Computer Vision and Image Understanding **77,** 1–24 (2000)
doi:10.1006/cviu.1999.0801, available online at http://www.idealibrary.com on IDEAL®

Segmentation and Interpretation of Multicolored Objects with Highlights

Bruce A. Maxwell

Department of Engineering, Swarthmore College, 500 College Ave., Swarthmore, Pennsylvania 19081
E-mail: maxwell@swarthmore.edu

and

Steven A. Shafer

Microsoft Research, 1 Microsoft Way, Redmond, Washington 98052

Received September 13, 1999; accepted September 13, 1999

This paper presents a segmentation system, based on a general framework for segmentation, that returns not only regions that correspond to coherent surfaces in an image, but also low-level interpretations of those regions' physical characteristics. This system is valid for images of piecewise uniform dielectric objects with highlights, moving it beyond the capabilities of previous physics-based segmentation algorithms which assume uniformly colored objects. This paper presents a summary of the complete system and focuses on two extensions of it that demonstrate its interpretive capacity and applicability to more complex scenes. The first extension provides interpretations of a scene by reasoning about the likelihood of different physical characteristics of simple image regions. The second extension allows the system to handle highlights within the general framework for segmentation. The resulting segmentations and interpretations more closely match our perceptions of objects since the resulting regions correspond to coherent surfaces, even when those surfaces have multiple colors and highlights.

1. INTRODUCTION

Images containing multicolored objects and/or multiple materials are difficult to interpret and segment intelligently. Simpler scenes containing uniformly colored objects of known material type can be segmented into regions corresponding to those objects using color and one or two physical models to account for color and intensity variations due to geometry and phenomena such as highlights [1, 4, 5]. These methods assume that a discontinuity in color between two regions of an image implies different surfaces.

1077-3142/00 $35.00

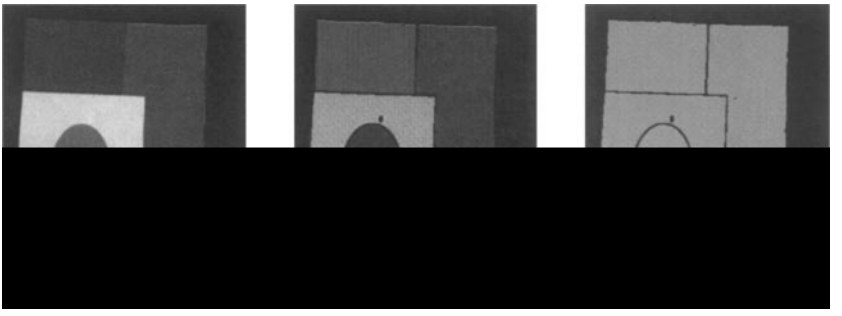

FIG. 1. (a) Several multicolored objects, (b) a segmentation based solely on color, and (c) a segmentation based on characteristics such as shape and material type. (Note: the black regions on the mug were too dark and fell below the background threshold.)

Multicolored objects, such as the ones shown in Fig. 1a, violate this assumption. Thus, segmentations based on color divide the objects into multiple small regions as in Fig. 1b, making the task of interpretation more difficult.

The change in color between two image regions, however, does not imply a discontinuity in shape, illumination, or other scene characteristics. To correctly group and interpret regions in scenes containing multicolored objects, multiple physical characteristics must be examined to determine whether two image regions of differing color are part of the same surface. By considering factors other than color we can achieve the segmentation shown in Fig. 1c, which groups the small multicolored regions of the cup's body, the plane, and the Lego into coherent surfaces [8, 9]. This segmentation corresponds more closely to our perception of objects than the segmentation based solely on color.

The challenge when segmenting images containing multicolored objects is that numerous interpretations can account for the same physical appearance. For example, taken by itself a red region on the mug could be a red dielectric under white light, a white dielectric under red light, or even metal reflecting a red light or a red object. These different interpretations of the region do not necessarily rule out its merger with adjacent regions. The different interpretations do, however, require different methods of analysis. Previously we presented a framework that proposed using reasoning about multiple individual region interpretations and physics-based measures of region similarity to achieve a segmentation such as Fig. 1c [7]. Initial results on real images using this framework were presented in [6]. This initial instantiation of the segmentation framework, however, was limited to objects without highlights. Furthermore, it did not take advantage of the inherent consideration of multiple region interpretations to provide a higher-level interpretation of the coherent surfaces in the image.

This paper addresses both of these issues and presents a more complete system that not only provides intelligent segmentations of more complex images, but also proposes interpretations of the segmentations it provides. Herein, we first present a review of the segmentation framework using multiple hypotheses for simple image regions. This review provides refinements and details of the segmentation system not previously presented in [8] or [9]. Next, we show how we can use these multiple hypotheses to prefer certain explanations for certain image regions. This moves the system from a pure segmentation algorithm to the realm of image interpretation. Finally, we show how we can deal with more complex scenes within the same interpretation framework by adding additional hypotheses

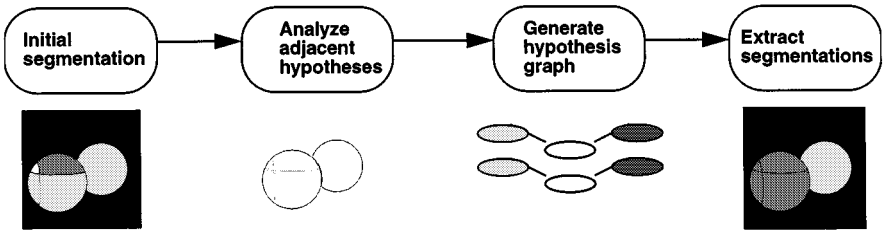

FIG. 2. Algorithm outline: initial segmentation, analysis of adjacent region compatibility, hypothesis graph generation, final segmentation extraction.

to the list of explanations considered for each simple image region. This expanded system can not only segment scenes into coherent surfaces when they contain multicolored objects containing highlights but also provide a simple interpretation of those surfaces in terms of their physical characteristics.

2. SEGMENTATION FRAMEWORK

Our approach to segmentation divides into four parts as shown in Fig. 2: an initial segmentation based on normalized color, compatibility analysis of adjacent hypotheses, generation of the hypothesis graph, and extraction of the final segmentations. The foundational concepts and segmentation framework were presented in [9]. This section summarizes the first two steps and details the implementation and concepts involved in the latter two.

2.1. Initial Segmentation

The first task is to find simple regions in the image. The objective of this step is to find groups of adjacent pixels that can reasonably be assumed to belong to a single object. We use a region-growing algorithm with normalized color as the feature of interest. This algorithm uses local information to grow the regions and stops growing when it reaches discontinuities in the normalized color.

Once the algorithm has identified the simple image regions it attaches the list of initial hypotheses to each one. We describe a hypothesis as an ordered tuple H = (<material type>, <illumination type>, <shape type>). See [9] for a detailed description of the different classes for each scene element. The implementation described in [8] used the hypothesis set Hc = {(Colored dielectric, White uniform, Curved), (Colored dielectric, White uniform, Planar)} for colored regions and the hypothesis set Hw = {(White dielectric, White uniform, Curved), (White dielectric, White uniform, Planar)} for white and grey regions. White uniform illumination is defined as illumination from one or more uniform spectrum area or point sources, with some level of uniform spectrum ambient illumination. The complexity of the initial hypothesis lists determines the complexity of the images for which the system will return a valid response.

2.2. Hypothesis Analysis

The second task of the algorithm is to compare potentially compatible adjacent hypotheses. This step requires analysis of the compatibility of the individual hypothesis elements:

the material type, the illumination, and the shape. To make this determination, we turn to a physical analysis of the problem.

To determine hypothesis compatibility we identified physical attributes of an object's appearance that have predictable relationships between different regions of the same object. By looking for these relationships we can differentiate between regions that may be part of the same object and those that are not. We use local physical characteristics that are more appropriate for region-based analysis than global scene analysis methods such as shape-from-shading. The three characteristics we use to determine compatibility are reflectance ratio constancy [11], gradient direction continuity, and intensity profile continuity. Breton and Zucker have also proposed the first two characteristics as theoretically useful physical measures of surface continuity for dielectric objects [2] but have not implemented them within an image analysis algorithm. It is important to note that these compatibility measures are appropriate for the hypotheses listed in the initial hypothesis sets above. Other hypotheses may require different tests of compatibility.

2.2.1. Reflectance Ratio

In order to understand how we use the reflectance ratio, consider two adjacent hypotheses h_1 and h_2 that both specify colored dielectrics under white illumination. If h_1 and h_2 are part of the same piecewise uniform object and have a different color, then the discontinuity at the border must be due to a change in the transfer function, and this change must be constant along the border between the two regions. Furthermore, along the border the two regions must share similar shape and illumination. If h_1 and h_2 belong to different objects, than the shape and illumination do not have to be the same.

