NASA Technical Memorandum 104022

A Summary of Image Segmentation Techniques

Lilly Spirkovska, Ames Research Center, Moffett Field, California

June 1993

NASA
National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035-1000

Summary

Machine vision systems are often considered to be composed of two subsystems: low-level vision and high-level vision. Low-level vision consists primarily of image processing operations performed on the input image to produce another image with more favorable characteristics. These operations may yield images with reduced noise or cause certain features of the image to be emphasized (such as edges). High-level vision includes object recognition and, at the highest level, scene interpretation. The bridge between these two subsystems is the segmentation system. Through segmentation, the enhanced input image is mapped into a description involving regions with common features which can be used by the higher level vision tasks.

There is no theory on image segmentation. Instead, image segmentation techniques are basically ad hoc and differ mostly in the way they emphasize one or more of the desired properties of an ideal segmenter and in the way they balance and compromise one desired property against another.

These techniques can be categorized in a number of different groups including local vs. global, parallel vs. sequential, contextual vs. non contextual, interactive vs. automatic. In this paper, we categorize the schemes into three main groups: pixel-based, edge-based, and region-based. Pixel-based segmentation schemes classify pixels based solely on their gray levels. Edge-based schemes first detect local discontinuities (edges) and then use that information to separate the image into regions. Finally, region-based schemes start with a seed pixel (or group of pixels) and then grow or split the seed until the original image is composed of only homogeneous regions.

Because there are a number of survey papers available, we will not discuss all segmentation schemes. Rather than a survey, we take the approach of a detailed overview. We focus only on the more common approaches in order to give the reader a flavor for the variety of techniques available yet present enough details to facilitate implementation and experimentation.

Introduction

Machine vision systems are often considered to be composed of two sub-systems: low-level vision and high-level vision. Low-level vision consists primarily of image processing operations performed on the input image to produce another image with more favorable characteristics. These operations may yield images with reduced noise or cause certain features of the image to be emphasized (such as edges). High-level vision includes object recognition and, at the highest level, scene interpretation. The bridge between these two subsystems is the segmentation system. Through segmentation, the enhanced input image is mapped into a description involving regions with common features which can be used by the higher level vision tasks. On one hand, this procedure should be sensitive enough to extract those areas of interest in the image. On the other hand, it should be immune to the disturbances of irrelevant objects and noise in the system.

Ideally, a good segmenter should produce regions which are uniform and homogeneous with respect to some characteristic such as gray tone or texture yet simple, without many small holes. Further, the boundaries of each segment should be spatially accurate yet smooth, not ragged. And finally, adjacent regions should have significantly different values with respect to the characteristics on which region uniformity is based. This situation can be represented mathematically as follows:

If I is the set of all pixels and $P()$ is a uniformity predicate defined on groups of connected pixels, a segmentation of I is a partitioning set of connected subsets or image regions $\{R_1, R_2, ..., R_n\}$ such that

$$\bigcup_{l=1}^{n} Rl = I, \text{ where } Rl \bigcap Rm = \varnothing \quad \forall\ l \neq m \qquad (1)$$

and the uniformity predicate (such as nearly equal gray level) satisfies

$$P(Rl) = \text{True}\ \forall l \qquad (2)$$

$$P(Rl \bigcup Rm) = \mathit{False}, \ \forall R_l \text{ adjacent to } R_m \qquad (3)$$

$$(Rl \supset Rm) \wedge (Rm \neq \varnothing) \wedge (P(Rl) = \text{True}) \Rightarrow P(Rm) = \text{True} \qquad (4)$$

Because noise destroys homogeneity in a local context, it is not possible to determine a consistent homogeneity of larger regions, resulting in fragmented segmentation results. If noise characteristics are known, however, it is possible to determine homogeneity on statistical grounds. In this case, we must drop the consistency criterion given by equation (4) which states that if a region is homogeneous, then all subsets of this region will also be homogeneous. This means that a region may be determined to be homogeneous even when subsets of this region are not.

There is no theory on image segmentation. Instead, image segmentation techniques are basically ad hoc and differ mostly in the way they emphasize one or more of the desired properties of an ideal segmenter and in the way they balance and compromise one desired property against another. These techniques can be categorized in a number of different groups including local vs. global, parallel vs. sequential, contextual vs. non contextual, interactive vs. automatic. In this paper, we categorize the schemes into three main groups: pixel-based, edge-based, and region-based. Pixel-based segmentation schemes classify pixels based solely on their gray levels. Edge-based schemes first detect local discontinuities (edges) and then use that information to separate the image into regions. Finally, region-based schemes start with a seed pixel (or group of pixels) and then grow or split the seed until the original image is composed of only homogeneous regions.

Because there are a number of survey papers available (Sahoo et al., 1988; Weszka, 1978; Haralick and Shapiro, 1985), we will not discuss all segmentation schemes. Rather than a survey, we take the approach of a detailed overview. We focus only on the more common approaches in order to give the reader a flavor for the variety of techniques available yet present enough details to facilitate implementation and experimentation.

Pixel-Based Segmentation Schemes

Mode Method

The most widely used segmentation technique is the mode method which is applicable to images with bimodal histograms, as shown in figure 1. One mode of the histogram corresponds to the gray levels of the object pixels while the other mode captures the gray levels of the background pixels. It is assumed that a fixed threshold level exists that separates the background area from the objects. The threshold level is chosen to be the gray level in between the two modes using any of a number of different methods. The two most popular methods are Gaussian filtering (Jain and Dubuisson, 1992) and Otsu's method based on discriminant analysis (Otsu, 1979).

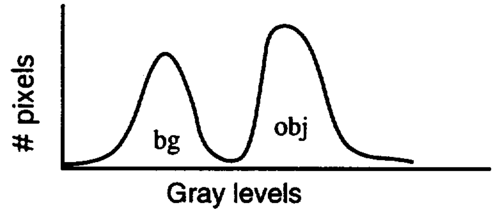

Figure 1: A bimodal histogram. One mode represents the background pixels while the other represents the object pixels.

