A Physics-Based Approach to Interactive Segmentation

Bruce A. Maxwell

Swarthmore College, Dept. of Engineering
Swarthmore, PA 19081, USA
(610)328-8081 (phone)
(610)328-8082 (fax)
maxwell@swarthmore.edu

Abstract. Interactive segmentation for image manipulation and composition is becoming increasingly important as digital imaging devices become ubiquitous. Users want to select objects in one image to put them in another. This paper proposes a method for selecting image regions that are likely to correspond to multi-colored objects rather than just regions of similar color and/or texture. The method is based upon a physical analysis of neighboring regions using the reflectance ratio along the region borders. This measure provides an illumination and geometry invariant cue to scene coherence between regions of different color. Using this measure and a user-selected seed region, the computer can automatically or interactively grow the region to include neighboring regions of different color that are likely to be part of the same object.

1 Introduction

The tremendous amount of digital imagery now available to anyone with a computer has significantly increased the need for both automatic and interactive image manipulation tools. One of the most important steps in image manipulation is segmentation: selecting coherent areas of the image for manipulation and modification. Automatic segmentation tools have been studied and developed within the computer vision community since its inception. Since computer vision is focused on the use of images by computers, however, interactive segmentation techniques have only recently been developed, mostly in the graphics and medical imaging community where image manipulation for use by people is a primary goal (see, for example [4]).

Previous work on interactive segmentation falls into three categories: boundary following, active contour and volume models, and region growing. Boundary following methods interact with the user as he or she traces a region of interest in an image. The computer may learn the boundary features from the user and/or use precomputed features to snap the traced boundary to the edge of the desired image region. An example of boundary following is the intelligent scissors method of Mortensen and Barrett [10].

Active contour models generally require the user to provide an approximate boundary tracing. Then the contour fits itself onto the image according to the interactions of the model parameters and calculated image features, such as gradient magnitude. Examples of active contour models are snakes in 2D, and balloons in 3D [12][4].

Finally, region growing methods, such as the Adobe Photoshop™ magic wand tool allow the user to select a seed region and then the computer expands the user-selected region according to a calculated similarity measure [1]. Similarity is normally based on color and/or small-scale texture features, although both Nayar and Bolle and Gevers and Smeulders have shown that the reflectance ratio/color ratio is useful for identifying regions of similarity and region boundaries [11][7].

Each of these methods has benefits and drawbacks to it. Region growing methods do not require the user to trace a boundary, but they can only select regions that have similar appearance. Therefore, they will not adequately segment a multi-colored object, for example, without further user interaction. Boundary following methods, on the other hand, will allow the user to select a multi-colored area of an image so long as the boundary function is adaptive along its length or the differently colored areas have similar boundary features. Active contours allow the user to provide approximate shape information, but rely heavily on computed features which may vary significantly across a multi-colored region. The active contour method is also more difficult than the other two methods to make an interactive process beyond the initial approximate boundary tracing.

This paper presents a new method of interactive segmentation that is geared towards selecting single or multi-colored coherent surfaces. By using a physical analysis of adjacent regions, the computer hypothesizes whether two adjacent regions are part of the same coherent surface in the scene. Combining this physical analysis with standard region growing segmentation methods allows the computer to grow a region that covers a coherent surface even though that surface contains multiple colors. The user can interact with this process in at least two ways. First, they can select adjacent regions and indicate they are part of the same surface. Second, the computer can query

the user when it adds a new section of the image to the selected region to make sure the addition is correct. This process moves interactive segmentation beyond the limitations of current region growing methods and allows the user to select coherent surfaces that will more often correspond to the object of interest rather than a region of similar color or texture.

2 Analysis of Piece-wise Uniform Surfaces

For the computer to make a decision about whether neighboring regions are part of the same surface it must somehow estimate their similarity in the scene. Coherence between adjoining surfaces in a scene, however, does not necessarily map to identifiable coherence in an image. Likewise, incoherent surfaces in a scene do not necessarily map to incoherent regions in an image. As an example of the former, consider a sphere that is half green plastic and half a shiny metal. While the sphere forms a coherent surface in the scene, its projection into an image does not necessarily reflect this coherence. Likewise, a red ball sitting on a red plane will display coherence in color across both regions despite the fact they are incoherent surfaces.