The reflectance ratio is a measure of the change in transfer function between two pixels that is invariant to illumination and shape if the latter two elements are similar at each pixel. If the shape and illumination of two pixels p_1 and p_2 are similar, then the reflectance ratio, defined in Eq. (1), where I_1 and I_2 are the intensity values of pixels p_1 and p_2, reflects the change in albedo between the two pixels [9]:

$$r = \left(\frac{I_1 - I_2}{I_1 + I_2} \right). \tag{1}$$

For each border pixel p_{1i} in h_1 that borders on h_2 we find the nearest pixel p_{2i} in h_2. If the regions belong to the same object, the reflectance ratio should be the same for all pixel pairs (p_{1i}, p_{2i}) along the h_1, h_2 border. A simple measure of constancy is the variance of the reflectance ratio. If h_1 and h_2 are part of the same object, this variance should be small, due mostly to the quantization of pixels, noise in the image, and small-scale texture in the scene.

If, however, h_1 and h_2 are not part of the same object, then the illumination and shape are not guaranteed to be similar for each pixel pair, violating the specified conditions for the characteristic. Differing shape and illumination should result in larger variance in the reflectance ratio. We can select an expected variance based upon the noise, variance in the object's transfer functions, and quantization effects and use this expected variance to differentiate between these two cases. Table 1 shows the variances in the border reflectance ratios of the region pairs for the test image of the stop sign and cup (shown in Fig. 11a).

We then use a χ^2 test to compare the variance for a particular region pair to the expected variance. The result of the χ^2 test is the likelihood that the variance in the reflectance ratio

TABLE 1
Results of Reflectance Ratio Test for the Stop-Sign and Cup Image with $\Sigma = 0.004$

Region 1	Region 2	μ	σ^2	$P(\sigma^2 < \Sigma^2)$
Sign	Letter S	0.4463	0.0004	1.0
Sign	Letter T	0.4449	0.0005	1.0
Sign	Letter O	0.4503	0.0004	1.0
Sign	Letter P	0.4541	0.0006	1.0
Sign	Cup	0.2107	0.0125	0.0
Sign	Pole	0.1709	0.0710	0.0
Letter O	O hole	−0.4358	0.0008	1.0
Letter P	P hole	−0.4562	0.0004	1.0

along the border is not caused by differing shape and illumination. While this test does not directly compare the shape and illumination of the two regions, the variance of the reflectance ratio along the border does implicitly measure their similarity. The reflectance ratio can be used to compare all of the hypothesis pairs in our current implementation.

2.2.2. Gradient Direction

The direction of the gradient of image intensity can be used in a manner similar to that for the reflectance ratio. The direction of the gradient is invariant to the transfer function for piecewise uniform dielectric objects—except due to border effects at region boundaries. Therefore, by comparing the gradient directions of border pixel pairs for adjacent regions we obtain an estimate of the similarity of the shape and illumination. To avoid border effects, the algorithm first calculates the gradient directions of nonborder pixels and then grows the results outwards.

As with the reflectance ratio, we sum the squared difference of the gradient directions of adjacent border pixels to determine a sample variance for each hypothesis pair. We then use the χ^2 test to compare the sample variance to an expected variance chosen based on numerous test images. Because of the conditions required for the gradient directions of adjacent borders to be similar, we interpret the result as a likelihood that the illumination and shape are similar along the border of the two regions.

Not surprisingly, the effectiveness of this characteristic is limited to regions with well-defined gradient directions. For planar or almost uniform surfaces with small gradients the angle of the gradient is very sensitive to noise and quantization errors. To avoid using noisy direction values, we only use border pixel pairs whose gradient intensities are above a selected threshold. Furthermore, for each hypothesis pair we require that a sufficient percentage of the applicable border pixels be above the gradient intensity threshold or we consider the results inconclusive and do not use them.

Like the reflectance ratio, the gradient direction can be used to compare all of the current hypothesis pairs.

2.2.3. Intensity Profile Analysis

So far, we have examined only calculated characteristics of the image, not the actual image intensities. The intensity profiles contain a significant amount of information, however,

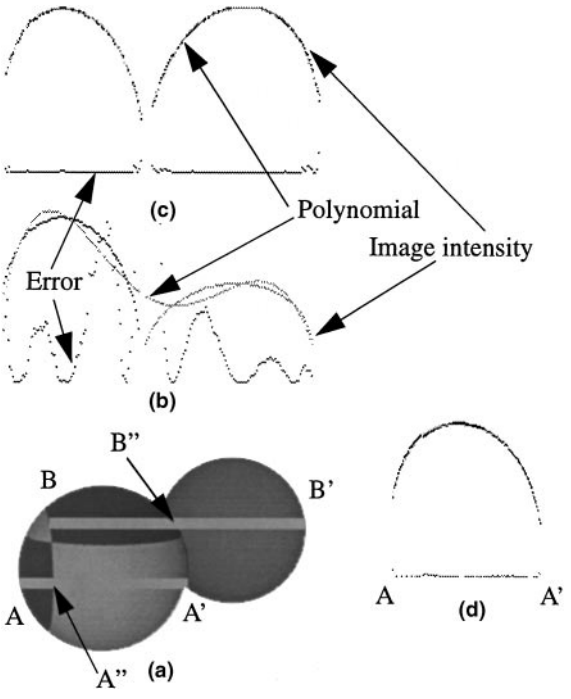

FIG. 3. (a) Test image, (b) fitting one polynomial to B-B', (c) fitting two polynomials to B-B', (d) fitting one polynomial to A-A'. In (b), (c), and (d), the bottom line shows the error in the fit, while the top two lines show the image intensity overlaid by the least-squares polynomial.

which we attempt to exploit with the following assertion: if two hypotheses are part of the same object and the illumination and shape match at the boundary of the hypotheses, then, if the scale change due to the albedo difference is taken into account, the intensity profile along a scanline crossing both hypotheses should be continuous. Furthermore, we should be able to effectively represent the intensity profile across both regions with a single model. If two hypotheses are not part of the same object, however, then the intensity profile along a scanline containing both hypotheses is more likely to be discontinuous and representing the profile with two models should be more appropriate.

To demonstrate this property, consider Fig. 3d, which shows the intensity profile for the scanline from A to A'. We can calculate the average reflectance ratio along the border to obtain the change in albedo between the two image regions. By multiplying the intensities from A" to A' by the average reflectance ratio we adjust for the difference in albedo. As a result, for this particular case the intensity profile becomes smooth and a single model is a good representation. On the other hand, for the scanline B to B', the scaled curves are disjoint, and two models are a better representation as shown in Figs. 3b and 3c.

Rather than use the first or second derivatives of the image intensities to find discontinuities, we take a more general approach which maximizes the amount of information used and is not as sensitive to noise or small-scale texture in the image. Our method is based upon the idea that if two hypotheses are part of the same object then it should require less information to represent the intensity profile for both regions with a single model than to represent the regions individually. We use the Minimum Description Length [MDL], as defined by Rissanen [14], to measure complexity, and we use polynomials of up to order

5 to approximate the intensity profiles. The formula we use to calculate the description length of a polynomial model, where x^n is the data, θ is the set of model parameters, k is the number of model parameters, and n is the number of data points [12], is given as

$$DL = -\log P(x^n \mid \theta) + \frac{k}{2}\log n. \tag{2}$$

Our method for a single scanline s_0 is as follows:

1. Model the intensity profile on scanline s_0 for hypothesis h_1 as a polynomial. Use the MDL principle to find the best order polynomial (we stop looking after order 5). Assign M_a the minimum description length.

2. Model the intensity profile on scanline s_0 for hypothesis h_2 as a polynomial. Again, use the MDL principle to find the best order and assign its MDL to M_b.

3. Model the scaled intensity profiles of scanline s_0 for both h_1 and h_2 as a polynomial, find the best order using MDL, and assign the smallest MDL to M_c.

4. Compare $(M_a + M_b)$ to M_c. To normalize the results of this test to the range [0, 1], we use the measure of merit given by (3),

$$L_m = 1 - \frac{M_c - (M_a + M_b)}{M_c + (M_a + M_b)}, \tag{3}$$

and any result >1.0 gets set to 1.0.

To obtain a robust measure for a region pair, we average the result of this procedure over all scanlines containing a border pixel, looking either vertically or horizontally depending upon the local border tangent. We then compare this average to the median likelihood and take the more extreme value (towards 0 or 1 depending on whether the average is less than or greater than 0.5, respectively). For more discussion of the profile analysis, see [10].

2.3. Creating the Hypothesis Graph

The system organizes the information it obtains in the analysis stage by generating a graph representation of the space of possible segmentations. This allows both local and global reasoning about how the hypotheses fit together. It is important to realize that the analysis stage is only a local analysis, focusing on a single pair of hypotheses at a time. What we are looking for, however, is a good global solution which maximizes some function of the local results. The hypothesis graph allows us to search the set of possible segmentations and find good solutions.

