative estimate of the noise power measured with $(2p + 1) \times (2p + 1)$ windows. Note that the noise includes any deviation from the piecewise constant image structure model.

A window is assumed to contain homogeneous data if its local variance is less than the corresponding global threshold Γ_p^2 . To select the largest window containing homogeneous data, the criterion is recursively examined from the $p_{\text{max}} = 3$ to $p_{\text{min}} = 1$. This technique was also used in [33] to determine the optimal window size for edge detection. Whenever none of the three windows was chosen (the pixel is close to an edge), only the differences between the pixel and its eight neighbors are considered



FIG. 15. Segmentation with locally estimated cooccurrence probabilities. (a) Segmented image. (b) Boundary image.

in the connectivity threshold computation. Classification of pixels in the *aerial* image according to the size of the associated window is shown in Fig. 14.

4.2. Connectivity Threshold

After the optimal window size is determined for the (i, j)th pixel, the *local* component of the connectivity threshold LT(i, j) can be defined as in Section 2. This time, however, the ordered sequence of gray-level differences is analyzed not for the neighbors on the region adjacency graph but for the entire window.

To avoid segmentation artifacts two global thresholds, GT^{l} and GT^{u} , are computed. The upper bound, GT^{u} , is used to prevent undersegmentation, while the lower bound, GT^{l} , is used to control oversegmentation. These global thresholds are derived from the global distribution of the gray-level differences of adjacent pixels from the entire image. The mode of the distribution, M, and the robust scale estimate $\hat{\sigma}$ (size of the mode search window) [24] are used to define the global thresholds as $GT^{l} = M + \hat{\sigma}$ and $GT^{u} = 2 \cdot GT^{l}$. These empirical thresholds seem to account for the strong skewness of the global distribution and were appropriate for a wide range of images [10]. To decide whether two adjacent pixels in the window centered on (i, j) belong to the same region, the connectivity threshold CT(i, j) is defined as

• if $LT(i, j) \leq GT^{l}$ then $CT(i, j) = GT^{l}$;

• if
$$GT^{l} < LT(i, j) \le GT^{u}$$
 then $CT(i, j) = LT(i, j)$;

• if $LT(i, j) \ge GT^u$ then $CT(i, j) = GT^u$.

Two adjacent pixels in the window are assumed to belong to the same region if their gray-level difference is less than CT(i, j).

Since windows are centered on every pixel, the connectivity of a pixel pair is examined in several windows. These windows can have different connectivity thresholds due to the local structure changes and thus the connectivity decision about the pixel pair is reassessed many times. The cooccurrence probability is defined as the ratio between the number of times the pixel pair was connected together, and the number of times it was examined. Note that the cooccurrence probabilities are now estimated from a small neighborhood. However, when this cooccurrence probability field is analyzed with the hierarchical technique discussed in Section 3.1 the output of the segmentation is not significantly different (Fig. 15).

The artifacts of employing rigid windows in the estimation of the probabilities are eliminated by using the RAG pyramid for the segmentation. For postprocessing $T_{gf} = GT^{l}$, and regions having area less than nine pixels are fused into their most similar neighbor, as discussed in Section 3.3. There are 99 regions in the segmented image (Fig. 15a).

4.3. Further Experimental Results

The two methods of obtaining the cooccurrence probability field were also compared for the *boat* image (Fig. 16). In Fig. 17, the cooccurrence probability field for the results on the left was generated from 20 initial segmenta-



FIG. 16. Boat image.



FIG. 17. Segmentations based on cooccurrence probability fields generated from 20 initial segmentations (left column) and with local windows (right column). (a, b) Consensus image. (c, d) Consensus segmentation and postprocessing. (e, f) The boundary image.

tions, while for the results on the right using the local windows. In the latter case $GT^{l} = 11$ and $GT^{u} = 22$. The thresholds $T_{co} = 0.8$, $T_{gf} = 11$, and $T_{A} = 9$ were used for both segmentations. The number of delineated regions is 600 in the left segmentation and 463 in the right segmentation.