Gaussian filtering algorithm– The simplest segmentation method is based on the Bayes decision theory in pattern recognition. The gray level histogram of the image is computed and then two component densities are extracted (corresponding to the object and the background) from the mixture density associated with the histogram. It is commonly assumed that both the background and the object densities are Gaussian.

Algorithm:

1. Compute the mean (μ) and standard deviation (σ) of the histogram:

$$\mu = \frac{1}{N}\sum F(i) * i \tag{5}$$

$$\sigma = \sqrt{\frac{1}{N}\sum F(i) * (i-\mu)^2} \tag{6}$$

where $F(i)$ is the histogram value for gray level i (out of L gray levels) and N is the number of points in the window.

To avoid the problem of division by 0 (for the deviation is necessarily 0 for 1-pixel regions and regions having identically valued pixels), a small positive constant can be added to σ.

2. Find a least-squares fit of

$$f(i) = \frac{P_1}{\sigma_1} e^{-(i-\mu_1)^2/2\sigma_1^2} + \frac{P_2}{\sigma_2} e^{-(i-\mu_2)^2/2\sigma_2^2} \tag{7}$$

to the histogram F(i) by adjusting the parameters $P_1, \mu_1, \sigma_1, P_2, \mu_2, \sigma_2$ as follows:

(i) Smooth the histogram by taking a local weighted average:

$$F'(i) = \frac{F(i-2) + 2F(i-1) + 3F(i) + 2F(i+1) + F(i+2)}{9} \tag{8}$$

On the smoothed histogram, find the deepest valley v (= lowest value) and use it to divide the histogram into two parts. Compute initial estimates of $P_1, \mu_1, \sigma_1, P_2, \mu_2$, and σ_2 on these two parts (for the original histogram F(i)):

$$N_1 = \sum_{i=1}^{v} F(i), \qquad N_2 = \sum_{i=v+1}^{L} F(i) \tag{9}$$

$$\mu_1 = \frac{1}{N_1}\sum_{i=1}^{v} F(i) * i, \quad \mu_2 = \frac{1}{N_2}\sum_{i=v+1}^{L} F(i) * i \tag{10}$$

$$\sigma_1 = \sqrt{\frac{1}{N_1}\sum_{i=1}^{v} F(i) * (i-\mu_1)^2} \quad (11)$$

$$\sigma_2 = \sqrt{\frac{1}{N_2}\sum_{i=v+1}^{L} F(i) * (i-\mu_2)^2} \quad (12)$$

$$\sigma_2 = \sqrt{\frac{1}{N_2}\sum_{i=v+1}^{L} F(i) * (i-\mu_2)^2} \quad (13)$$

(ii) Use a hill-climbing method to minimize:

$$\sum_{i=1}^{L} [f(i) - F(i)]^2 \quad (14)$$

(a) Calculate: val = |f(i) – F(i)| for i = deepest valley (v) chosen in step (i) or (ii.d).

(b) Calculate: left_val = |f(i – 1) – F(i – 1)|.

(c) Calculate: right_val = |f(i + 1) – F(i + 1)|.

(d) If (left_val ≤ val), set the estimate for deepest valley at *i – 1*.

Else if (right_val ≤ val), set the estimate for deepest valley at *i + 1*.

Else deepest valley found at *i*.

(e) If the deepest valley value was changed in step (d), reestimate P_1, μ_1, σ_1, P_2, μ_2, and σ_2 using equations (9–13) and the new value of v. Repeat steps (a–d).

3. After the parameters of the mixture density have been estimated, a pixel with gray level *x* is assigned to the object if

$$\frac{P_1}{\sigma_1} e^{-(x-\mu_1)^2/2\sigma_1^2} > \frac{P_2}{\sigma_2} e^{-(x-\mu_2)^2/2\sigma_2^2} \quad (15)$$

The threshold value t is then defined as

$$\frac{P_1}{\sigma_1} e^{-(t-\mu_1)^2/2\sigma_1^2} = \frac{P_2}{\sigma_2} e^{-(t-\mu_2)^2/2\sigma_2^2} \quad (16)$$

and satisfies:

$$\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) t^2 + 2\left(\frac{\mu_2}{\sigma_2^2} - \frac{\mu_1}{\sigma_1^2}\right) t + \frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2} + 2\ln\frac{P_2\sigma_1}{P_1\sigma_2} = 0 \quad (17)$$

Otsu's algorithm– Otsu's method of determining a threshold in a bimodal histogram is based on discriminant analysis in which thresholding is regarded as the partitioning of the pixels of an image into two classes C_0 and C_1 at gray level *t*.

Algorithm:

n_i = number of pixels at level i (from L gray levels)

N = total number of pixels = $n_1 + n_2 + \ldots + n_L$

1. The gray level histogram is normalized and regarded as a probability distribution:

$$p_i = n_i / N \quad (18)$$

$$p_i \geq 0 \quad (19)$$

$$\sum_{i=1}^{L} p_i = 1 \quad (20)$$

2. Dichotomize pixels into two classes C_0 and C_1 by a threshold at level *k*.

3. Calculate the probabilities of class occurrence:

$$w_0 = \Pr(C_0) = \sum_{i=1}^{k} p_i = w(k) \quad (21)$$

$$w_1 = \Pr(C_1) = \sum_{i=k+1}^{L} p_i = 1 - w(k) \quad (22)$$

4. Calculate the class mean levels:

$$\mu_0 = \sum_{i=1}^{k} i \Pr(i \mid C_0) = \sum_{i=1}^{k} ip_i / w_0 = \mu(k) / w(k) \quad (23)$$

$$\mu_1 = \sum_{i=k+1}^{L} i \Pr(i \mid C_1) = \sum_{i=k+1}^{L} ip_i / w_1 = (\mu_T - \mu(k)) / (1 - w(k)) \quad (24)$$

where w(k) and μ(k) are the zeroeth and first-order cumulative moments of the histogram up to the kth level and

$$\mu_1 = \sum_{i=k+1}^{L} i \Pr(i \mid C_1) = \sum_{i=k+1}^{L} ip_i / w_1 = (\mu_T - \mu(k)) / (1 - w(k)) \quad (25)$$

is the total mean level of the original picture.