The structure of the hypothesis graph is as follows. Each hypothesis in the image forms a node, and edges connect all hypotheses that are adjacent in the image as shown in Fig. 4. If two adjacent hypotheses cannot merge, then a single discontinuity edge connects them as between hypotheses A1 and B2 in Fig. 4. Alternatively, if the adjacent hypotheses potentially can merge, then two edges connect the corresponding pair of nodes as between the hypotheses A1 and B1 in Fig. 4. The dashed edge is a discontinuity edge, indicating the cost of not merging the two hypotheses. The solid edge is a merge edge, whose value is based upon the results of the pairwise local analysis. All edges have values in the range [0, 1].

Given a complete hypothesis graph of this form, the set of possible segmentations of the image is the set of subgraphs such that each subgraph includes exactly one hypothesis, or node from each region and exactly one edge connecting each adjacent pair of hypotheses.

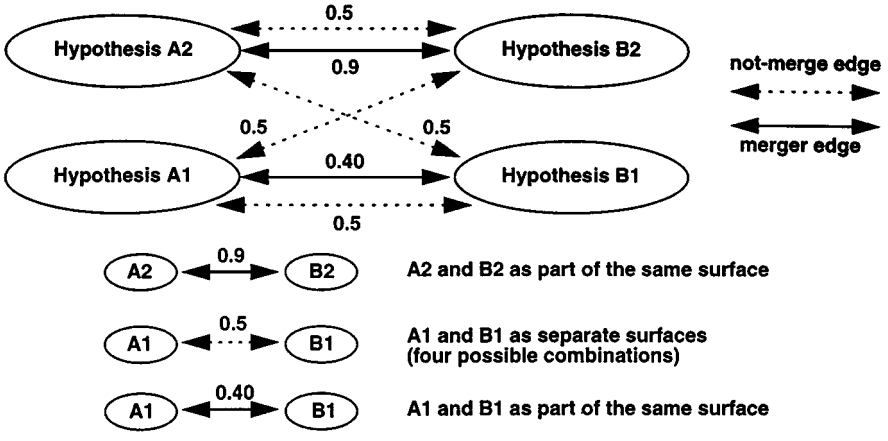

FIG. 4. Example hypothesis graph and the set of valid segmentations ranked according to the sum of the edge values in each segmentation.

To find a segmentation, therefore, requires the selection not only of a single hypothesis to represent each region, but also the type of edges connecting each hypothesis to its neighbors. For example, one valid segmentation of the hypothesis graph in Fig. 4 would be hypotheses A2 and B2 connected by a merge edge. This hypothesis graph contains six valid segmentations in all, one for each edge in the graph.

Assigning weights to the merge edges is straightforward; the weight of a merge edge is a weighted average of the results of the compatibility tests. Given that these tests return a value closer to 1.0 for compatible surfaces, good segmentations will be those that maximize the values of the edges in the final segmentation. The not-merge edges receive a value of 0.5. For more insight on this determination see [9]. Figure 4 shows a hypothesis graph and its valid segmentations in order of their likelihood.

Note that if we have a rank ordering of the hypotheses for a region we can adjust the edge values to reflect the ordering. Such a rank ordering must be based upon a comparison of each hypothesis to its region's image data. For example, we might prefer a hypothesis proposing a planar surface over a hypothesis proposing a curved surface for a region of uniform intensity. However, in the interest of maintaining the tuning between the discontinuity and merge edges, such adjustments must be small and the values should remain close to the original values. This modification is the topic of Section 3, which shows how comparing the hypotheses to the region data results in greater interpretive capacity.

The hypothesis graph for the stop sign and cup image is shown in Fig. 5 without the discontinuity edges for clarity. Weak merge edges connect the pole and the sign, and the cup and the sign, indicating they are likely different surfaces. Strong merge edges connect the sign and its letters, indicating they are likely to be part of the same surface. The next section describes how we can automatically extract good segmentations of the image given the hypothesis graph.

2.4. Extracting Segmentations

Before an approach is selected to extracting segmentations from the hypothesis graph, it is informative to determine the extent of the space of segmentations. If it is computationally easy to search the space of possible segmentations, then we can always find the optimal, or

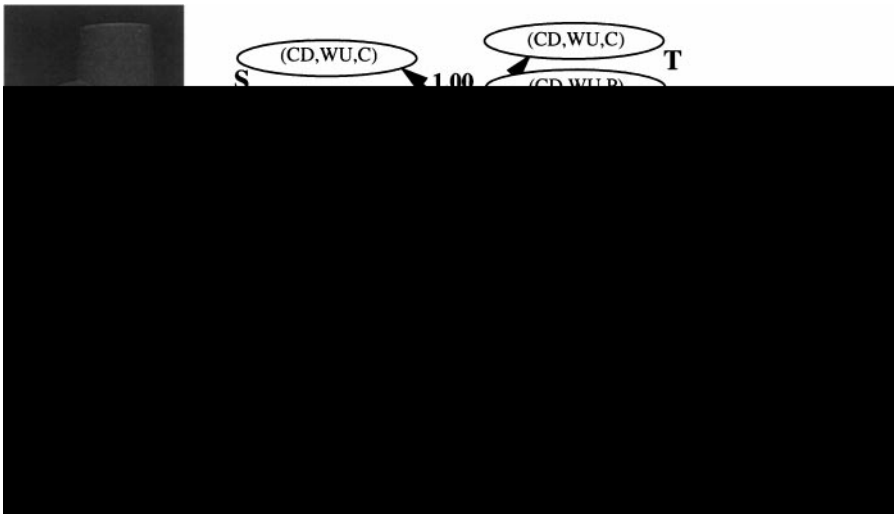

FIG. 5. Hypothesis graph for the stop-sign and cup image. For clarity, only the merge edges and their values are shown.

most likely segmentation of the image. If, on the other hand, the space is too large, then we must turn to heuristic methods to find a solution.

2.4.1. Characterizing the Space of Segmentations

The most complex, or worst case hypothesis graph is one where any combination of hypotheses is a legal and consistent segmentation. The set of hypothesis graphs where any combination of hypotheses is legal is the set of graphs containing no loops.

Figure 5 is an example of a worst case graph because it contains no loops and any combination of hypotheses is possible. Given that the stop-sign and cup image is not an unusual image, examining the worst-case graph is a reasonable way to understand the complexity of the segmentation space. The worst-case scenario also gives an upper bound on the size of the segmentation space in the more general case of graphs with or without loops. The following theorem gives the upper bound on the size of the space of segmentations.

THEOREM 1. *The number of segmentations S for a graph without loops is given by*

$$S = NE^{R-1} \tag{4}$$

where N is the number of hypotheses per region, E is the number of edges connecting a single hypothesis to all of the hypotheses in an adjacent region, and R is the total number of regions in the graph.

Proof. We can use induction to show Theorem 1. First, consider a hypothesis graph with a single region. As there are N hypotheses per region, there are N possible segmentations for this graph.

Now consider a hypothesis graph containing R regions $Q_1 \ldots Q_R$ and no loops. Let the number of segmentations contained in this graph be S_R. If we add a new region Q_{R+1} as a neighbor to any single existing region Q_i, then each of the S_R segmentations of the R regions

attaches to E hypotheses in region Q_{R+1}. This multiplies S_R by E, giving $S_{R+1} = S_R * E$ as the number of distinct segmentations of the new graph with $R + 1$ regions.

By combining these two observations we get (4) as the general formula for the number of possible segmentations contained in a hypothesis graph with no loops. ■

What Theorem 1 tells us is that testing all possible segmentations of an image is undesirable even for fairly simple images. It is beyond reasonable current technology to examine all possible segmentations of images with more than 30–35 initial regions. As a reference, the image in Fig. 1, which contains only 3 objects, has 28 initial regions.

2.4.2. Review of Clustering Techniques

Unable to test all possible cases, we must turn to heuristic methods to extract segmentations from the hypothesis graph. Agglomerative clustering, or region merging, is a common technique that resembles extracting segmentations from the hypothesis graph. In both cases we search for the best groupings of pixels, regions, or nodes.

Region-merging algorithms calculate distance metrics between neighboring elements and then group similar elements using an iterative process. A number of researchers use clustering techniques for image segmentation.

The most relevant recent work is by LeValle and Hutchinson [7]. They developed a stepwise optimal algorithm, which they call the highest-probability-first algorithm, to extract segmentations from a single-layer graph of nodes and probabilities. Both they and Panjwani and Healey have used it to segment images containing texture [7, 13]. At each step the algorithm merges the most likely two nodes until it reaches a threshold based on the number of regions or the likelihood of the best merge.

