Towards Object-Based Retrieval for Image Libraries

Bruce A. Maxwell
Swarthmore College
500 College Ave.
Swarthmore, PA 19081
maxwell@swarthmore.edu

Abstract

Object discovery is one of the steps to effective content-based image retrieval. If a vision system has the ability to identify potential objects in an image, it enables the user to phrase queries in terms of objects, and it enables the computer to search the image library for similar objects. This paper describes a physics-based approach to object discovery. The approach grows seed regions to include neighboring regions when they are deemed coherent according to a suite of physics-based and heuristic measures. Using this approach, the computer is able to extract coherent surfaces from an image library and use the resulting object images to execute searches. Initial results indicate higher accuracy and precision for object-based image retrieval when compared to traditional histogram-based retrieval. The object-based method is able to find and retrieve images that are considered poor matches by other methods.

Keywords: content-based image retrieval, physics-based vision, segmentation, grouping

1 Introduction and Goals

Object discovery in a single image is one of the stepping stones to effective content-based image retrieval [CBIR]. If we can identify image regions that correspond to objects, then we can undertake object-based image retrieval. This will permit a user to more effectively phrase queries and search for images that contain objects of interest [5].

Object-based CBIR can be viewed as object recognition. If we acquire or build object-models, then we can use object recognition techniques to look for new instances of those objects. The limitation to this approach is that the search is not general, but limited to the object models [9].

Recent approaches to autonomous object class development look for recurring patterns in multiple images that contain the objects of interest. Weber et. al., for example, look for clusters of small templates that occur often in a series of training images [11]. Hong et. al. identify small recurring color adjacency subgraphs to generate the equivalent of object models [4]. The strength of these approaches is that they are not dependent upon particular features or assumptions, but are data driven. This is also their weakness. They will never work on a single image because there is no inherent understanding of the image formation process, and thus a single image does not provide enough data.

An alternative to an object recognition approach is to approximate object-based CBIR using generic spatial features such as variance, region, or color adjacency graph relationships [5]. This approach does not try to determine whether adjacent regions should be linked together, but only looks for similar patterns between images. While this approach is computationally fast, it does not tell the difference between a red car next to a yellow fire-hydrant and a red and yellow box.

A third approach to object-based CBIR is to try grouping low-level regions or features into coherent regions or surfaces. Perceptual grouping is one way to approach this problem [1].

The approach developed in this paper is grounded in the physics of appearance. It uses knowledge about the appearance of surfaces to build them up from simple regions. These coherent surfaces then form the basis for defining objects and conducting queries.

This approach is taken with the realization that none of these approaches will probably be sufficient by themselves for robust CBIR. Instead some combination of the above is probably necessary for robust object-based retrieval in arbitrary image data sets.

This paper focuses on a physically-based approach and brings in other tools--primarily from perceptual grouping--as necessary in order to deal with difficulties in the physical analysis and to provide robustness under noisy or ill-constrained situations.

2 Methods and Algorithms

The underlying physical analysis is based on the work of Maxwell & Shafer, Breton & Zucker, and Nayar & Bolle [6][3][8]. The process is an iterative one that begins with a single pixel, selected either by the user or an automatic click selection program that selects likely regions based on a simple image segmentation.

Using region growing, the system converts the seed

pixel into a simple image region--defined as a set of pixels that are similar in a given feature space such as hue or normalized color. The ultimate goal is to grow this initial region into a single coherent surface, even if that surface contains multiple simple regions that are different in the fundamental feature space. Regions that represent different surfaces in the scene should not be merged with the initial seed region.

Identifying neighboring seed regions is straightforward. The system first looks for coherent runs of pixels parallel to but outside the region boundary. It then selects a coherent boundary segment and grows it into a neighboring region.

The next step is to determine whether the two regions are part of the same coherent surface. This system uses three physically-based measures of coherence--continuity in the reflectance ratio, continuity in the image profile, and continuity of the direction of the image gradient--and two heuristic measures similar to Beveridge et. al--relative size, and connectedness [6][2]. The system combines these five measures using a weighted average of the three physically-based measures, with the size and connectedness measures modifying the reflectance ratio value.