5. Calculate class variances:

$$\sigma_0^2 = \sum_{i=1}^{k} (i-\mu_0)^2 \Pr(i \mid C_0) = \sum_{i=1}^{k} (i-\mu_0)^2 p_i / w_0 \quad (26)$$

$$\sigma_1^2 = \sum_{i=k+1}^{L} (i-\mu_1)^2 \Pr(i \mid C_1) = \sum_{i=k+1}^{L} (i-\mu_1)^2 p_i / w_1 \quad (27)$$

6. In order to evaluate the "goodness" of the threshold k, we can use the following discriminant criterion measures (or measures of class separability):

$$\lambda = \frac{\sigma_B^2}{\sigma_W^2}, \quad \kappa = \frac{\sigma_T^2}{\sigma_W^2}, \quad \eta = \frac{\sigma_B^2}{\sigma_T^2} \qquad (28)$$

where

$$\sigma_W^2 = w_0\sigma_0^2 + w_1\sigma_1^2 \qquad (29)$$

is the within-class variance,

$$\begin{aligned} \sigma_B^2 &= w_0(\mu_0 - \mu_T)^2 + w_1(\mu_1 - \mu_T)^2 \\ &= w_0 w_1 (\mu_1 - \mu_0)^2 \end{aligned} \qquad (30)$$

is the between-class variance, and

$$\sigma_T^2 = \sum_{i=1}^{L} (i - \mu_T)^2 p_i \qquad (31)$$

is the total variance.

Note that λ, κ, and η are equivalent to one another for a given k because

$$\sigma_W^2 + \sigma_B^2 = \sigma_T^2 \qquad (32)$$

7. The problem is now reduced to an optimization problem to search for a threshold k that maximizes one of the object functions (the criterion measures). Note that σ_W^2 and σ_B^2 are functions of threshold level k, whereas σ_T^2 is independent of k. Further, σ_W^2 is based on second-order statistics while σ_B^2 is based on first-order statistics. Thus, we use η since it is the simplest measure with respect to k:

$$\eta = \frac{\sigma_B^2}{\sigma_T^2} \qquad (33)$$

Since σ_T^2 is independent of k, we can maximize η by maximizing $\sigma_B^2(k)$:

$$\sigma_B^2(k) = \frac{[\mu_T w(k) - \mu(k)]^2}{w(k)[1 - w(k)]} \qquad (34)$$

Thus, the optimal threshold t is chosen to be that k which maximizes $\sigma_B^2(k)$.

The threshold determination methods discussed above work well when the object size is large enough to make a distinct mode in the histogram, the gray level noise distribution (intensity noise) is independent of the gray level, and the noise is spatially uncorrelated. The methods fail when it is difficult to detect the valley bottom, as in images with extremely unequal peaks or in those with broad and flat valleys. Since peaks tend to become wider and lower with an increasing amount of intensity noise, some sharpening of the peaks and valleys can be accomplished by applying noise reduction preprocessing procedures.

Another approach to sharpening peaks and valleys is to weigh the influence of individual pixels and not count them all equally when calculating the histogram, as in the gradient-guided methods. Gradient guided histograms take one of two forms, interior only or boundary only. The interior-only methods (Mason et al., 1975; Panda and Rosenfeld, 1978) take into account only pixels belonging to either the objects or the background (i.e., those pixels having low gradient values), disregarding pixels belonging to boundaries where the gray level varies rapidly. This should yield a histogram with sharper peaks and deeper valleys. In contrast, the boundary-only methods (Weszka, Nagel, and Rosenfeld, 1974; Watanabe et al., 1974) take into account only pixels belonging to boundaries (i.e., those pixels having high gradient values). This should yield a well-defined unimodal histogram, the peak value of which is a proper constant threshold level.

Finally, instead of computing a 1D histogram of gray level values, a 2D histogram or "scatter" diagram can be computed with gray level and gradient as its coordinates. In this case, a good threshold can be selected using clusters of points rather than the modes of a histogram (Katz, 1965; Weszka and Rosenfeld, 1979).

Local methods– Global segmentation techniques such as the mode method are notoriously sensitive to parameters like ambient illumination, object shape and size, noise level, variance of gray levels within the object and background, and contrast (Taxt et al., 1989). When there is a large range of variation in gray values from one part of the image to the other, a single threshold value cannot be used. Further, objects may legitimately have widely different albedos and, as a result, an object in one part of an image may be lighter than the background in another part. Local methods attempt to eliminate these disadvantages by partitioning the image into subimages, determining a threshold for each of these subimages, and then smoothing between subimages to eliminate discontinuities. An example of this group of methods is the

Chow–Kaneko adaptive thresholding method (Chow and Kaneko, 1972). This method assigns a different threshold value to each pixel.

Chow–Kaneko adaptive thresholding algorithm–

1. Subdivide the image into several subimages.

2. For each subimage, compute the histogram, smooth it, and determine a threshold using the mode method.

3. Smooth the thresholds among the neighboring subimages.

4. Determine a threshold for each pixel by interpolation. For example, to interpolate the 2×2 image:

$$\begin{bmatrix} 4 & 7 \\ 7 & 10 \end{bmatrix}$$

into a 4×4 image, begin in the columns direction and form a 2×4 image:

$$\begin{bmatrix} 4 & 5 & 6 & 7 \\ 7 & 8 & 9 & 10 \end{bmatrix}$$

Then interpolating in the rows direction, form the desired 4×4 image:

$$\begin{bmatrix} 4 & 5 & 6 & 7 \\ 5 & 6 & 7 & 8 \\ 6 & 7 & 8 & 9 \\ 7 & 8 & 9 & 10 \end{bmatrix}$$

5. Threshold the image using the threshold value assigned to each pixel.

The biggest determinant of whether a local thresholding method produces an acceptable segmentation is the size of the subwindows. If it is chosen to be too large, the algorithm will not focus on local properties and will not perform significantly better than global techniques. On the other hand, if it is chosen to be too small, the histograms produced for each subwindow would yield meaningless statistics since the number of pixels participating in the process would be reduced substantially. Unfortunately, the best method of choosing an appropriate window size is by trial-and-error.