Roughly equivalent to the highest-probability-first algorithm is the use of dendograms. As described by Duda and Hart [3], a dendogram is a tree data structure used for unsupervised, hierarchical clustering. The leaves of the tree are the individual regions, and at each level two regions merge to form a new node. Each layer of the dendogram is a different clustering, or segmentation, of the data. Which two regions merge at each layer depends upon the task. A common method is to merge the two nodes which most belong together, which is the same as the highest-probability-first algorithm. Which layer to choose as the final segmentation is the equivalent of deciding when to stop merging with the highest-probability-first algorithm. Both Wallace [16] and Krumm [6] used dendograms for the segmentation of data, Krumm for the segmentation of images with texture. Note that neither the highest-probability-first algorithm nor the dendogram approach is guaranteed to encounter the globally optimal solution.

2.4.3. Modified Highest-Probability First Algorithm

Our segmentation extraction algorithm is basically the highest-probability-first algorithm modified for the multilayer hypothesis graphs. Although it may be a direction of future research, we do not maintain the different layers during the segmentation as with the dendogram approach. Figure 6 gives a graphic demonstration of the modified highest-probability-first algorithm for the case of three regions all adjacent to one another.

The segmentation extraction algorithm begins with an initialization step that sets up two lists: one contains all of the initial hypotheses, the other contains all of the edges in the initial hypothesis graph sorted in descending order. This step also initializes the

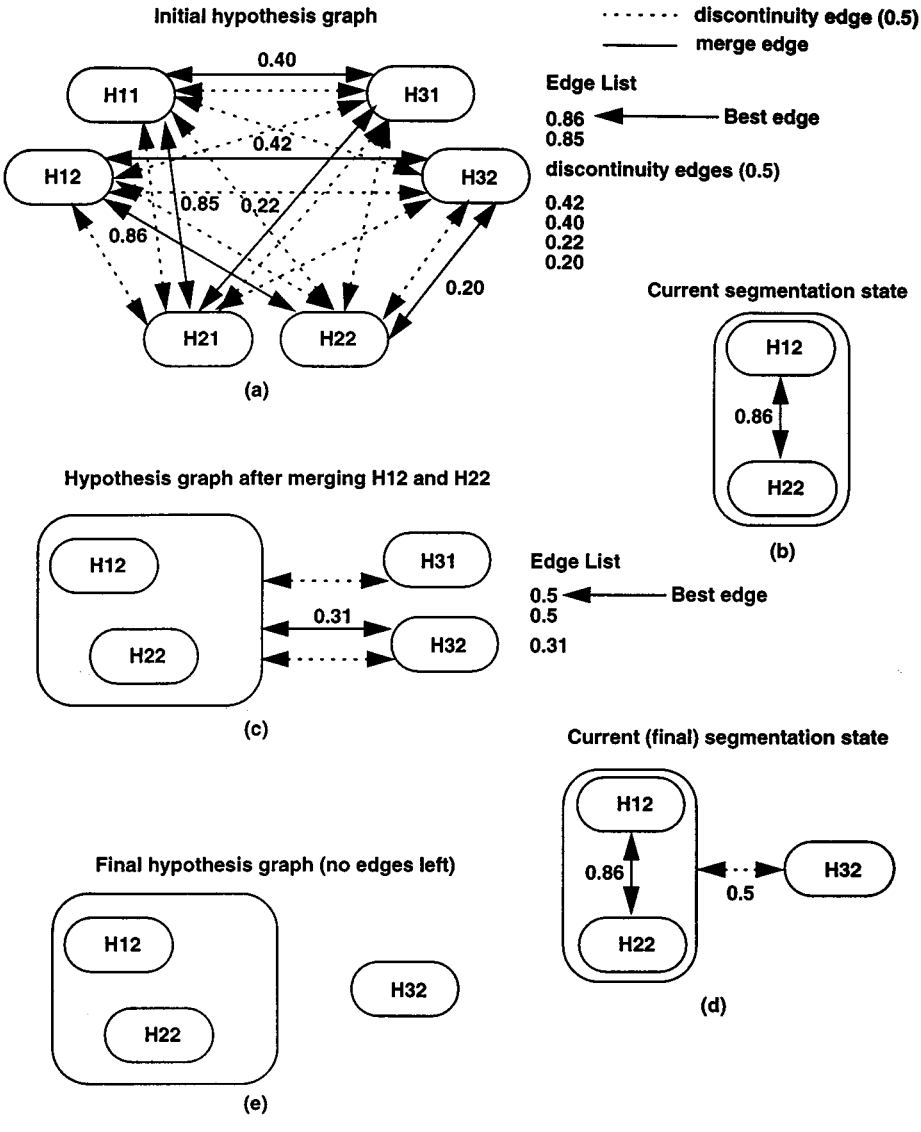

FIG. 6. (a) Initial hypothesis graph, (b) segmentation state after merging H12 and H22, (c) hypothesis graph after merging H12 and H22 and updating the edges (d) segmentation state after adding H32 and the discontinuity edge, (e) final hypothesis graph state. After (e) there are no edges left so the algorithm terminates.

segmentation map. This map contains one hypothesis opening for each initial image region and one edge opening for each adjacent hypothesis pair. These openings are initially empty. Figure 6a shows the initial hypothesis graph and edge list.

The main loop begins by taking the best edge from the edge list. In Fig. 6a the best edge is a merge edge connecting H12 and H22 with a value of 0.86. If either or both of the hypotheses connected by the edge are not already in the segmentation map, then the algorithm puts them there. It also places the selected edge into its appropriate position in the segmentation map. Figure 6b shows the state of the segmentation map after this step.

Note that after placing one of the hypotheses for a region into the segmentation map the algorithm must remove all other hypotheses for that region from the hypothesis list.

Furthermore, it must remove all edges associated with those alternative hypotheses from the edge list. This happens whether or not the edge indicates a merge or a discontinuity. For example, in Fig. 6 we see that the hypothesis graph no longer contains H11 and H21 as alternative explanations, for those regions are already in the current segmentation state.

In addition to the alternative hypotheses, if the selected edge is a merge edge, then the algorithm removes the individual hypotheses connected by that edge from the hypothesis list and replaces them with a single aggregate hypothesis. It then removes all edges connected to the individual hypotheses, recalculates the edge values between the aggregate hypothesis and its neighbors, and places the new edges into the edge list.

For example, in Fig. 6 the algorithm selects the merge edge connecting H12 and H22. It then removes H12 and H22 from the graph, as well as the merge and discontinuity edges connecting H31 and H32 to H12 and H22. The algorithm also removes H11 and H21 and all of their connecting edges from the graph, as they are alternative explanations for the regions represented by H12 and H22. After these removals, the algorithm replaces H12 and H22 with a single aggregate region containing both and recalculates all of the edges connecting the new aggregate region to its neighbors. Figure 6c shows the state of the hypothesis graph after H12 and H22 are merged.

If, on the other hand, the edge is a discontinuity edge, then the algorithm leaves the individual hypotheses in the hypothesis list and only removes the discontinuity edge placed in the segmentation and any other edges connecting the same two hypotheses. In Fig. 6c, for example, the best edge is one of the two discontinuity edges connecting the aggregate hypothesis with H31 or H32. In this case we select the discontinuity edge connecting H32 with the aggregate hypothesis. Thus, the algorithm places H32 and the not-merge edge into the segmentation and removes the alternative merge edge. Figure 6d shows the state of the segmentation after this second pass.

One important side note is that we use averaging to combine edges together. To get the value of the merge edge between the aggregate region and H31 and H32 in Fig. 6c, for example, the algorithm averages the values of the merge edges connecting the aggregate region's component hypotheses to H32, in this case giving a value of 0.31. What is important to note is that this can alter the relationship between regions. For example, if a 0.60 edge connected H12 and H32, indicating a likely merge, the average value between H12 and H3 after the merge would be 0.4, indicating a likely discontinuity. In this way, hypothesis pairs with strong merge or discontinuity values can influence more ambivalent hypothesis pairs.

This set of actions repeats until the edge list is empty. At that point the segmentation map may or may not be complete. In Fig. 6d the segmentation is complete and, as shown in Fig. 6e, there are no edges left in the hypothesis graph, terminating the algorithm. An isolated region in an image, however, will have no neighbors and, therefore, no edges connected to its hypotheses. If the segmentation is incomplete, the algorithm completes it by randomly choosing one hypothesis from each isolated region's hypothesis list. Note that if the algorithm is given a rank ordering for the hypotheses, then it could use that to make a selection instead of randomly choosing.

Running the algorithm once on the entire hypothesis graph returns a single segmentation. The sum of the values of all of the edges contained in the segmentation measures its quality. In most cases, this segmentation should be good. Given that the segmentation extraction algorithm is not globally optimal, however, this segmentation may or may not be the best, where the best segmentation is defined as the segmentation with the maximum possible sum of the edge values within it. Furthermore, this segmentation does not include a large percentage of the possible hypotheses.