If the two neighboring regions match, then the system combines them and restarts the process of finding new neighboring regions. This process terminates when there are no neighboring regions that match according to the criteria listed above. At this point the system returns the resulting region. An example output of the system is given in Figure 1. This iterative growing process was first described in [7] but used only one physically-based measure of coherence.

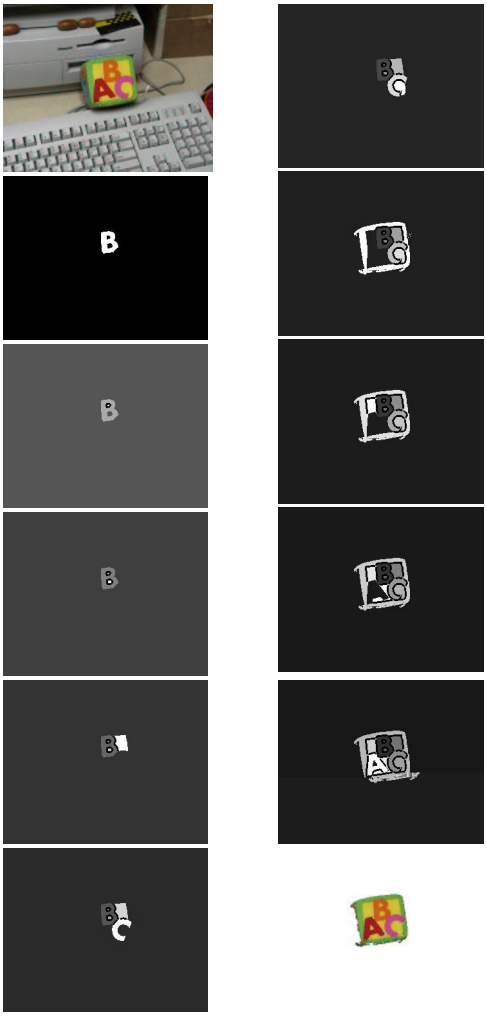


Figure 1 Example input image, sequence of region additions, and the resulting surface extraction for an initial seed point on the B.

2.1 Physically-based coherence measures

One of the most difficult aspects of this process is determining low-level coherence. This system uses a combination of three physically-based methods.

The reflectance ratio, developed by Nayar and Bolle, looks at the ratio of pixel intensity between two nearby pixels [8]. If the pixels are close to one another in the scene, then the geometric and illumination dependent terms cancel out and the result is dependent only upon object albedo. If the surfaces are not close to one another, then this is not necessarily the case.

The reflectance ratio measure is based upon the idea that, for objects with piece-wise uniform albedos, pixel pairs--one from each region--along the boundary of two regions that correspond to the same object should have a constant reflectance ratio. This can be measured explicitly through a χ^2 test of variance of the reflectance ratio along the border.

The profile analysis measure is based upon coherence in the profile of pixel intensities perpendicular to the border between two regions. A plot of intensities perpendicular to a border between two regions that are part of the same object will usually exhibit coherence except for the change in albedo between the two regions. If we can correct for this change in albedo--for example by computing the reflectance ratio--then we can align the profiles from each region in an optimal fashion.

To keep track of the albedo corrections at each pixel, the system maintains a correction map for the image. All regions get a correction coefficient relative to the initial region in the object. The system calculates the correction coefficients for a region when it calculates that region's reflectance ratio during the matching process. Having a global correction coefficient map permits the profile analysis to integrate pixels from multiple regions of different albedos.

The profile analysis measure reflects the degree of

coherence in this corrected intensity profile across two adjacent regions. The system measures coherence by fitting one or two models to the intensity profiles and testing the error in each representation. If a single model is a good fit for the profile of both regions, then it is likely that the two regions are coherent. If two independent models are a better match for the intensity profiles, then it is likely that the two regions are not coherent.

One way to automatically answer the question of which is better--one or two models--is to use an information theoretic criterion such as minimum description length [MDL]. MDL balances the cost of representing a set of data with the error in that representation. The hypothesis is that the minimum length representation of both the model parameters and the error--expressed in bits--is the best solution.

Specifically, the system fits polynomials from order 1-5 to each region's intensity profile independently, and then polynomials from order 1-5 to the combined, corrected profile. The best description length is calculated for the sum of the two independent profiles and compared to the best description length for the combined profile model.

To create a single measure for a border between two regions, at each border pixel the system calculates a measure of merit. The intensity profile used to calculate the measure is perpendicular to the border, but restricted to 45 angle increments. The measure of merit for each border pixel is given by (1), and then limited to the range [0, 1].

$$PR = \left[\frac{\mathrm{MDL}_A + \mathrm{MDL}_B}{\mathrm{MDL}_C}\right]^2 \quad (1)$$

Each MDL value is calculated using (2), where x^n is the data, θ is the model, k is the number of model parameters, and n is the number of points.

$$\mathrm{MDL} = -\log P(x^n \mid \theta) + \frac{k}{2}\log n \quad (2)$$

As a final check, the profile analysis is declared invalid if there is more than a factor of four difference between the average $\mathrm{MDL_A}$ and $\mathrm{MDL_B}$ values for the entire border, which means that the polynomial model is not a good fit for the high MDL region. In practice, this situation gives a high measure of merit because the high MDL value dominates (1). The final number returned is the median merit value for the border, which makes the result more robust to outliers.

The third measure examines continuity in the direction of the gradient of image intensity. One way to do this is to measure the sum-squared error between the angles, allowing for opposite directions to be equivalent (as they are in the case of a region boundary on the long side of a cylinder). In theory, this measure should be invariant to illumination and geometry along the border of two regions that are part of the same surface [3]. Borders between two surfaces that have different illuminations or geometries should not have continuous gradient directions.

Based on previous experience, this measure works reasonably well on smooth curved objects in high quality pictures [6]. In practice, the gradient direction is not well behaved at low gradient magnitude values, nor is it easily calculated along a region border. To overcome these difficulties, the system grows gradients outward to the border and discards border pairs with small gradient magnitudes. If fewer than 20% of the border pairs give a valid result, the measure is declared invalid for that border.

Even with these precautions, the gradient direction, while used for these experiments, does not give robust results. Further experimentation is necessary to see if a better calculation is possible and whether use of this measure actually benefits the overall process.

3 Experiments

Two experiments explore the utility of coherent surface discovery in CBIR. The first demonstrates that, for real piece-wise uniform objects, the method does a reasonable job of extracting them from complex backgrounds.

The second experiment automatically analyzes a database of images and extracts coherent surfaces. The system then combines similar image regions from the same image and generates a set of object images for each original image. We can then execute CBIR using standard methods, such as histogram matching. Since the coherent surfaces exist separately in each object image, it is possible to directly match an object in a target image with objects in the database without the effects of background clutter.

The results of the first experiment are given in Figure 2. The system successfully extracts the soft block from five images, four of which have complex backgrounds. On the sixth, the system includes one background region that is actually not that different from the block itself in terms of overall shape and appearance. The parameter configuration is the same for each image; the only thing that changes is the initial seed pixel.

The parameter configuration includes the region growing thresholds, the population variance for the reflectance ratio, the maximum order polynomial for the profile analysis, the threshold error for the gradient direction analysis, and the weighting function that combines all of these together into a single measure. Based on experimental evaluation, most of the parameters are robust within a range of values. The most sensitive parameters are the weights that combine the coherence measures. These experiments used a fixed set of these parameters, found by



Figure 2 Results of object extraction on six images with varying backgrounds.

experimentation with a set of images.

To ensure that seed pixel selection is not biased, the initial seed pixel comes from a set of automatically generated seed pixels. The seed pixel selection is based on a simple normalized color image segmentation. Each resulting region of sufficient size gets a seed pixel at its centroid (if the centroid is in the region). Figure 3 shows an example of an image and its initial seed pixels.

For the second experiment, the system uses the automatic seed pixel selection algorithm to select a set of seed pixels. For each of these seed pixels the system generates a resulting object image. A number of these results are repetitive and the system can combine similar images together into a representative set of object images. Currently, the system uses histogram matching to combine them, but a sum-squared difference approach may be more appropriate. Figure 4 shows the compressed results for one of the images.



Figure 3 Automatic seed pixel selection.



Figure 4 Input image (upper left) and all of the automatically generated unique object images.

We can use this database of object images for object-based CBIR using standard methods. Initial experiments show that a match using standard RGB histograms and the object image database works better than simple whole-image histogramming, and can find small objects in a variety of background fields. Figure 5 shows comparative results for a small image data set (23 images). The whole-image matching used the entire target image, while the object-based retrieval used an automatically extracted region from the target image (which could be the result of a mouse click by the user).

Note that the accuracy and precision are much higher for the object-based CBIR. For the top 10 images, the whole-image histogram matching gets a 30% precision and 43% accuracy (3 of 7 images containing the soft

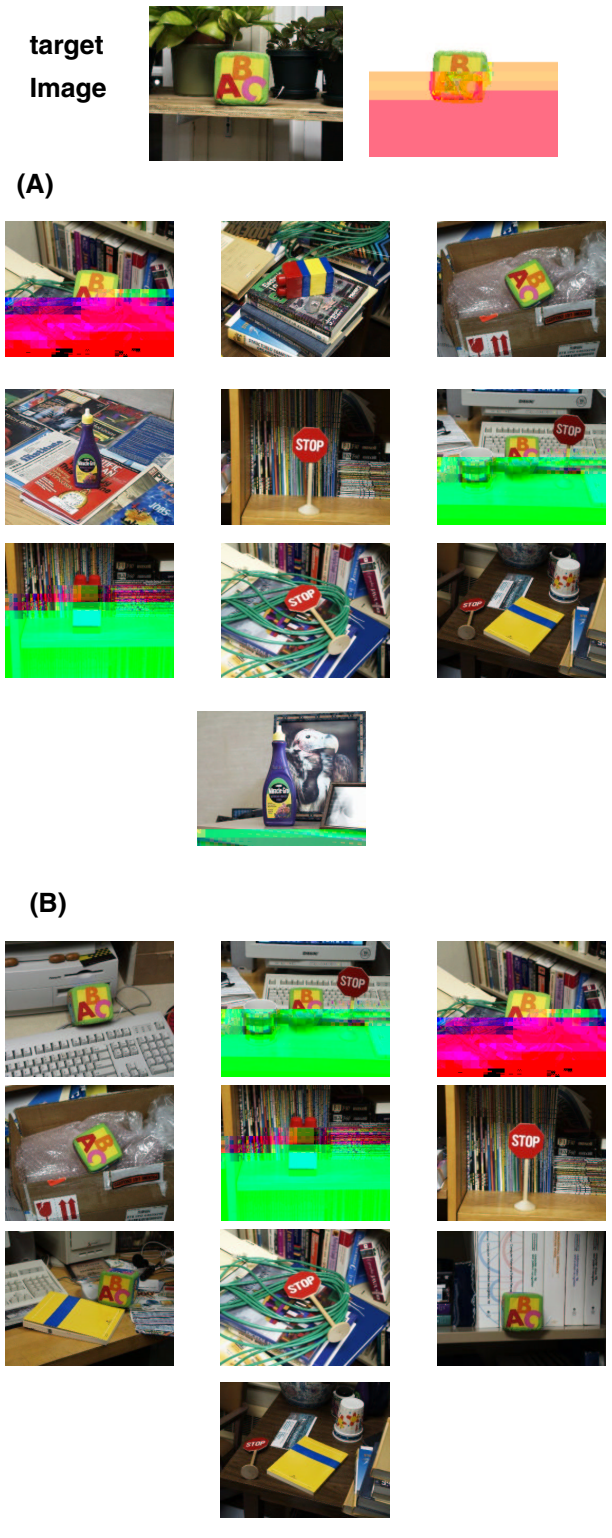


Figure 5 Comparative results for histogram-based matching: A) whole image matching, B) object-based matching. Each case shows the top ten matches from left to right, top to bottom.

block). The object-based CBIR gets 60% precision and 86% accuracy (6 of 7). Interestingly, the top match for the object-based retrieval does not even appear in the top ten images for whole image matching. This demonstrates the importance of isolating the object of interest in content-based retrieval.

4 Discussion

Low-level object extraction is a challenging problem. This is an initial attempt at a solution that tries to explicitly identify coherent surfaces with no a priori knowledge about the objects themselves. The primary success of this approach is its performance on objects that fit the basic physics behind the methodology (matte, piece-wise uniform objects, including single-color objects).

The success of the method is, in part, dependent upon the initial segmentation. If the initial segmentation under-segments the image, or combines two regions that should not be combined, then the physics-based analysis does not come into play. On the other hand, if the initial segmentation is too strict then the regions tend to be small and there may not be enough information in them to do an accurate analysis.

For these experiments, the images were run with fixed parameters. It is important to note that all of the pictures used the same parameters. Hand tuning the parameters can result in improved performance on some images. One direction of future research would be to explore automatic tuning of the parameters on an image by image basis. Clearly, with the large number of potential parameters, reducing, automatically setting, or fixing some set of those parameters is necessary to make the method usable.

With respect to the low-level measures of coherence, the different analysis methods show different strengths on real images. The profile analysis tends to make the most mistakes when adding small, simple regions to larger more complex regions because the complexity of the larger region dominates the measures (see equation (1)). The threshold on the difference in complexity helps this, but this area merits further exploration.

The gradient analysis, unfortunately, seems to be somewhat random in its assessments of coherence, and it invalidates itself much of the time. One potential solution is to explore more robust estimates of gradient direction. However, it may be the case that in real images, local gradient directions are simply too noisy to use effectively.

Finally, the reflectance ratio analysis is sensitive to small ridges or bumps that occur on one side of a region boundary but not the other. This occurs in the soft block with the creases on the corners. Most of the time, however, the reflectance ratio is more prone to false positives, which is why the heuristic size and connectivity measures multi-

ply it, but not the other two measures. This ensures that the reflectance ratio does not falsely encourage coherence when there is a small amount of data or an oddly connected region.

5 Conclusions

Object-based image retrieval has the potential to greatly improve CBIR systems. It gives users the ability to phrase queries on a more natural way. It gives the CBIR system the ability to locate objects of interest even in the face of background clutter.

The physics-based approach to object detection shows promise, and permits us to use general knowledge about the appearance of coherent surfaces. This paper takes the first step towards making this approach robust and applicable to real image data sets.

References

[1] A. Amir and M. Lindenbaum, (1999) "Ground from Figure Discrimination", *Computer Vision and Image Understanding*, Vol. 76, No. 1, pp 7-18.

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

[3] P. Breton, and S. W. Zucker, (1996) "Shadows and Shading Flow Fields," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, June, pp. 782-789.

[4] P. Hong, R. Wang, and T Huang, (2000) "Learning Patterns from Images by Combining Soft Decisions and Hard Decisions", in *Proceedings of IEEE CVPR'00*, vol I, pp 78-83.

[5] W. Hsu, T. S. Chua, and H. K. Pung (2000) "Approximating Content-Based Object-Level Image Retrieval", *Multimedia Tools and Applications*, 12(1), pp. 59-79.

[6] B. A. Maxwell and S. A. Shafer, "Segmentation and Interpretation of Multi-colored Objects with Highlights", *Computer Vision and Image Understanding*, Vol.77, No.1 January 2000, pp. 1-24.

[7] B. A. Maxwell, "A Physics-Based Approach to Interactive Segmentation", in *Visual Information and Information Systems*, D. Huijsmans and A. Smeulders (ed.), Springer-Verlag, Berlin, pp. 517-524, June 1999.

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

[9] S. Sclaroff (1997) "Deformable Prototypes for Encoding Shape Categories in Image Databases", *Pattern Recognition*, 30(4), pp. 627-41, April 1997.

[10] M. Swain and D. Ballard, (1991) "Color Indexing", Int'l Journal of Computer Vision, vol. 7, pp 11-32.

[11] M. Weber, M. Welling, and P. Perona, (2000) "Towards Automatic Discovery of Object Categories", in *Proceedings of IEEE CVPR'00*, vol II, pp 101-109.