Even if window size is chosen well, the grid imposed on the image may not be coherent with the image contents and thus the threshold values determined within a subwindow would be set at arbitrary locations instead of being placed in truly meaningful positions. Purely local techniques are blind to trends in the data that are significantly larger than their element size. Finally, serious errors can occur if, due to noise and bad lighting conditions, grid windows placed entirely on object areas or entirely on background areas, by chance yield subhistograms that are judged to be bimodal.

Edge-Based Segmentation Schemes

Edge-based segmentation schemes also take local information into account but do it relative to the contents of the image, not based on an arbitrary grid. Each of the methods in this category involves finding the edges in an image and then using that information to separate the regions. The most direct method is the detect and link method in which local discontinuities are first detected and then connected to form longer, hopefully complete, boundaries. The disadvantage of this approach is that the edges are not guaranteed to form closed boundaries and thus the subsequent thresholding scheme merges regions which may not be uniform (relative to the uniformity predicate discussed in the introduction).

An improvement over this method is Yanowitz and Bruckstein's adaptive thresholding method (Yanowitz and Bruckstein, 1989). Similar to the detect and link method, the adaptive thresholding method locates objects in an image by using the intensity gradient. These edges are used as a guide to determine an initial threshold level for various areas of the image. Local image properties are then used to assign thresholds to the remainder of the image.

Another improvement of the detect and link method is the local intensity gradient (LIG) method (Parker, 1991). It is similar to Yanowitz and Bruckstein's algorithm and works well in practice. We present each algorithm below.

Yanowitz and Bruckstein's Adaptive Thresholding Algorithm

1. Smooth the image, replacing every pixel by the average gray-level values of some small neighborhood of it.

2. Derive the gray-level gradient magnitude image from the smoothed original. In discrete images, the gradient is actually computed as an intensity difference over a small distance:

$$G(i,j) = \min(I(i,j) - I(i+\delta_i, j+\delta_j)) \qquad \delta_i = -1,1, \quad \delta_j = -1,1 \tag{35}$$

where I is the image being examined and G is the resulting image consisting of differences.

3. Thin the gradient image, keeping only points in the image which have local gradient maxima. This should produce a one-pixel wide edge.

4. Sample the smoothed image at these maximal gradient (or edge) points. These points are assumed to be correct.

5. Interpolate the sampled gray levels over the image. The result is a threshold surface, with a (possibly) different threshold value for each pixel.

6. Using the obtained threshold surface, segment the image. If the original pixel value is greater than or equal to the threshold value at that location, set the thresholded value to 1 (or white). Otherwise, set the value to 0 (or black). Thus, objects will be set to white and background to black.

Local Intensity Gradient Method Algorithm

The local intensity gradient method (Parker, 1991) is based on the fact that objects in an image will produce small regions with a relatively large intensity gradient (at the boundaries of objects) whereas other areas ought to have a relatively small gradient. It uses small subimages of the gradient image to find local means and deviations. As in the local mode techniques, these regions must be small enough so that the illumination effects can be ignored.

1. Compute a smooth gradient of the image.

- For all pixels in the N × N image (IM1), compute the minimum difference between the pixel and all eight neighbors. (See gradient computation in step 2 of Yanowitz and Bruckstein's algorithm.) Store in IM2.

- Break up IM2 (gradient array) into subregions, each composed of M × M pixels. Compute the mean value (QIM) and mean deviation (QDEV) for each subregion *k*:

$$\mathrm{QIM}[k] = \frac{1}{M*M}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathrm{IM2}[i][j] \tag{38}$$

$$\mathrm{QDEV}[k] = \sqrt{\frac{1}{M*M}\sum_{i=1}^{M}\sum_{j=1}^{M}(\mathrm{IM2}[i][j]-\mathrm{QIM}[k])^2} \tag{39}$$

- Smooth both QIM and QDEV. The value of QIM (and QDEV) at each point is replaced by the weighted mean of all the neighboring subregions using the following weight matrix:

$$\begin{bmatrix} 0.7 & 1.0 & 0.7 \\ 1.0 & 1.5 & 1.0 \\ 0.7 & 1.0 & 0.7 \end{bmatrix}$$

- Interpolate/extrapolate the values of QIM and QDEV to fill an N × N region again. Linear interpolation is acceptable. This results in an image containing estimates of the gradient and deviation at each pixel.

2. Find the object pixels in the gradient image. Object pixels are defined as outliers in the smoothed image. That is, pixel [i,j] is an object pixel if IM2[i,j] ≤ QIM[i,j] + 3*QDEV[i,j]. Otherwise [i,j] is unclassified.

3. After sample object pixels are found, thresholds can be identified for the remaining pixels. This can be done using a region growing procedure based on gray levels in the local surroundings, and begins at pixels that are known to belong to the object.

- For all unclassified pixels [i,j], select a gray level threshold by finding the smallest gray level value in an 8-neighborhood (pixel aggregation). If IM1[i,j] is less than this value, it is an object pixel. Repeat until no more object pixels are found.

- (Optional) For all still unclassified pixels, compute a threshold as the mean of at least four neighboring object pixels (region growing). Repeat until no new object pixels are found.

- (Optional) For all still unclassified pixels, compute a threshold as the minimum of at least six neighbors (region growing). Repeat until no new object pixels are found.

4. Set all object pixels in IM1 to the value 0 and all unclassified pixels to a positive value. IM1 now contains the thresholded image.

Region-Based Segmentation Schemes

Region-based segmentation schemes attempt to group pixels with similar characteristics (such as approximate gray level equality) into regions. Conventionally, these are global hypothesis testing techniques. The process can start at the pixel level or at an intermediate level. Generation and filtering of good seed regions of high confidence is essential. Given initially poor or incorrect seed regions, these techniques usually do not provide any mechanism for detecting and rejecting local gross errors in situations such as when an initial seed region spans two separate surfaces. These techniques can also fail if the definition of region uniformity is too strict, such as when we insist on approximately constant brightness while in reality brightness may vary linearly. Another potential problem with region growing schemes is their inherent dependence on the order in which pixels and regions are examined. Usually, however, differences caused by scan order are minor.