2.4.4. Generating a Set of Representative Segmentations

What we would like to do is extract a set of segmentations that provides a representative sample of the space of good segmentations. We also would like to have each hypothesis represented in at least one segmentation within this set. This guarantees that our set of good segmentations contains at least one data point for each interpretation of each region. The hope is that the best interpretation of a region will produce better global segmentations.

Our solution is to run the extraction algorithm on N different graphs, where N is the number of hypotheses in the image. For each hypothesis $h \in H$, we set up a complete hypothesis graph, except that we modify it by only assigning the single hypothesis h to the region it represents. This forces each hypothesis to be included in at least one of the set of N segmentations. This set of segmentations may contain up to R duplicates, where R is the number of regions. This is because we can get the same segmentation from any two modified graphs so long as the fixed hypothesis is one of the R hypotheses that would be chosen when using the complete hypothesis graph.

If we remove the duplicates and rank-order the segmentations according to the sum of the edge values, we end up with a rank-ordered list where each interpretation of each region is contained in at least one of the segmentations. Note, if all discontinuous region pairs have a likelihood of 0.5; there will almost always be multiple equally likely segmentations with the same grouping of regions but different hypotheses for the individual groups. Figure 7 shows the best region grouping for the stop-sign and cup image.

To summarize, the output of the system is a set of rank-ordered segmentations, where each segmentation specifies a hypothesis for each region and how those hypotheses group together. For most of the results shown here we will show only the best segmentation for each image. It is important to remember, however, that the complete output of the algorithm is a rank-ordered set.

2.4.5. Avoiding and Getting out of Local Minima

The biggest problem with the highest-probability-first algorithm is that it tends to get stuck in local minima. This problem happens most often in images with loops, such as the image of two spheres. For example, if the best edge in pass one connects the curved hypotheses for the left two regions of the left-most sphere, then the algorithm places those

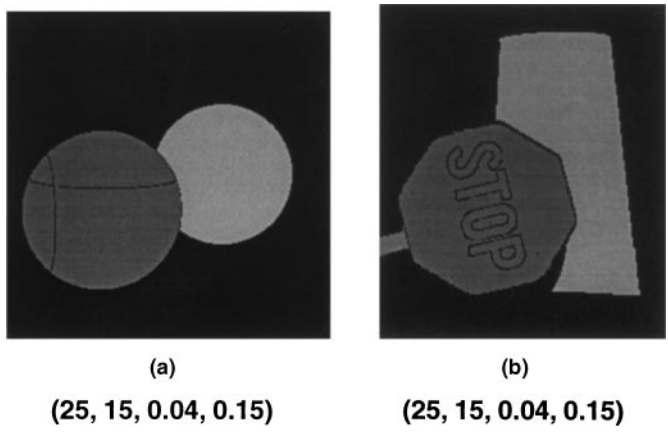

FIG. 7. Best region groupings for (a) the two-sphere image, and (b) the stop-sign and cup image.

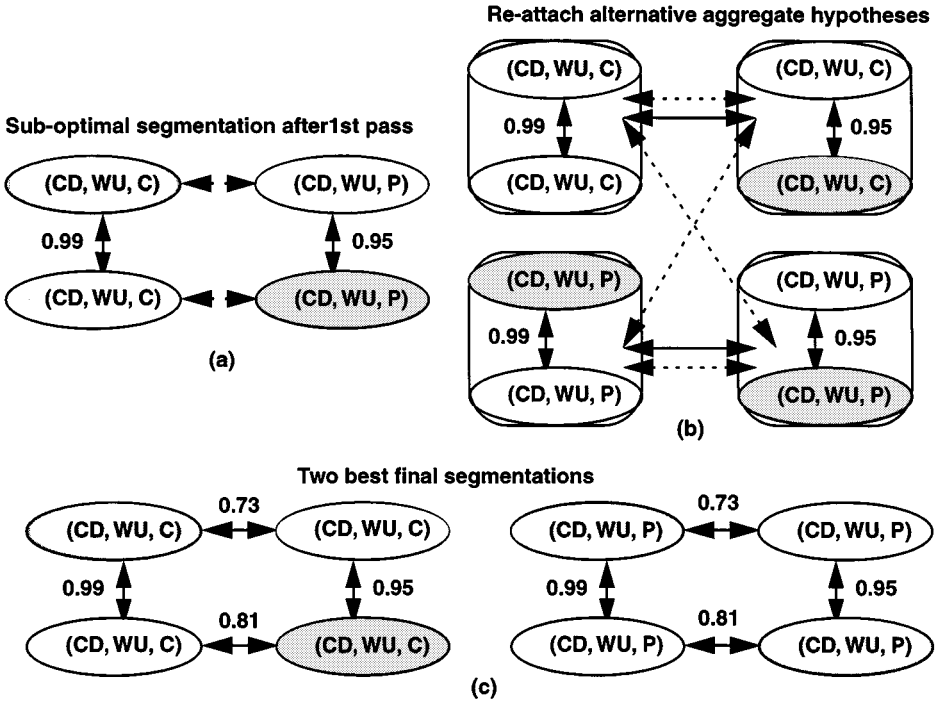

FIG. 8. (a) Suboptimal segmentation, (b) hypothesis graph after reattaching alternative aggregate hypotheses, (c) two best segmentations given (b).

two hypotheses in the segmentation map. If, however, in pass two the best edge connects the two planar hypotheses for the right two regions of the sphere, then the algorithm places those two hypotheses in the segmentation map. The only edges between the planar and curved hypotheses, however, are discontinuous edges. This produces a less than optimal segmentation, as shown in Fig. 8a.

We use two heuristics to avoid or get out of local minima. The first heuristic is that if there are several edges at the top of the edge list with close to the same likelihood, we pick the edge that connects hypotheses that are most like the mandatory hypothesis in the graph. In the case of the two-spheres image, if the upper left green region must be (Colored Dielectric, White Uniform Illumination, Curved), then we prefer the curved hypotheses for other regions as well so long as the edge values are within ε of one another. Alternatively, if the single hypothesis region must be planar, then we prefer the planar hypotheses.

This heuristic works so long as the likelihoods of the different hypothesis pairs have similar values and we want all of the surfaces in an image to be like the mandatory hypothesis. However, it does not work well if there are differences in the likelihoods of the hypothesis pairs for the same two regions. In this case, we need a different approach as the first heuristic only tries to avoid local minima, not get out of them.

To get out of local minima somehow we have to perturb things and restart the search process, which motivates the second heuristic. First, run the algorithm to get an initial segmentation, or set of hypothesis groupings. These hypothesis groupings are also region groupings and tend to contain most of the strong edges in the hypothesis graph. The heuristic is based upon the observation that there are other, almost as likely interpretations for the aggregate regions. For example, in the two-spheres image the interpretation of the left

sphere as four hypotheses proposing a curved surface has a likelihood similar to that for the interpretation of it as four hypotheses proposing a planar surface. However, any mixture of the planar and curved hypotheses is much less likely. In Fig. 8a, for example, there is a planar interpretation of the left side of the sphere that is equally as likely as the curved interpretation. The same situation exists for the right side of the sphere. If we reattach these alternative aggregate interpretations, then we have the new hypothesis graph shown in Fig. 8b. This perturbs the algorithm out of the local minimum and gives it larger groupings of hypotheses to work with. Now, when we run the modified highest-probability-first algorithm again, we get one of the final segmentations shown in Fig. 8c. Both of these segmentations are better than the first segmentation shown in Fig. 8c.

It turns out that a third pass is superfluous for these cases. For all of the test images, the second pass found the optimal segmentation in the judgement of human observers.

This heuristic works for the two-layer hypothesis graphs because there will always be two possible interpretations for each region grouping, except the region grouping containing the mandatory hypothesis for that graph. When we begin to add more hypotheses per region, however, there will be more interpretations for the aggregate regions. In this case, we will need to enumerate the possible alternative interpretations for each aggregate region. In the general case, a perturbation method could be used in an iterative loop until a measure of convergence was achieved (e.g., the resulting segmentation did not change).

It is interesting to note the approximate run-time of the modified highest-probability-first algorithm in comparison to the complexity of the segmentation space. The complexity depends on the number of regions R, the number of hypotheses per region N_r, and the total number of edges in the hypothesis graph E. These values are related by the expression $E = k_e R N_r$, where k_e is one-half the average number of regions adjacent to each region.

If we assume the edge list is sorted on every pass through the main loop of the algorithm, then a single loop takes $E \log E$ time to sort, linear time to update the segmentation map, and linear time to create the new edges. The algorithm goes through the main loop once for each adjacent region pair ($\propto E$) in order to fill the segmentation map, bounding the run-time by $O(E^2 \log E)$. If we run the extraction algorithm for each of the RN_r hypotheses, the total run time to extract the representative set of good segmentations is $O(RN_r E^2 \log E)$. Since RN_r is proportional to E, the complexity is a cubic polynomial in the number of edges in the graph. This is significantly better than having to search the entire space of segmentations, and it guarantees that each hypothesis is considered in at least one interpretation of the image.