For certain classes of objects we can, however, make predictions about coherence in the scene by analyzing relationships in the image. In particular, we can use physical analysis of appearance to examine surfaces that are either piece-wise uniform, or that vary slowly relative to their image projection. Nayar and Bolle, for example, have developed the reflectance ratio as an illumination and shape invariant measure of relative albedo [11]. A similar measure, the color ratio, has been used by Gevers and Smeulders for illumination and geometry invariant segmentation and recognition [6]. Breton and Zucker have shown that the shading flow field--related to the direction of the gradient of image intensity--is related to surface coherence [1]. Finally, Maxwell and Shafer formulated a test of surface coherence based on three measures: reflectance ratio, direction of the gradient of image intensity, and the smoothness of the image profile across adjacent regions [9].

This work on interactive segmentation is based upon the reflectance ratio similarity measure used by Maxwell & Shafer [9]. At is essence, the reflectance ratio is a measure of the difference in transfer function between two pixels that is invariant to illumination and shape so long as the latter two elements are similar. Nayar and Bolle developed the concept and have shown it to be effective for both segmentation and object recognition [11], as have Gevers and Smeulders [6]. The reflectance ratio was originally defined for intensity images and measures the ratio in albedo between two points. The albedo of a point is the percentage of light reflected by the surface.

The principle underlying the reflectance ratio is that two nearby points in an image are likely to be nearby points in the scene. Therefore, they most likely possess similar illumination environments and geometric characteristics as shown in Fig. 1(a).

Nayar and Bolle represent the intensity I of a pixel in an image as

$$I_i = k\rho_i R(s, v, n) \tag{1}$$

where k represents the sensor response and light source intensity, ρ is the albedo of the surface, and $R(s, v, n)$ is a scattering function representing the geometry dependent aspects of the transfer function and exitant illumination field. If p_1 and p_2 are nearby

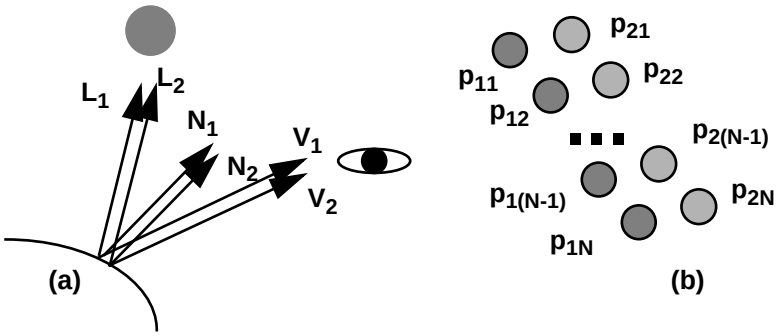


Fig. 1: (a) Two nearby points. Note that the geometry is approximately the same for both points. The ratio of the intensities is, therefore, a function of the albedos of the two points. (b) Pixel pairs along the border of two regions. Two regions that are part of the same object should have a constant reflectance ratio along their boundary.

points in a scene, then the geometry dependent functions should be similar, as will the light source brightness and sensor response. Therefore, if we take the ratio of the intensities of two nearby pixels in an image, then we obtain the ratio of albedos, or the reflectance ratio π as shown in (2).

$$\pi = \frac{I_1}{I_2} = \frac{k\rho_1 R(s, v, n)}{k\rho_2 R(s, v, n)} = \frac{\rho_1}{\rho_2} \tag{2}$$

Multiple light sources do not affect the reflectance ratio so long as the assumption of similar geometries and illumination environments holds. A well-behaved version of the reflectance ratio is the difference in intensities divided by their sum, as in (3) [11].

$$r = \left(\frac{I_1 - I_2}{I_1 + I_2}\right) \tag{3}$$

Note that the geometry and sensor dependent terms still cancel out. Unlike the ratio in (2), however, this measure of the reflectance ratio ranges from [-1,1].

Given this measure of the difference between the transfer functions of adjacent pixels, how do we use it to measure surface coherence? The basic idea is to look for constant reflectance ratios along region boundaries. If the reflectance ratio along the boundary connecting two regions of different intensity is not constant, then either the shape or illumination are incompatible in addition to the regions' colors.