There are two approaches in region-based methods: region growing and region splitting. In the region growing methods, the evaluated sets are very small at the start of the segmentation process. The iterative process of region growing must then be applied in order to recover the surfaces of interest. In the region growing process, the seed region is expanded to include all homogeneous neighbors and the process is repeated. The process ends when there are no more pixels to be classified. Because initial seeds are very small, the processing time can be minimized by minimizing the number of times an image element is used to determine the homogeneity of a region.

In region splitting methods, on the other hand, the evaluation of homogeneity is made on the basis of large sets of image elements. The process starts with the entire image as the seed. If the seed is inhomogeneous, it is split into a predetermined number of subregions, typically four. The region splitting process is then repeated using each subregion as a seed. The process ends when all subregions are homogeneous. Because the seeds being processed at each step contain many pixels, region splitting methods are less sensitive to noise than the region growing methods. In both approaches, their iterative structure leads to computationally intensive algorithms.

In the late 70s, Horowitz and Pavlidis developed a hybrid algorithm, the split, merge, and group (SMG) algorithm (Horowitz and Pavlidis, 1976; Chen and Lin, 1991), which incorporates the advantages of both approaches. Because the SMG algorithm begins at an intermediate resolution level, it is more efficient than either the pure split algorithms or the pure merge algorithms. The major disadvantage, however, is that the resulting image tends to mimic the data structure used to represent the image (a quadtree for 2D images or an octree or K-tree for 3D images) and comes out too square.

Split, Merge, and Group Algorithm

1. Initialization phase:

Divide the image into subimages using a quadtree structure, as shown in figure 2. The root of the quadtree corresponds to the whole image. Each node in the quadtree has only one parent (except for the root) and four children (except for the leaves). The four children are denoted by the quadrant within the parent that they correspond to (NW, NE, SW, SE). Thus, the image must be $2^n \times 2^n$ pixels. The leaves are at node level 0. The root is at level n. During initialization, the quadtree is built from the root down to a heuristically set initialization level L_S. The choice of the initialization level L_S can be selected in terms of minimizing the expected number of splits and merges.

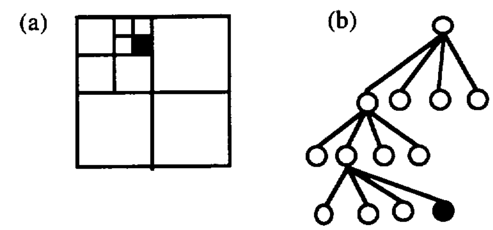

Figure 2: (a) Original image split using quadtree.
(b) Quadtree representation of image.

2. Merge phase:

If level L_S is homogeneous, evaluate the homogeneity of level $L_S + 1$. If the node is homogeneous, the four children are cut and the node becomes an endnode. Repeat until no merges take place or level n is reached (a homogeneous image).

3. Split phase:

If level L_S is inhomogeneous, split the nodes into four children and add them to the quadtree. Evaluate the new endnodes and if necessary, split again until the quadtree has homogeneous endnodes only.

4. Conversion from quadtree to RAG phase:

• A RAG is a Region Adjacency Graph. It allows different subimages that are adjacent but cannot be merged in the quadtree to be merged.

• This phase consists of extracting the implicit adjacency relations from quadtree endnodes needed to construct the branches of the RAG. Two neighboring subimages in the quadtree will have common ancestors, i.e., nodes in the quadtree on a higher level from which both endnodes can be reached.

• First, the nearest common ancestor is determined that connects the current endnode with the neighbor. Next, the path is mirrored about an axis formed by the common boundary between the adjacent subimages.

5. Grouping phase:

The now explicit neighbor relationships can be used to merge adjacent nodes which have a homogeneous union. Grouping strategies include:

a. Assign the first node of the RAG (corresponding with the subimage in the top left corner) the status of seed. The neighbors of the seed are then evaluated on homogeneity together with the seed. A merge of a neighbor with the seed produces new neighbors, which are evaluated. When no more grouping takes place, the seed is rendered inactive and a new seed (the first unprocessed node) is assigned. The grouping phase ends if all remaining RAG nodes have become inactive.

b. Sequential grouping: the seeds are chosen based on their size with the first seed being the largest sub-image, etc. A disadvantage of this approach is that because of the size of the first seed, these regions tend to grow beyond their "actual" boundaries, annexing all fuzzy border areas.

c. Parallel grouping: assign a number of active seeds at the start of the grouping phase. Now only direct neighbors of a seed are grouped if possible. New neighbors have to wait for evaluation until the seed is processed again. Active seeds are processed successively until none remain. The growing of seeds will be bounded by other seeds.

Grouping strategy (a) is sufficient in practice.

6. Postprocessing of the RAG phase:

- If subimages are too small, merge them with their nearest neighbor. It is difficult to interpret very small regions as objects and since there is usually a relatively large number of them, their presence increases the computational burden on later stages of processing.

- Exploit prior knowledge about the application problem to improve the segmentation.

Concluding Remarks

The goal of image segmentation is a domain-independent decomposition of an image into distinct regions which are uniform in some measurable property such as brightness, color, or texture. Unfortunately, natural scenes often contain feature gradients, highlights, shadows, textures, and small objects with fine geometric structure, all of which make the process of producing useful segmentations extremely difficult. We have presented some of the techniques which attempt to deal with these difficulties. Although they produce reasonable segmentations in many situations, at some point local ambiguities and errors introduced by the segmentation process can only be resolved by application specific knowledge.

Since the quality of the above segmentation techniques depends on the type of image the technique is being applied to, we will end this overview with a summary of what type of image each technique works best on.

- The mode method is applicable to images with bimodal histograms where the modes are fairly distinct (well separated and sharp) and of nearly equal length. It does not work well if the gray level noise distribution is dependent on the gray level or is spatially correlated.