3. PREFERRING HYPOTHESES

The output of the system described so far is a set of segmentations which specify both a region grouping for the image and the possible high-level explanations for each aggregate region. However, the quality of a segmentation is based only on how well the individual regions match. What this implies is that a segmentation that groups regions that are likely to be part of the same surface will be preferred to a segmentation that does not. However, the basic system will not prefer one explanation to another if they are based upon the same tests of compatibility.

As the two hypotheses considered by the previous implementation in [8] differ only in their geometry, the same analysis tools apply to both hypotheses. Therefore, while that implementation does prefer certain segmentations, it does not prefer one interpretation over another so long as they provide the same region groupings. This paper presents a new

extension to the basic framework that allows the system to prefer certain hypotheses over others based on the image data.

Certain hypotheses require "weirder" or more improbable components to describe a specific image region. For example, for a planar surface to generate the shading in the mug image in Fig. 1a, either the transfer function must change intensity over the surface, the illumination must be extremely complex, or some combination of the two must occur. However, for a curved surface to generate the mug image requires only a piecewise uniform object and a single light source. If we use the principle of simplicity to guide our analysis, then we should prefer curved hypotheses over planar hypotheses for the body of the mug because that allows for a simpler description.

It is true that we cannot rank hypotheses in the absence of other information except by reasoning about their likelihood in the world. However, once we have attached a set of hypotheses to a specific image region, we can talk about the complexity of the hypothesis elements required to explain that particular region.

To implement this concept, we focus on ranking the curved and planar hypotheses specifying dielectrics under uniform illumination. What this requires is a method of comparing the likelihood that a particular image region was generated by a planar surface to the likelihood that it was generated by a curved surface. To develop this method we would like to make as few assumptions as possible about the image and the scene.

There are three major issues to solve in order to implement a method of ranking hypotheses. The first is how to determine the likelihood that a particular hypothesis generated a particular image region. In this presentation we focus on the likelihood that a particular region was generated by a planar surface. If this likelihood is high, we prefer a planar hypothesis; if it is very low, we prefer a curved hypothesis.

The second issue is how to put this information into the hypothesis graph. Somehow we need to modify the edge values within the graph to reflect these likelihoods while not significantly modifying the final region groupings. In this implementation we use modifiers to the edge likelihoods that are based on the test of the planarity of an image region.

Finally, we have to decide how to update the edges of the hypothesis graph after merging two regions. For example, if the algorithm merges a hypothesis from a region better described by a curved hypothesis with one from a region better described by a planar hypothesis, how should it rank the fit of the aggregate hypothesis with the image data?

3.1. A Test of Planarity

We focus upon the likelihood that a planar surface created a particular image region because it is the most restrictive case in terms of the geometric complexity of the surface. We base this test on the assumption that the global illumination for the scene is distant with respect to the extent of the objects. This assumption implies that the angle between the surface normal of a plane and the light source does not change significantly over the surface's extent. If we model the body reflection of dielectrics as Lambertian, which assumes that an object's intensity depends only upon the angle between the light source and the surface normal, then the apparent intensity of a planar surface patch will be constant within a given threshold over its extent. This implementation does not consider the issue of shadows.

The combination of the distant-light-source assumption with the system's assumption of piecewise uniform objects implies that a planar surface patch will possess approximately uniform intensity over its extent. On the other hand, a sufficiently curved surface will

contain variations in intensity. Clearly, we cannot differentiate between small regions of slowly curving surfaces and regions of planar surfaces, but this reasoning does allow us to differentiate between surfaces with significant curvature and surfaces approximating a plane.

To make this differentiation in a given image we look at the uniformity of the intensity of each image region. We model a planar surface's appearance with a normal distribution and assume that most of the pixel intensities will fall within a range around the mean intensity. A threshold variance based on test images of planar surfaces defines this range.

It is important to note that we cannot use the same variance for each region because of the different relative intensities. What we must do is compare the variances relative to the average brightness of the region. To accomplish this for a given region the algorithm first calculates the sample mean and variance of the absolute pixel values. It then divides the sample variance by the square of the average pixel value for the region, giving a relative variance. The complete formula for the relative variance is given by the following equation where $\bar{I}$ is the average pixel intensity and N is the number of pixels in the region:

$$\sigma_{\text{rel}}^2 = \frac{1}{I^2} \sum_{i=1}^{N} \frac{(I_i - \bar{I})^2}{N-1}. \tag{5}$$

After calculating the relative variance for a given region, we compare it to the threshold variance using a χ^2 test. We specify the threshold variance with respect to the maximum possible pixel value. The result of the χ^2 test indicates the likelihood that the sample is from the population of surfaces with uniform appearance.

We computed the relative variance of 76 curved regions and 27 planar regions in 10 test images. Except for images containing the mug object, the difference between the relative variances of the curved and planar regions was an order of magnitude or more. Based on these results we selected a threshold variance of 180, which translates to a standard deviation of 13.4 pixel values out of 255, and a relative threshold variance of 0.0029. This threshold gave the best possible performance on a set of 172 regions on dielectric multicolored objects.

We take the additional step of only making a ranking decision when the results are very strong or very weak. If the χ^2 test returns a value between 0.25 and 0.75, then we claim that the results are ambiguous and are not a good basis for ranking the hypotheses. Given the large average difference between the relative thresholds of the curved and planar regions, however, very few of the regions fall within this middle range. Figure 9 shows the results of the test of planarity on the stop sign and cup and mug images.

3.2. Modifying the Hypothesis Graph

Once we have the results of the planarity test we want to incorporate this information into the hypothesis graph. To accomplish this we modify the edge values in the graph by devaluing the edge likelihoods connecting hypotheses that do not agree with the results of the planarity test.

After running the compatibility tests—reflectance ratio, gradient direction, and profile analysis—on each hypothesis pair we also run the planarity test on each hypothesis's region. If both hypotheses match the results of the planarity test on their respective regions, then we do not modify the edge value connecting the two. If one of the hypotheses does not match its planarity result, the algorithm reduces the connecting edge's value by multiplying it by 98.5%. If both the hypotheses disagree with the planarity results, the algorithm reduces the connecting edge's value by multiplying it by 98%. This gives a hierarchy of edge values,

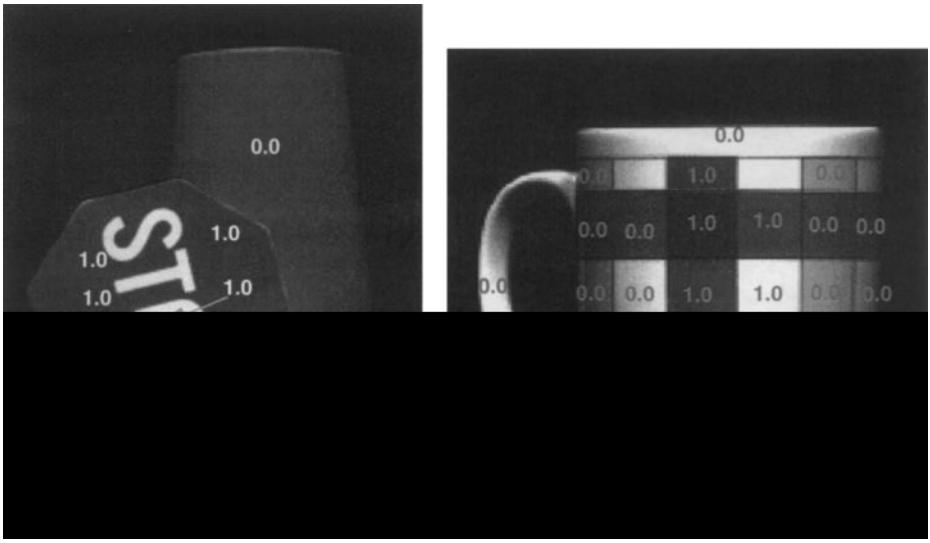

FIG. 9. Planarity test results on (a) the stop-sign and cup image and (b) the mug image. A high value indicates that the region has approximately uniform intensity.

penalizing most the edges connecting the hypotheses that most disagree with the planarity test, but not significantly modifying the overall belief that the two regions are somehow related. Figure 10 gives a graphic demonstration of this hierarchy.

While this may not seem like a significant change in the edge values, the best-first merge algorithm chooses the highest likelihood edge at each step. Therefore, these slight differentiations in edge values cause the algorithm to prefer some hypothesis pairs over others without significantly changing the likely region groupings.

3.3. Modifying the Extraction Algorithm

As specified above, the final task is to dynamically update the edges as the algorithm merges hypotheses into larger clusters. The first task is to define the planarity of an aggregate hypothesis that covers multiple image regions. We define the planarity value for an aggregate region as the average of the planarity likelihoods. Using this new planarity value, the algorithm can modify all of the connecting edges as described above.