The algorithm is as follows. As shown in Fig. 1(b), for each border pixel p_{1i} in h_1 that borders on h_2 the algorithm finds the nearest pixel p_{2i} in h_2 and calculates the reflectance ratio between them if both pixels are equal to or brighter than a dark threshold d_{rr}. Since the reflectance ratio is designed to work on intensity images, not color images, the algorithm uses the pixel intensity values.

Summing up the ratios of all of the sufficiently bright pixel pairs the algorithm calculates the mean reflectance ratio r_{avg}. It then finds the variance of the reflectance ratio along the border. 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, regardless of the shape or illumination. The variance is a simple measure of constancy. If h_1 and h_2 are part of the

same object, this variance should be small, with variation existing because of 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 assumption underlying the reflectance ratio. This should result in a larger variance. We can differentiate between these two cases using a threshold variance and chi-square test. Region pairs with a small variance may be coherent surfaces; region pairs with large variances are not coherent surfaces. This test, therefore, will rule out merging some region pairs and build confidence for merging others [9].

It is important to note that not all discontinuous surfaces will produce a significant variance in the reflectance ratio. However, the expectation is that for most camera positions the reflectance ratio between two different surfaces will vary. The algorithm uses a variance of 0.008, or a standard deviation of 0.09, as a threshold for all of the cases in this experiment. This is approximately 4.5% of the range [-1,1].

3 Heuristic Segmentation Tools

In addition to the physical analysis, we can use heuristic region merging techniques to improve the results for most cases. Beveridge *et. al.*, for example, proposed several useful heuristic measures based on region connectivity and region size [2]. For this work, we used their connectivity score, and our own size-based heuristic.

The connectivity score is based on the idea that regions that share a large percentage of their boundary are more likely to be connected than regions that share a small percentage. This heuristic is quite useful with the reflectance ratio estimate as small segments of a border are likely to have little variation in the reflectance ratio along that border simply because of its length. The connectivity score is based upon the expression in (4), which says that the connectivity score increases with increasing shared boundary.

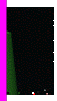
$$conn(a, b) = \frac{4 \times shared(a, b)}{min(length(a), length(b)))} \tag{4}$$

The connectivity score is then bounded in the range [0.5, 1.0], which means that any shared boundary greater than 25% of the smaller region is not penalized.

The size heuristic for the interactive segmentation task is based on the idea that regions should not be too large: in general we don't want to select more than half the picture if we are trying to select objects in a scene. The purpose is to stop the segmentation tool from selecting the background. The size heuristic is based upon the relationship given in (5), bounded to the range [0.5, 1.0].

$$size(a, b) = \sqrt{1 - \frac{(size(a) + size(b))}{size(\text{image})}} \tag{5}$$

This relationship says that increasing size increases the penalty for merging. The penalty flattens out for any region above half the size of the image. The square root ensures that medium and small regions are not heavily penalized. With the physical and heuristic merging techniques in place, we can now turn to the interactive segmentation and object selection process.









ws. The session begins when the user
r example, in the image of a stop-sign
in the sign, hoping to select the entire

stop-sign. Using this point as a seed area, the computer grows a region using normalized color as the similarity feature. This region, shown in Fig. 3a, becomes the seed region that starts an iterative process of looking for adjacent regions that might be part of the same surface.

Step two is to find a boundary for the seed region and its neighbors. This is accomplished in three stages. First a grassfire transform of the inner region identifies all pixels that are a distance d_I away from the edge of the region. These pixels form the inner border. Second, a grassfire transform of the outer region identifies all pixels that are a distance d_O from the background border; these pixels form the outer border. The distance values d_I and d_O are user-supplied constants that are chosen to avoid pixels that overlap different regions in the scene. Finally, the two borders are related by finding adjacent pixel pairs, one from the in-region border and one from the out-region border. Adjacent pairs are found by following a line perpendicular to the tangent of the in-region border until it hits an out-region border pixel. The borders are shown in Fig. 3b.

Step three uses the reflectance ratio analysis of section 2 to divide the border into regions where the reflectance ratios are similar. This step segments the boundary into sections as in Fig. 3c. The largest of these sections corresponds to the adjacent region that shares the greatest boundary similarity--in a physical sense--with the seed region.

Step four is grow an adjacent region using the largest exterior boundary section as a seed. In the stop-sign and cup example, this results in the red portion of the stop-sign being selected as shown in Fig. 4.