- Local methods, such as the Chow–Kaneko method, are applicable to images in which the ambient illumination may vary in gray level from one part of the image to another or when one part of an image may be lighter than the background in another (as long as the contrast in each area is adequate). The major disadvantage of local methods is that it is difficult to choose an appropriate window size which localizes the illumination variation yet considers a large enough area to yield meaningful statistics. Also, even if the window size is chosen well, the grid imposed on the image may not be coherent with the image contents and thus threshold values determined within a subwindow would be set at arbitrary positions instead of being placed in truly meaningful positions.

- Because the quality of the segmentation depends on the quality of the edge detector, edge-based schemes work best on images in which the edges are easily detectable—that is, images which have good local (5×5 pixel area or less) contrast. They do not work well with images in which the noise forms well-defined edges.

- Region-based schemes work well for images with an obvious homogeneity criteria (such as nearly equal gray level). Also, these schemes tend to be less sensitive to noise since homogeneity is typically determined statistically. Their disadvantages are that an initial split-level must be chosen well (else the technique could be very slow) and the segmented image tends to mimic the data structure used to represent the image and is thus too square.

References

Ballard, D. H.; and Brown, C. M.: Principles of Animate Vision. CVGIP: Image Understanding, vol. 56, no. 1, 1992, pp. 3–21.

Basu, S.: Image Segmentation by Semantic Method. Pattern Recognition, vol. 20, no. 5, 1987, pp. 497–511.

Besl, P. J.; and Jain, R. C.: Segmentation through Variable-Order Surface Fitting. IEEE PAMI, vol. 10, no. 2, 1988, pp. 167–192.

Berg, T. B.; Kim, S-D.; and Siegel, H. J.: Limitations Imposed on Mixed-Mode Performance of Optimized Phases Due to Temporal Juxtaposition. Journal of Parallel and Distributed Computing, vol. 13, 1991, pp. 154–169.

Beveridge, J. R.; Griffith, J.; Kohler, R. R.; Hanson, A. R.; and Riseman, E. M.: Segmenting Images Using Localized Histograms and Region Merging. International Journal of Computer Vision, vol. 2, 1989, pp. 311–347.

Bhanu, B.; and Parvin, B. A.: Segmentation of Natural Scenes. Pattern Recognition, vol. 20, no. 5, 1987, pp. 487–496.

Brink, A. D.: Comments on Grey-Level Thresholding of Images Using a Correlation Criterion. Pattern Recognition Letters, vol. 12, 1991, pp. 91–92.

Burt, P. J.; Hong, T.-H.; and Rosenfeld, A.: Segmentation and Estimation of Image Region Properties through Cooperative Hierarchical Computation. IEEE Transactions on Systems, Man, and Cybernetics, SMC, vol. 11, no. 12, 1981, pp. 802–809.

Chen, S-Y.; and Lin, W-C.: Split-and-Merge Image Segmentation Based on Localized Feature Analysis and Statistical Tests. CVGIP: Graphical Models and Image Processing, vol. 53, no. 5, 1991, pp. 457–475.

Chow, C. K.; and Kaneko, T.: Boundary Detection of Radiographic Images by a Thresholding Method. Frontiers of Pattern Recognition, S. Watanabe, ed., Academic Press, New York, 1972, pp. 61–82.

Daniell, C.; Kemsley, D.; and Bouyssounouse, X.: Comparative Evaluation of Neural Based Versus Conventional Segmentors. SPIE, vol. 1471, 1991, pp. 436–451.

Dickson, W.: Feature Grouping in a Hierarchical Probabilistic Network. Image and Vision Computing, vol. 9, 1991, pp. 51–57.

Doyle, W.: Operation Useful for Similarity-Invariant Pattern Recognition. J. Assoc. Computing Machinery, vol. 9, 1962, pp. 259–267.

du Buf, J. M. H.; Kardan, M.; and Spann, M.: Texture Feature Performance for Image Segmentation. Pattern Recognition, vol. 23, nos. 3 and 4, 1990, pp. 291–309.

Dubes, R. C.; Jain, A. K.; Nadabar, S. G.; and Chen, C. C.: MRF Model-Based Algorithms for Image Segmentation. Proc. 10th Int. Conf. on Pattern Recognition, Atlantic City, NJ, 1990, pp. 808–814.

Freeman, M. O.; and Saleh, B. E. A.: Moment Invariants in the Space and Frequency Domains. J. Opt. Soc. Am. A, vol. 5, no. 7, 1988, pp. 1073–1084.

Freeman, M. O.; and Saleh, B. E. A.: Optical Location of Centroids of Nonoverlapping Objects. Applied Optics, vol. 26, no. 14, 1987, pp. 2752–2759.

Fong, Y-S.; and Brown, D. H.: A Centroid Tracking Scheme in a Weighted Coordinate System. IEEE 1985 CVPR, 1985, pp. 219–221.

Geman, S.; and Geman, D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE PAMI PAMI, vol. 6, no. 6, 1984, pp. 721–741.

Geiger, D.; and Yuille, A.: A Common Framework for Image Segmentation. Int. Journal of Computer Vision, vol. 6, no. 3, 1991, pp. 227–243.

Grossberg, S.; and Mingolla, E.: Neural Dynamics of Perceptual Grouping: Textures, Boundaries, and Emergent Segmentations. Perception and Psychophysics, vol. 38, no. 2, 1985, pp. 141–171.

Haralick, R. M.; and Shapiro, L. G.: Survey: Image Segmentation Techniques. CVGIP, vol. 29, 1985, pp. 100–132.

Horowitz, A. L.; and Pavlidis, T.: Picture Segmentation by a Tree Traversal Algorithm. JACM, vol. 23, no. 2, 1976, pp. 368–388.

Jain, A. K.; and Dubuisson, M-P.: Segmentation of X-Ray and C-Scan Images of Fiber Reinforced Composite Materials. Pattern Recognition, vol. 25, no. 3, 1992, pp. 257–270.

Johannsen, G.; and Bille, J.: A Threshold Selection Method Using Information Measures. Proc. 6th Int. Conf. on Pattern Recognition, Munich, Germany, 1982, pp. 140–143.