The result of making these modifications to the system is shown in Fig. 11. Figure 11c, for example, shows the best segmentation and interpretation of the stop-sign and cup image. Unlike the mug's segmentation in Fig. 1c, this segmentation also contains an interpretation

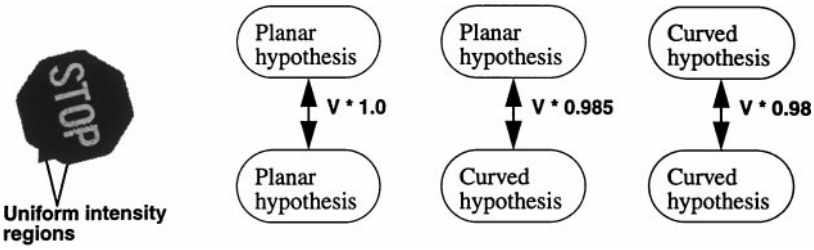

FIG. 10. The red sign body and the white "P" form two approximately uniform intensity regions. Each region has a curved and a planar hypothesis. The diagrams show the edge multipliers that modify the weights according to the likelihood of a uniform intensity surface given the hypothesis.

of the image as consisting of two curved objects—the cup and pole—and one planar object—the sign. Results are also given for two other images containing multicolored objects. Note that these results are unique in the field of physics-based vision because they provide interpretations as well as segmentations of multicolored dielectric objects.

4. EXTENDING THE HYPOTHESIS LIST

One of the strengths of this framework is that the same general system can handle increasingly complex images by increasing the number of hypotheses considered for each image region. In this section we expand the initial hypothesis list to include (colored dielectric displaying surface reflection, white uniform, curved) and (colored dielectric displaying surface reflection, white uniform, planar) when appropriate. We then show the result of this system on an image of a multicolored object displaying a highlight.

By adding these hypotheses to the initial hypothesis list the system moves beyond previous physics-based segmentation algorithms because it can not only correctly merge the highlight regions with their body reflection counterparts, but also correctly merge differently colored regions into coherent surfaces.

4.1. Characterizing Specular Regions

The key to developing a compatibility test for regions displaying surface reflection and their body reflection counterparts is understanding the relationship of these two types of reflection. Shafer first proposed the physically based dichromatic reflection model [15]. Klinker, Shafer, and Kanade then refined this model to identify highlight regions in an image and the surfaces to which they belonged [5]. Novak continued this line of inquiry by undertaking a detailed characterization and parameterization of surface reflection and body reflection for dielectric objects [12].

The dichromatic reflection model and the work of Klinker and Novak in characterizing and refining that model provides a body of knowledge upon which we can build a compatibility test. There are a number of characteristics of this relationship that we can measure. We approach this task in the same manner as the other compatibility tests, looking for characteristics that must be true for two regions to have the surface/body reflection relationship and be part of the same object.

To this end, we now give a brief description of the appearance of a highlight region on an inhomogeneous dielectric in terms of its characteristic shape in color space (RGB). Based on this characteristic shape we then describe a set of attributes we can measure and test. We conclude this section by making several observations about a highlight's appearance in the image and how this adds additional constraints to the problem.

The colors of the surface reflection and the body reflection of a single-colored inhomogeneous dielectric form a skewed T in RGB color space [5]. Inhomogeneous dielectrics include paper, plastic, ceramic, paint, and numerous other materials. RGB color space is equivalent to a 3-D histogram, where a bucket is filled if that color exists in the image.

The important characteristics of the skewed T are highlighted in Fig. 12. First, the body and surface reflection of a single color surface form a plane in color space. The body color and light source color define this plane. Second, most of the pixel colors lie very close to either the body or light source color vectors. Third, these two vectors meet in a well-defined manner. Finally, the camera usually clips the colors of some of the surface reflection pixels.

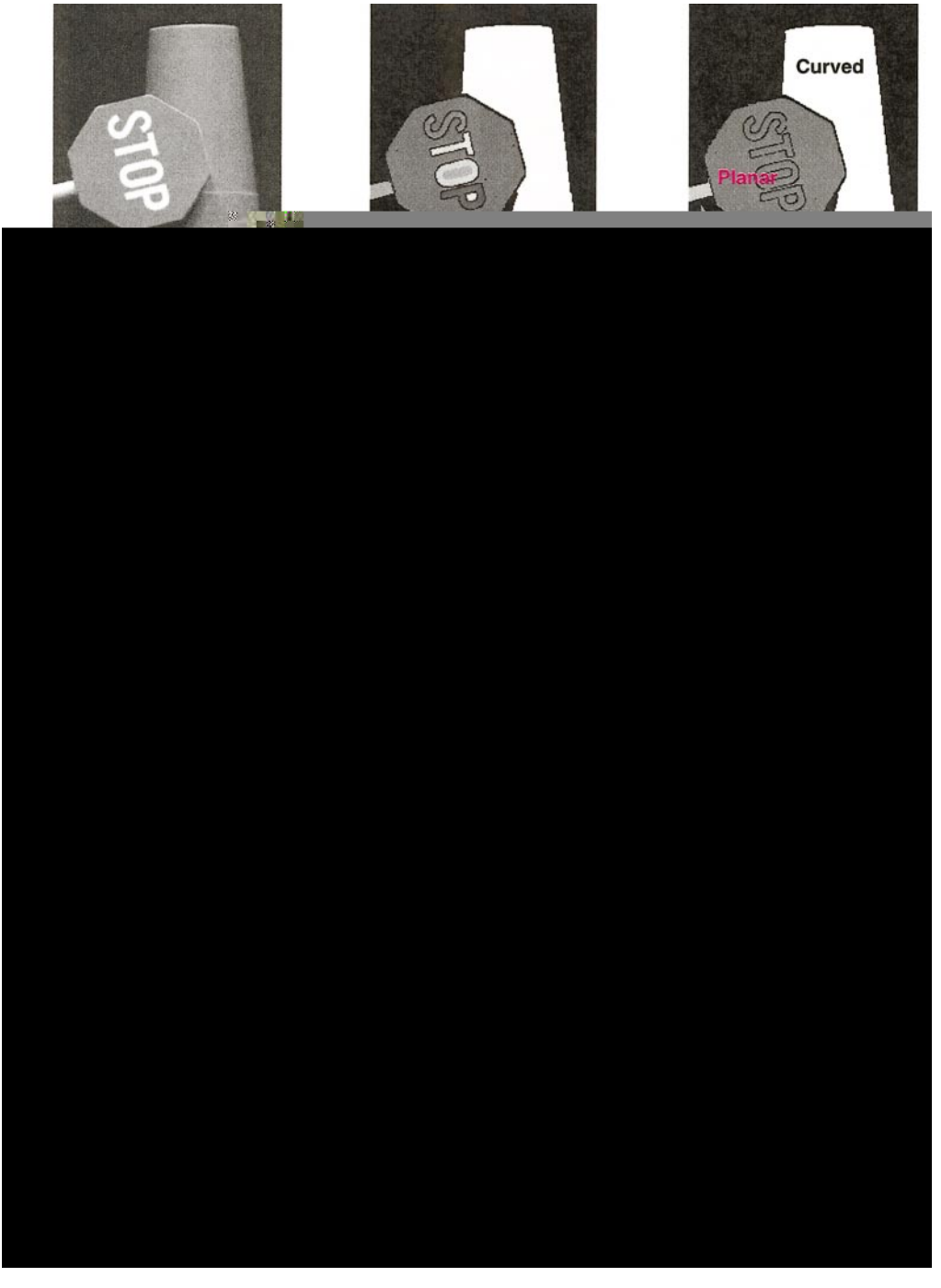

FIG. 11. (a) Image of a stop-sign and cup; (b) segmentation based on color; (c) final region grouping and best interpretation of the image showing the cup as a curved surface and the stop-sign as a planar surface; (d) initial image and (e) best final segmentation and interpretation of the cylinder–plane image; (f) initial image and (g) best final segmentation and interpretation of the Pepsi™ image. Pepsi™ trademarked logo used with permission of PepsiCo, Inc.

FIG. 12. Specular reflection (in yellow) and body reflection (in magenta) as they would appear in the RGB histogram space.

It is important to explain the "well-defined manner" in which the body and surface reflection vectors meet. Put simply, these two vectors must intersect in the upper 50% of the body reflection vector, where the pixel exhibiting the maximum intensity body reflection defines the upper end of this vector and the black point defines the lower end. Klinker initially proposed this rule as a heuristic, and the rule was later proven by Novak [4, 10].

The 50% rule immediately suggests several characteristics to test for compatibility. We can look at the minimum and maximum intensity values in the hypothesized highlight region and compare them to the body reflection intensities. We can also measure whether the body and surface reflection vectors intersect in the upper 50% of the body reflection vector.