The only significant difference between the two segmentations is that the local method may fail to delineate two regions sharing a blurred boundary. Since the analysis is restricted to windows of maximum size 7×7 , relatively high cooccurrence probabilities can be assigned to pixel pairs in the regions with slowly changing gray levels. The

IMAGE SEGMENTATION



FIG. 18. Consensus segmentation of three well known images. (a) The *pentagon* image. (d) The *house* image. (g) The *bulkhead* image. (b, e, h) The corresponding segmented images after postprocessing. (c, f, i) The corresponding boundary images.

delineation of the clouds (Fig. 17e and 17f) is an example of such cases. A speedup of about five times is obtained for the method using local windows. The exact processing times depend on the complexity of the segmented image.

Quantitative assessment of the quality of an image segmentation algorithm is a challenging open question, e.g., [18, 42]. To compare the proposed method with others, in Fig. 18 the segmentation of three well-known images is given. The segmented images are in the center and on the right the corresponding boundary images are shown. The original hierarchical procedure was used with the same thresholds as in Section 3.4, except for those recovered from the image itself.

The *pentagon* image (Fig. 18a) has relative low contrast and the detection of the fine features (some of the roads for example) appears to be somewhat random. However, the structure of the image is correctly recovered, in spite of using a hierarchical procedure for closely spaced narrow regions. The extracted 3768 regions were reduced to 408 after postprocessing.



FIG. 19. Roadmap detected from the segmented image in Fig. 13c with a probability of at least 0.6.

The *house* image (Fig. 18d) is representative of the class of outdoor scenes often used in segmentation papers. The presence of textured areas (trees, bushes, grass) challenges any segmentation algorithm. Note that they are satisfactorily recovered together with the features of the house. Some of the low contrast features on the lawn were fused. The extracted 4509 regions were reduced to 608 after postprocessing.

The *bulkhead* image (Fig. 18g) was extensively used in [17], as the reference image for presenting different segmentation algorithms ([17], Section 10). (There are two similar images used; ours is the one in Fig. 10.34.) The result in Fig. 18i is about the same quality as the output of a sophisticated edge detector (Fig. 10.27), which then was used to obtain a strong undersegmentation of the image (Fig. 10.28). The single hierarchical, split and merge, example (Fig. 10.40) is clearly inferior, producing severe artifacts. The extracted 3914 regions were reduced to 357 after postprocessing. It must be emphasized that the goal of the processing is to automatically obtain a correct oversegmentation of the image. The final segmentation must always be goal oriented, i.e., a knowledge guided top-down procedure.

5. DISCUSSION

The obtained segmentation is a slight oversegmentation into "puzzle pieces." The procedure is exclusively bottomup and to obtain regions corresponding to real parts of objects, top-down processing is also required. As an example of top-down processing, we used a simple Bayesian network [35] to extract the roads from the segmented image shown in Fig. 13c. Bayesian networks were successfully used to extract features from images [36]. Similar to [2] a region is characterized as a road candidate by simple features. Three such features were defined: the variance of the width of the region, the ratio of average width to length, and the average gray level. The width of a region was computed by using the distance transform ([17], Section 5.8). The Bayesian network computed the probability of a region as a road element using prior probabilities of 0.5 and the conditional (expert) probabilities derived from the image itself. The regions declared as road elements with a probability larger than or equal to 0.6 are shown in Fig. 19.

Image segmentation is a deceptively complex problem. Simple homogeneity criteria, like constant gray level cannot provide a reliable decomposition of a real image. The tradeoff between over and undersegmentation was discussed in Section 1. To take optimal decisions about fusing the delineated regions, the confidence in the homogeneity of the regions must be quantized. The consensus technique described in this paper is an effective way to obtain these quantitative measures. The differences in the initial segmentations are more significant for image parts where the employed homogeneity criterion is not adequate. These differences yield small cooccurrence probabilities and therefore can prevent undersegmentation.