Jolion, J-M.; Meer, P.; and Bataouche, S.: Robust Clustering with Applications in Computer Vision. IEEE PAMI , vol. 13, vol. 8, 1991, pp. 791–802.

Jumarie, G.: Contour Detection by Using Information Theory of Deterministic Functions. Pattern Recognition Letters, vol. 12, 1991, pp. 25–29.

Kapur, J. N.; Sahoo, P. K.; and Wong, A. K. C.: A New Method for Gray Level Picture Thresholding Using the Entropy of the Histogram. CVGIP, vol. 29, 1985, pp. 273–285.

Katz, Y. H.: Pattern Recognition of Meteorological Satellite Cloud Photography. Proc. Third Symp. on Remote Sensing of Environment, 1965, pp. 173–214.

Keeler, K.: Map Representations and Coding-Based Priors for Segmentation. CVPR, 1991, pp. 420–425.

Koplowitz, J.; and Lee, X.: Edge Detection with Subpixel Accuracy. SPIE, vol. 1471, 1991, pp. 452–463.

Lee, D.; and Wasilkowski, G. W.: Discontinuity Detection and Thresholding—A Stochastic Approach. CVPR, 1991, pp. 208–214.

Lee, C-H.: Recursive Region Splitting at Hierarchical Scope Views. CVGIP, vol. 33, 1986, pp. 237–258.

Lin, W.-C.; Tsao, E. C-K.; and Chen, C-T.: Constraint Satisfaction Neural Networks for Image Segmentation. Pattern Recognition, vol. 25, no. 7, 1992, pp. 679–693.

Liou, S-P.; Chiu, A. H.; and Jain, R. C.: A Parallel Technique for Signal-Level Perceptual Organization. IEEE PAMI, vol. 13, no. 4, 1991, pp. 317–325.

Liou, S-P.; and Jain, R. C.: An Approach to Three-Dimensional Image Segmentation. CVGIP: Image Understanding, vol. 53, no. 3, 1991, pp. 237–252.

Luijendijk, H.: Automatic Threshold Selection Using Histograms Based on the Count of 4-Connected Regions. Pattern Recognition Letters, vol. 12, 1991, pp. 219–228.

Mason, D.; Lauder, I. J.; Rutoritz, D.; and Spowart, G.: Measurement of C-Bands in Human Chromosomes. Comput. Biol. Med., vol. 5, 1975, pp. 179–201.

Matsuyama, T.: Expert Systems for Image Processing: Knowledge-Based Composition of Image Analysis Processes. CVGIP, vol. 48, 1989, pp. 22–49.

Mingsheng, Z.; Zhenkang, S.; and Huihang, C.: A New Edge Enhancement Method Based on the Deviation of the Local Image Grey Center. SPIE, vol. 1471, 1991, pp. 464–473.

Modestino, J. W.; and Zhang, J.: A Markov Random Field Model-Based Approach to Image Interpretation. IEEE PAMI, vol. 14, no. 6, 1992, pp. 606–615.

Mohan, R.; and Nevatia, R.: Perceptual Organization for Scene Segmentation and Description. IEEE PAMI, vol. 14, no. 6, 1992, pp. 616–635.

Monga, O.; Deriche, R.; and Rocchisani, J-M.: 3D Edge Detection Using Recursive Filtering: Application to Scanner Images. CVGIP: Image Understanding, vol. 53, no. 1, 1991, pp. 76–87.

Morii, F.: A Note on Minimum Error Thresholding. Pattern Recognition Letters, vol. 12, 1991, pp. 349–351.

Nakagawa, Y.; and Rosenfeld, A.: Some Experiments on Variable Thresholding. Pattern Recognition, vol. 11, 1979, pp. 191–204.

Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. on Systems, Man, and Cybernetics SMC, vol. 9, no. 1, 1979, pp. 62–66.

Pal, S. K.; and Ghosh, A.: Index of Area Coverage of Fuzzy Image Subsets and Object Extraction. Pattern Recognition Letters, vol. 11, 1990, pp. 831–841.

Pal, N. R.; and Pal, S. K.: Image model, Poisson Distribution and Object Extraction. Int. Journal of Pattern Recognition and Artificial Intelligence, vol. 5, 1991, pp. 459–483.

Panda, D. P.; and Rosenfeld, A.: Image Segmentation by Pixel Classification in (Gray Level, Edge Value) Space. IEEE Trans. on Computers, vol. C-27, no. 9, 1978, pp. 875–879.

Paranjape, R. B.; Morrow, W. M.; and Rangeyyan, R. M.: Adaptive-Neighborhood Histogram Equalization for Image Enhancement. CVGIP: Graphical Models and Image Processing, vol. 54, no. 3, 1992, pp. 259–267.

Parker, J. R.: Gray Level Thresholding in Badly Illuminated Images. IEEE PAMI, vol. 13, no. 8, 1991, pp. 813–819.

Pavlidis, T.; and Liow, Y-T.: Integrating Region Growing and Edge Detection. IEEE PAMI, vol. 12, no. 3, 1990, pp. 225–233.

Perkins, W. A.: Area Segmentation of Images Using Edge Points. IEEE PAMI PAMI, vol. 2, no. 1, 1980, pp. 8–15.

Petkovic, D.; and Wilder, J.: Machine Vision in the 1990s: Applications and How to Get There. Machine Vision and Applications, vol. 4, 1991, pp. 113–126.

Pun, T.: A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram. Signal Processing, vol. 2, 1980, pp. 223–237.

Pun, T.: Entropic Thresholding: A New Approach. CVGIP, vol. 16, 1981, pp. 210–239.

Reed, T.; and Wechsler, H.: Spatial/Spatial-Frequency Representations for Image Segmentation and Grouping. Image and Vision Computing, vol. 9, no. 3, 1991, pp. 175–193.

Rodriguez, A. A.; and Mitchell, O. R.: Image Segmentation by Successive Background Extraction. Pattern Recognition, vol. 24, no. 5, 1991, pp. 409–420.