4.2. Specular Compatibility Test

The compatibility test takes as input a proposed highlight region and a proposed body reflection region. It returns a likelihood that they fit the skewed-T model based on the previously described physical model. A sketch of the algorithm follows.

First, using the methods of Klinker *et al.*, we process the highlight region to remove the effects of color clipping [5]. We accomplish this by estimating the surface reflection vector in RGB color space using unclipped pixels and then projecting the clipped pixels onto that vector.

Second, using the two primary vectors in the RGB color space, we calculate where on the body reflection vector the surface reflection vector intersects. According to the 50% rule, this intersection must take place in the upper half of the body reflection vector. The two primary vectors are found using a least-squares regression on the unclipped pixels in the proposed highlight and body reflection regions.

Third, we use three other characteristics of the regions in color space as pass/fail tests. These tests are, in part, based upon the detailed analysis of highlights by Novak [10]. They are designed to robustly differentiate between highlight regions and regions with white body reflection.

- All of the surface reflection pixels must be brighter than 50% of the maximum body reflection brightness.
- The average intensity of the surface reflection pixels must be greater than the average intensity of the body reflection pixels.
- The brightest body reflection pixel must be less than the brightest surface reflection pixel.

If the proposed body reflection and highlight regions pass all of these requirements, then the algorithm needs to return a likelihood of a merger for the hypothesis pair. Given that it is unlikely for a region pair that is not a body reflection and surface reflection pair to have these characteristics, we would like the likelihood to be high. Given d_x, the distance along the body reflection vector to the intersection, the algorithm uses the heuristic likelihood equation

$$L(\text{merge}) = 1 - (d_x - 0.75)^2 \tag{6}$$

to obtain the likelihood of a merge. Otherwise, the algorithm returns a likelihood of 1×10^{-7} (zero).

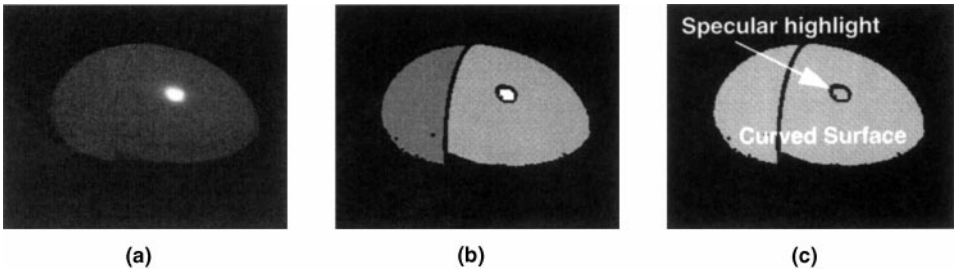

FIG. 13. (a) Image of a two-colored plastic egg with a highlight; (b) segmentation based on normalized color; (c) final region grouping, best interpretation of the image.

4.3. Integrating the Specular Hypotheses

Integrating the specular hypothesis into the segmentation framework is a straightforward process. To consider specular hypotheses for each region we simply add them to the list of initial hypotheses proposed for each region. Next, using the previously developed compatibility tests and the surface/body reflection test just described, we analyze each adjacent hypothesis pair for compatibility and use the results to generate the hypothesis graph. In this case, it is a four-layer graph. Note, in order to deal with the four hypotheses per regions we had to modify the ranking system to be a four-tiered system based on how likely a hypothesis is for a given region. Finally, we apply the segmentation extraction algorithm to the graph to find the best interpretations.

Figure 13 shows the best segmentation and interpretation of an image of a two-colored plastic egg with a highlight. An important test of the updated system is that the segmentation and interpretation of the stop sign and cup image in Fig. 11 does not change, even though that image contains small white regions that might be mistaken for highlights. Thus, the test of surface/body reflection region compatibility can discriminate between highlights, such as on the egg, and small white regions, such as on the stop sign.

5. LIMITATIONS OF THE PHYSICS-BASED APPROACH

The primary limitation of a physics-based approach to image analysis is the need for sufficient information to analyze. If a region is very small (<100 pixels) then its boundaries with adjacent regions will be short, and there will be few pixels available on any one scanline for use in the intensity profile analysis. This lack of information can lead to false positive matches in all of the three compatibility tests. The reflectance ratio and gradient direction tests, as they are variance-based, are not accurate for small numbers. The intensity profile analysis also tends to return false positives for small regions because of the limited variation that can appear in them. However, it is still the most robust of the three compatibility tests.

There are two directions from which a solution to the problem can come. The first is a bottom-up heuristic measure that provides guidance in cases where there is insufficient information to calculate physical attributes. The second is top-down guidance based on a current model of the scene that uses knowledge to suggest the most likely assignment for the small region.

Secondary limitations in this approach are due to the assumptions necessary for the tests of compatibility. The primary assumption is that the image contains only piecewise uniform

objects. While this is a looser assumption than previous physics-based methods—uniformly colored objects are assumed in [1, 4]—it is still a strong restriction on the class of objects that will be correctly handled.

Finally, implementation-dependent limitations include the fact that the analysis of highlights will not work correctly when the highlight region crosses a diffuse boundary. While the highlight may correctly split at the boundary due to the color change in the underlying body reflection, and the body reflection-highlight regions will be correctly handled, the current implementation does not handle highlight–highlight compatibility tests, which are an area of future work.

6. SUMMARY

This paper demonstrates that a physics-based framework allows us to segment images of complexity increased over previous segmentation algorithms, physics-based or not. The key improvement of the segmentation aspect of this system is that it provides a grouping of regions that corresponds more closely to the kind of perceptual grouping a person would make; it finds coherent surfaces. This paper shows how we can take the basic framework outlined in [6] and add to its capabilities by supporting more hypotheses per initial region than the initial results reported in [7]. Future work on the bottom-up aspects of the system can draw from the success in dealing with images of objects with highlights.

In addition, this paper shows that by considering multiple hypotheses and preferring certain hypotheses over others based on the image data we have created a new and powerful approach to image interpretation. These interpretations provide basic explanations for the material, shape, and illumination of coherent surfaces in a scene. These interpretations, in turn, have the potential to offer improved input to object recognition and selection algorithms for higher-level tasks.

REFERENCES

1. R. Bajcsy, S. W. Lee, and A. Leonardis, Color image segmentation with detection of highlights and local illumination induced by inter-reflection, in *Proc. International Conference on Pattern Recognition, Atlantic City, NJ, 1990*, pp. 785–790.
2. P. Breton and S. W. Zucker, Shadows and shading flow fields, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, June, 1996*, pp. 782–789.
3. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, 1973.
4. G. Healey, Using color for geometry-insensitive segmentation, *J. Opt. Soc. Amer. A* **6**(6), 1989, 920–937.
5. G. J. Klinker, S. A. Shafer, and T. Kanade, A physical approach to color image understanding, *Int. J. Comput. Vis.* **4**(1), 1990, 7–38.
6. J. Krumm, *Space Frequency Shape Inference and Segmentation of 3D Surfaces*, Ph.D. Dissertation, CMU-RI-TR-93-32, Carnegie Mellon University, December 1993.
7. S. M. LaValle and S. A. Hutchinson, *A Bayesian Segmentation Methodology for Parametric Image Models*, Technical Report UIUC-BI-AI-RCV-93-06, University of Illinois at Urbana–Champaign Robotics/Computer Vision Series.
8. B. A. Maxwell and S. A. Shafer, Physics-based segmentation: Moving beyond color, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1996*, pp. 742–749.
9. B. A. Maxwell and S. A. Shafer, Physics-based segmentation of complex objects using multiple hypotheses of image formation, *Comput. Vision Image Understand*. 1997.

10. B. A. Maxwell, *Segmentation and Interpretation Using Multiple Physical Hypotheses of Image Formation*, Ph.D. thesis, CMU-RI-TR-96-28, Robotics Institute, Carnegie Mellon University, 1996.
11. S. K. Nayar and R. M. Bolle, Reflectance based object recognition, *Int. J. Comput. Vision* 1996.
12. C. Novak, *Estimating Scene Properties by Analyzing Color Histograms with Physics-Based Models*, Ph.D. thesis, School of Computer Science, Carnegie Mellon University, 1992.
13. D. Panjwani and G. Healey, Results using random field models for the segmentation of color images of natural scenes, in *Proceedings of International Conference on Computer Vision, June 1995*, pp. 714–719.
14. J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore, 1989.
15. S. A. Shafer, *Using Color to Separate Reflection Components*, COLOR Research and Application, No. 10, pp. 210–218, 1985.
16. R. S. Wallace, Finding Natural Clusters through Entropy Minimization, Ph.D. Dissertation, School of Computer Science, Carnegie Mellon University, CMU-CS-89-183, June, 1989.