After step four, we can recalculate the segmentation of the inner and outer borders to account for the new region information, resulting in the Fig. 3d. Note that the entire border between the seed region and the adjacent region is now marked as coherent. In step five, we test for coherence between the two regions to see if they are likely to be part of the same surface. This test uses the methods outlined in sections 2 and 3 to determine a match likelihood.

If the two regions demonstrate significant coherence, the computer will join them together into a new seed region as shown in Fig. 5. If they do not demonstrate coherence, then the process will either stop, if there are no unexplored sections of the border, or it will repeat from step 4 and try a different adjacent region.

If the two regions are joined, then the process will repeat from step one with the aggregate region. When the algorithm cannot find any more coherent adjacent regions it stops and returns the current region. Extracted regions in the stop-sign and cup example image are shown in Fig. 6.

5 Results and Conclusion

The results shown in Fig. 6 demonstrate that this approach provides a segmentation corresponding more closely to objects in the scene than methods based primarily on color and/or texture. Fig. 7 shows a more complex scene, where the algorithm is able to extract the plastic Lego™ object from a colored planar background.

The key to making this approach more robust is twofold. First, as shown in [9], using three measures of similarity is more robust than using just the reflectance ratio. Second, in order to make the region's edges cleaner a method such as snakes or the intelligent scissors method of [10] could be given the region's boundary as an input.

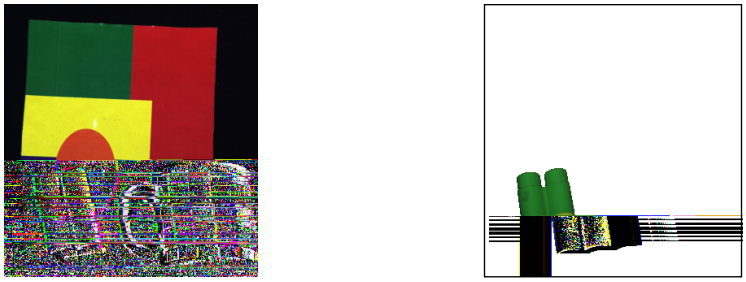


Fig. 7. Selecting an object against a complex background. The region on the right was selected by the computer after clicking in the green plastic lego.

The strength of an approach based on a physical analysis of the scene is that it gives the computer a scene-based estimation of similarity which corresponds more closely to how people want to select and manipulate the world.

References

[1] Adobe Systems, Inc., Photoshop™ 4.0, computer program, 1997.

[2] J. R. Beveridge, J. Griffith, R. Kohler, A. Hanson, and E. Riseman, "Segmenting Images Using Localized Histograms and Region Merging," *Int'l J. of Computer Vision*, 2, 1989, pp. 311-47.

[3] 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.

[4] S. Cagnoni, A. B. Dobrzeniecki, J. C. Yanch, and R. Poli, "Interactive segmentation of multi-dimensional medical data with contour-based application of genetic algorithms", in *Proceedings of IEEE Int'l Conference on Image Processing*, pp. 498-502, 1994.

[5] L. D. Cohen, "On Active Contour Models and Balloons", *CVGIP: Image Understanding*, vol. 52, no. 2, pp. 211-218, March 1991.

[6] T. Gevers and A. W. M. Smeulders, "Color Based Object Recognition, in *Proceedings of ICIAP*, Florence, Italy, 1997.

[7] T. Gevers and A. W. M. Smeulders, "Edge Steered Region Segmentation by Photometric Color Invariant", in *Proceedings of SCIA*, Lappeenranta, Finland, 1997.

[8] L. Lapin, *Probability and Statistics for Modern Engineering*, PWS Engineering, Boston, 1983.

[9] B. A. Maxwell and S. A. Shafer, "Physics-Based Segmentation: Moving Beyond Color," in *Proceedings of Conference on Computer Vision and Pattern Recognition*, IEEE, 1996.

[10] E. N. Mortensen and W. A. Barrett, "Intelligent Scissors for Image Composition", in *Proceedings of SIGGRAPH '95*, pp. 191-198, 1995.

[11] S. K. Nayar and R. M. Bolle, "Reflectance Based Object Recognition," *International Journal of Computer Vision*, 1996.

[12] D. Terzopoulos, A. Witkin, and M. Kass, "Constraints on Deformable Models: Recovering 3D Shape and Nonrigid Motion", *Artificial Intelligence*, 36, pp. 91-123, 1988.