Besides the cooccurrence probabilities, other measures can also be extracted from the initial segmentations. For example, from the empirical distribution of the N values of a pixel, the confidence in the stability of the segmentation *as seen by this pixel* can be obtained. Such measure can be used to replace locally the homogeneity criterion and use a higher order facet model.

The consensus methodology is not restricted to image segmentation. It is motivated by a resampling technique introduced recently in statistics, bootstrap [14]. Bootstrap can be used for a large variety of tasks in computer vision. We have already employed it for performance evaluation of a complete edge detection system [9] and for bias reduction in conic fitting [7]. Recently, a bootstrap based technique was proposed to improve the performance in machine learning [6]. The consensus paradigm may also have a biological justification. Information processing in the lower levels of the visual system is parallel with largely overlapping inputs, recalling information flow similar to the one in Fig. 1.

ACKNOWLEDGMENTS

We thank K. L. Boyer for providing the image in Fig. 13a, V. Ramesh for the image in Fig. 18g, and I. Marsic for the Bayesian network implementation. The support by the National Science Foundation under Grant IRI-9210861 is gratefully acknowledged.

REFERENCES

- S. Baronti, A. Casini, F. Lotti, L. Favaro, and V. Roberto, Variable pyramid structure for image segmentation, *Comput. Vision Graphics Image Process.* 49, 1990, 346–356.
- M. Barzohar and D. B. Cooper, Automatic finding of main roads in aerial images by using geometric—stochastic models and estimation, in *Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition '93, New York, 1993*, pp. 459–464.

- M. Basseville and A. Benveniste, Design and comparative study of some sequential jump detection algorithms for digital signals, *IEEE Trans. Acoustics, Speech, Signal Process.* 31, 1983, 521–535.
- J. R. Beveridge, J. Griffith, R. R. Kohler, A. R. Hanson, and E. M. Riseman, Segmenting images using localized histograms and region merging, *Int. J. Comput. Vision* 2, 1989, 311–347.
- M. Bister, J. Cornelis, and A. Rosenfeld, A critical view of pyramid segmentation algorithms, *Pattern Recognition Lett.* 11, 1990, 605–617.
- 6. L. Breiman, Bagging predictors, Mach. Learning 24, 1996, 123-140.
- 7. J. Cabrera and P. Meer, Unbiased estimation of ellipses by bootstrapping, *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 1996, 752–756.
- P. Chen and T. Pavlidis, Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm, *Comput. Graphics Image Process.* 10, 1979, 172–182.
- K. Cho, P. Meer, and J. Cabrera, Quantitative evaluation of performance through bootstrapping: Edge detection, in *Proceedings of IEEE International Symposium on Computer Vision, Coral Gables, 1995*, pp. 491–496.
- K. Cho, *Cooperative Processes in Computer Visions*, Ph.D. thesis, Department of Electrical Engineering, Rutgers University, NJ, October 1995.
- N. Christofides, Graph Theory: An Algorithmic Approach, Academic Press, London/New York/San Francisco, 1975.
- C. Chu and J. K. Aggarwal, The integration of image segmentation maps using region and edge information, *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 1993, 1241–1252.
- J. S. Duncan and T. Birkholzer, Reinforcement of linear structure using parametrized relaxation labeling, *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 1992, 502–515.
- 14. B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, London, 1993.
- D. Geiger and A. Yuille, A common framework for image segmentation, Int. J. Comput. Vision 6, 1991, 227–243.
- J. F. Haddon and J. F. Boyce, Image segmentation by unifying region and boundary information, *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 1990, 929–948.
- 17. R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, 2 volumes, Addison–Wesley, Reading, MA, 1992.
- A. Hoover, J.-B. Gillian, X. Jiang, P. J. Flynn, H. Bunke, D. Goldgof, and K. Bowyer, Range image segmentation: The user's dilemma, in *Proceedings of International Symposium on Computer Vision, Coral Gables, FL, November 1995*, pp. 323–328. [See also, TR 95-01, Computer Science and Electrical Engineering Department, University of South Florida, Tampa]
- R. A. Hummel and S. W. Zucker, On the foundations of relaxation labeling processes, *IEEE Trans. Pattern Anal. Mach. Intell.* 5, 1983, 267–286.
- J. Knapman and W. Dickson, Hierarchical probabilistic image segmentation, *Image Vision Computing* 12, 1994, 447–456.
- S. M. LaValle and S. A. Hutchinson, A framework for constructing probability distributions on the space of image segmentation, *Comput. Vision Image Understanding* 61, 1995, 203–230.
- 22. B. Lerner, W. J. Campbell, and J. LeMoigne, Image segmentation by integration of edge and region data: The influence of edge detection

algorithms, in Proceedings of 1994 ARPA Image Understanding Workshop, Monterey, 1994, pp. 1541–1545.

- 23. P. Meer, Stochastic image pyramids, *Comput. Vision Graphics Image Process* **45**, 1989, 269–294.
- P. Meer, D. Mintz, A. Rosenfeld, and D. Kim, Robust regression methods for computer vision: A review, *Int. J. Comput. Vision* 6, 1991, 59–70.
- P. Meer, R. H. Park, and K. Cho, Multiresolution adaptive image smoothing, *CVGIP: Graphical Models Image Process.* 56, 1994, 140–148.
- J. LeMoigne and J. C. Tilton, Refining image segmentation by integration of edge and region data, *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 1995, 605–615.
- A. Montanvert, P. Meer, and A. Rosenfeld, Hierarchical image analysis using irregular tessellations, *IEEE Trans. Pattern Anal. Mach. Intell.* 13, 1991, 307–316.
- O. Monga, An optimal region growing algorithm for image segmentation, Int. J. Pattern Recognition Artificial Intell. 1, 1987, 351–375.
- P. F. M. Nacken, Image segmentation by connectivity preserving relinking in hierarchical graph structures, *Pattern Recognition* 28, 1995, 907–920.
- A. M. Nazif and M. D. Levine, Low level image segmentation: an expert system, *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 1984, 555–577.
- I. Ng, J. Kittler, and J. Illingworth, Supervised segmentation using a multiresolution data representation, *Signal Process.* **31**, 1993, 133–163.
- N. R. Pal and S. K. Pal, A review on image segmentation techniques, *Pattern Recognition* 26, 1993, 1277–1294.
- D. J. Park, K. M. Nam, and R. H. Park, Multiresolution edge detection techniques, *Pattern Recognition* 28, 1995, 211–229.
- 34. T. Pavlidis and Y. Liow, Integrating region growing and edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 1990, 225–233.
- 35. J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
- S. Sarkar and K. L. Boyer, Intergration, inference, and management of spatial information using Bayesian networks: perceptual organization, *IEEE Trans. Pattern. Anal. Mach. Intell.* 15, 1993, 256–274.
- M. Spann and C. Horne, Image segmentation using a dynamic thresholding pyramid, *Pattern Recognition* 22, 1989, 719–732.
- D. Willersinn and W. G. Kropatsch, Dual graph construction for irregular pyramids, in *Proceedings of 12th International Conference* on Pattern Recognition, Jerusalem, 1994, pp. 251–256.
- D. Willersinn, Dual Irregular Pyramids, Ph.D. thesis, Technical University of Vienna, March, 1995.
- X. Wu, Adaptive split-and-merge segmentation based on piecewise least-square approximation, *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 1993, 808–815.
- Z. Wu and R. Leahy, An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 1993, 1101–1113.
- Y. L. Zhang and J. J. Gerbrands, Objective and quantitative segmentation evaluation and comparison, *Signal Process.* 39, 1994, 43–54.
- S. G. Ziavras and P. Meer, Adaptive multiresolution structures for image processing on parallel computers, *J. Parallel and Distrib. Computing* 23, 1994, 475–483.