Sadjadi, F. A.; and Bazakos, M.: A Perspective on Automatic Target Recognition Evaluation Technology. Optical Engineering, vol. 30, no. 2, 1991, pp. 141–146.

Sahoo, P. K.; Soltani, S.; and Wong, A. K. C.: A Survey of Thresholding Techniques. CVGIP, vol. 41, 1988, pp. 233–260.

Selim, S. Z.; and Alsultan, K.: A Simulated Annealing Algorithm for the Clustering Problem. Pattern Recognition, vol. 24, no. 10, 1991, pp. 1003–1008.

Shah, J.: Segmentation by Nonlinear Diffusion. CVPR, 1991, pp. 202–207.

Sher, C. A.; and Rosenfeld, A.: Pyramid Cluster Detection and Deliniation by Consensus. Pattern Recognition Letters, vol. 12, 1991, pp. 477–482.

Snyder, W.; Bilbro, G.; Logenthiran, A.; and Rajala, S.: Optimal Thresholding—A New Approach. Pattern Recognition Letters, vol. 11, 1990, pp. 803–810.

Spann, M.: Figure/Ground Separation Using Stochastic Pyramid Relinking. Pattern Recognition, vol. 24, no. 10, 1991, pp. 993–10002.

Spann, M.; and Wilson, R.: A Quad-Tree Approach to Image Segmentation which Combines Statistical and Spatial Information. Pattern Recognition, vol. 18, nos. 3 and 4, 1985, pp. 257–269.

Strasters, K. C.; and Gerbrands, J. J.: Three-Dimensional Image Segmentation Using a Split, Merge, and Group Approach. Pattern Recognition Letters, vol. 12, 1991, pp. 307–325.

Taxt, T.; Flynn, P. J.; and Jain, A. K.: Segmentation of Document Images. IEEE PAMI, vol. 11, no. 12, 1989, pp. 1322–1329.

Thurfjell, L.; Bengtsson, E.; and Nordin, B.: A New Three-Dimensional Connected Components Labeling Algorithm with Simultaneous Object Feature Extraction Capability. CVGIP: Graphical Models and Image Processing, vol. 54, no. 4, 1992, pp. 357–364.

Watanabe, S.; and the CYBEST Group: An automated Apparatus for Cancer Prescreening: CYBEST. CVGIP, vol. 3, 1974, pp. 350–358.

Westberg, L.: Hierarchical Contour-Based Segmentation of Dynamic Scenes. IEEE PAMI, vol. 14, no. 9, 1992, pp. 946–952.

Weszka, J. S.: Survey: A Survey of Threshold Selection Techniques. Computer Graphics and Image Processing, vol. 7, 1978, pp. 259–265.

Weszka, J. S.; and Rosenfeld, A.: Histogram Modification for Threshold Selection. IEEE Trans. Systems, Man, and Cybernetics SMC, vol. 9, 1979, pp. 38–51.

Weszka, J. S.; Nagel, R. N.; and Rosenfeld, A.: A Threshold Selection Techniques. IEEE Trans. Computers, vol. C-23, 1974, pp. 1322–1326.

Whatmough, R. J.: Automatic Threshold Selection from a Histogram Using the "Exponential Hull." CVGIP: Graphical Models and Image Processing, vol. 53, no. 6, 1991, pp. 592–600.

Won, C. S.; and Derin, H.: Unsupervised Segmentation of Noisy and Textured Images Using Markov Random Fields. CVGIP: Graphical Models and Image Processing, vol. 54, no. 4, 1992, pp. 308–328.

Yanowitz, S. D.; and Bruckstein, A. M.: A New Method for Image Segmentation. CVGIP, vol. 46, 1989, pp. 82–95.

Zhang, Y. J.; and Gerbrands, J. J.: Transition Region Determination Based Thresholding. Pattern Recognition Letters, vol. 12, 1991, pp. 13–23.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1993	Technical Memorandum

4. TITLE AND SUBTITLE
A Summary of Image Segmentation Techniques

5. FUNDING NUMBERS
506-59-31

6. AUTHOR(S)
Lilly Spirkovska

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Ames Research Center
Moffett Field, CA 94035-1000

8. PERFORMING ORGANIZATION REPORT NUMBER
A-93082

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
NASA TM-104022

11. SUPPLEMENTARY NOTES
Point of Contact: Lilly Spirkovska, Ames Research Center, MS 269-3, Moffett Field, CA 94035-1000
(415) 604-4234

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified — Unlimited
Subject Category 61

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Machine vision systems are often considered to be composed of two subsystems: low-level vision and high-level vision. Low-level vision consists primarily of image processing operations performed on the input image to produce another image with more favorable characteristics. These operations may yield images with reduced noise or cause certain features of the image to be emphasized (such as edges). High-level vision includes object recognition and, at the highest level, scene interpretation. The bridge between these two subsystems is the segmentation system. Through segmentation, the enhanced input image is mapped into a description involving regions with common features which can be used by the higher level vision tasks.

There is no theory on image segmentation. Instead, image segmentation techniques are basically ad hoc and differ mostly in the way they emphasize one or more of the desired properties of an ideal segmenter and in the way they balance and compromise one desired property against another. These techniques can be categorized in a number of different groups including local vs. global, parallel vs. sequential, contextual vs. non contextual, interactive vs. automatic. In this paper, we categorize the schemes into three main groups: pixel-based, edge-based, and region-based. Pixel-based segmentation schemes classify pixels based solely on their gray levels. Edge-based schemes first detect local discontinuities (edges) and then use that information to separate the image into regions. Finally, region-based schemes start with a seed pixel (or group of pixels) and then grow or split the seed until the original image is composed of only homogeneous regions.

Because there are a number of survey papers available, we will not discuss all segmentation schemes. Rather than a survey, we take the approach of a detailed overview. We focus only on the more common approaches in order to give the reader a flavor for the variety of techniques available yet present enough details to facilitate implementation and experimentation.

14. SUBJECT TERMS
Segmentation, Vision, Images

15. NUMBER OF PAGES
14

16. PRICE CODE
A02

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified		