RI 00009, 13 April 2000 Computer Science

IBM Research Report

An Integrated Segmentation Technique for Interactive Image Retrieval

R. Aditya and Sugata Ghosal
IBM Research Division
IBM India Research Lab
Block I, I.I.T. Campus, Hauz Khas
New Delhi - 110016, India.

IBM Research Division
Almaden - Austin - Beijing - Delhi - Haifa - T.J. Watson - Tokyo - Zurich

Abstract

Segmentation-based emerging content-based image retrieval systems enable the user to perform object-based database querying. We propose an integrated segmentation technique for interactive image retrieval, that is reasonably accurate and fast. An initial oversegmentation is generated by finding the dominant color modes in the global histogram of the image using the mean-shift algorithm. Edge-based processing is performed at the initial segment boundaries to merge non-obvious segments. Finally segment shapes are regularized using the Hopfield network and competitive learning to improve their perceptual quality. Scalable implementations are presented for ensuring fast serial execution of neural networks. The entire segmentation process takes less than ten seconds to segment 128×192 *stock photos on a standard workstation.*

1 Introduction

There has been a renewed interest in the core problem of image segmentation due to the development of content-based image retrieval engines that employ segment-based querying of image databases. Segmentation of query image is particularly useful for pictures containing "things" (as opposed to "stuff"). Segmentation provides an effective means for object-oriented image querying (like in [1,2]) and learning user perception of the query object through efficient query object modification and new query redefinition [3,4]. However, most of the segmentation based retrieval engines such as BlobWorld and Netra assume that the user starts his query from a database image, for which the segmentation exists in the database. We believe that a retrieval engine must give the user a capability of starting the query from her own query image. In such an interactive scenario, both speed and accuracy of the segmentation are essential. Existing segmentation-based CBIR engines advocate the accuracy of segmentation process over speed. Thus they are not particularly suitable for interactive segmentation of query image. In fact the speed of segmentation is also important for image digestion into the database.

In this report, we propose a color-based integrated approach that can generate a reasonably accurate segmentation for a wide variety query images at a reasonable speed for interactive CBIR. The basic principle is to generate an initial over segmentation of the image by detecting the dominant color features, followed by modification of the initial segmentation using edge information and a shape regularization procedure. The edge-based processing and shape regularization can be performed simultaneously or the latter is performed on the edge processed, segmented image. Two neural net-based techniques, one using Hopfield network and the other using competitive learning are proposed for shape regularization. A novel sparse implementation is proposed for ensuring the execution speed of neural networks on a standard serial workstation.

2 Overview of the proposed approach

The proposed approach is based on the individual favorable characteristics of region-based and edge-based segmentations, and it utilizes the strengths of both in a complimentary manner. The shape of the resulting segments is regularized to improve the perceptual quality of the final segmented image.

2.1 Region-based oversegmentation

The initial region-based segmentation is generated by analyzing the global histogram of the intensities of the query image in LUV plane. LUV is chosen for its approximate uniformity in perceptual sense, and its ability to decouple illumination and color information. First, the mean shift algorithm, which is a simple nonparametric procedure for estimating density gradients is employed to robustly determine the dominant color modes of the histogram, as proposed in [5]. Only after then are the clusters containing the instances of these features recovered. This procedure explicitly employs image domain information and avoids the normality assumption. Additionally segmentation resolution, e.g., under- and over-segmentation can be controlled in this technique.

2.2 Edge-based postprocessing

Since a histogram essentially captures the global characteristics of an image, a relatively large segment with a gradual change in color is frequently segmented into multiple perceptually non-obvious segments due to inherent difficulty in finding modes of the histogram. Sufficient undersegmentation of the image can merge these segments, but only at the expense of undesirable merging of segments with distinct colors with one another. We propose an edge based processing to properly merge the segments generated by the histogram process. The proposed approach takes into account the gradient properties at the initially obtained segment boundaries. First a histogram of gradient magnitudes is computed using only the pixels at the segment boundaries (actually 1-3 pixels thick boundaries are used). A high threshold, α is obtained so that, 15% of the total boundary pixels have gradient magnitudes greater that that threshold). Pixels with gradient magnitude greater than this high threshold indeed correspond to perceptually obvious strong edges in a wide variety of images. Now a low threshold, β is chosen to be a moderate, say 40% of the high threshold, α. Any boundary pixel with gradient magnitude greater than the low threshold are marked as an edge pixel. β=0.4α seems to well capture all the perceptual edges at the initial segment boundaries for a large number of reasonable quality images. A lower β is needed only in poorly captured images with very high local contrast variations (i.e., partially over- and under-exposed). Next for each segment, the fraction, F of the boundary pixels with gradient magnitude greater than β is computed for each contiguous segment. If F is less than, say 50%, it is inferred that region-based and edge-based boundaries do not substantiate each other – thus the segment is merged with the contiguous segment. This process is repeated for all segments. While segment merging, it is ensured that vector difference of mean color vectors of the candidate segments do not differ by more than 15 degrees. Usually one iteration is sufficient for removing most of the perceptually non-obvious segments from an image.

2.3 Shape regularization using neural networks

Due to inherent noise associated with the imaging process, as well as difficulty in histogram mode finding and threshold setting during edge-based postprocessing, the resulting segmentation may not be perceptually desirable to the user. Typically what is observed is that the segments tend to have a zigzag non-uniform kind of boundary. The human eye generally tends to perceive smoother kind of shapes as against local sudden variations in shape. For example, when we look at a tree, we ignore to see individual leaves at its boundaries. Rather we tend to look at the tree in its entirety and see a smooth kind of a shape. In essence the human eye searches for "regularity" in an image segment more often than not.

Zigzag boundaries essentially mean that the amount of contact between segments is increased. Thus we need to reduce this boundary between the two segments to achieve shape regularization. We use two approaches:- (1) Hopfield network based optimization formulates this problem as an optimization problem, which tries to minimize a cost function subjected to a set of constraints. (2) Competitive learning which gives good solutions to the graph partitioning problem (graph partioning-based perceptual segmentation has been advocated in [7]). We can model the image as a graph with the pixels being the nodes of the graph. Each pixel is connected to its immediate neighbors.

2.3.1 Hopfield Network

Starting from the initial edge postprocessed image, the Hopfield network iterates to *regularize* the segment shapes, i.e., the initial segment labels are changed so that the total number of edge pixels between adjacent segments are minimized. The segment sizes are constrained to have approximately same number of pixels. Let $1 \le X \le N$ denote an image pixel, and $1 \le i \le M$ represent the segment label. Then, if the pixel Y belongs to the segment j, $V_{Yj} = 1$ and $V_{Yi} = 0 \; \forall \, i \neq j$. Then the regularized segments can be obtained by minimizing

$$E = A\sum_X \left(\sum_i V_{Xi} - 1 \right)^2 + A\sum_{X,i} V_{Xi}\left(V_{Xi} - 1\right) + B\sum_i \left(\sum_X V_{Xi} - N_i \right)^2 + C\sum_{X,i} \sum_{Y \in \mathrm{nbor}(X)} \left(V_{Xi} - V_{Yi}\right)^2 \quad (1)$$

where, A, B, and C are empirically determined constants, N_i is the number of pixels in the i-th segment, and nbor(X) denotes the set of eight-connected neighbors of pixel i. The first two terms in Equation (1) ensures that $V_{Xi} = 0 \text{ or } 1$ and pixel i belongs to only one segment. The third term ensures that the number of pixels in each segment remains the same, and the fourth term approximately minimizes the number of edges between segments (better measures can be used for achieving improved performance). Of course, the optimal solution of this

nonconvex problem is both difficult and computationally expensive to obtain. However, when we start from a reasonable edge-processed segmentation, good suboptimal solutions can be obtained within 15-30 Hopfield iterations. A=075, B=0.001, and C=0.5 give good results for a large number of images.

2.3.1.1 Scalable implementation of Hopfield network. The main motivation behind neural net-based computation is that they are simple and exhibit massive parallelism suitable for large-scale scientific problems. Neural computations can however be expensive for execution on serial machines. Experiments reported in this paper are all performed on a single processor machine. In this subsection, implementation issues are discussed for fast serial execution of neural networks, taking into account the special nature of the shape regularization problem. A straightforward serial implementation of the Hopfield network is of $O(NM)$ per iteration. For image segmentation, the number of image pixels, N may be in thousands and the number of segments, M can vary in ten's. Thus the complexity of the simple-minded implementation of the Hopfield network is not suitable for interactive image segmentation.

The pixel structure of an image can be utilized to develop sparse implementation of the Hopfield network, that in addition to reducing memory requirements, effectively lowers the computational expenses. The states of the network are initialized according to the initial edge-processed segmentation. Since the initial algorithm generates a reasonable segmentation for most problems, a pixel initially belonging to a segment stays in that segment during the entire time evolution of the network or may change its membership only to the neighboring segments present in the initial segmentation. That is, if any pixel Y belongs to segment k in the initial segmentation, then either Y remains in segment k during the evolution of the network, else Y moves to a segment k' such that $k' \in \{S_Y\}$ where S_Y is the set of segments, contiguous to segment k. This is an acceptable proposition since the number of inter-segment edges can be minimized by local adjustment of the membership of a given pixel. Upper limit of $|S_Y|$ rarely exceeds four in real images. Constraining the movement of a pixel to only neighboring segments in fact helps in the optimization process for some problems by cutting down the search space, and facilitates sparse implementation of the Hopfield network for large-scale segmentation problem.

A new state updating scheme is also developed utilizing the present as well as the past states, to further reduce the computational complexity. Since the initial state of the network corresponds to a reasonable segmentation, only a small fraction of pixels changes their memberships in each iteration as the network evolves in time. Consequently, for a given pixel X, $V_{Xi}(t) - V_{Xi}(t-1)$ remains zero for almost all i's, and thus does not enter into the computational process. Therefore, for any pixel X, the inputs-outputs of the Hopfield network need to be updated for only a small number of i's (compared to $|S_X|$ for regular updates). Also, it is found that for most images, only the pixels near the segment boundaries change their memberships as the inter-segment edge pixels are minimized, and sub-optimal segmentations can be obtained by updating the states of those pixels which are near the segment boundaries at each iteration. Thus, with sparse data structures and the proposed state updating rule, the Hopfield network can be implemented with complexity $O(N_I)$ per iteration on a serial machine, where N_I is the number of inter-segment edge pixels present in the initial edge-processed segmentation.

Note that the edge-based processing and the shape regularization process can be combined in the minimization functional, given by Equation (1). This can be accomplished by properly incorporating the image gradients in the fourth term of Equation (1). Effort is underway to develop such a constraint.

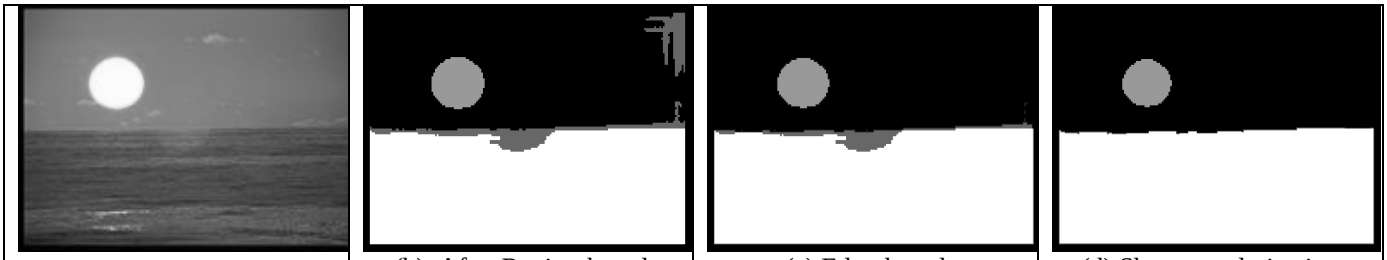
2.3.2 Competitive Learning

It is an unsupervised algorithm, based on the nonassociative statistical learning principle, and well suited for regularity detection. If the input patterns are sufficiently sparse, and/or there are sufficiently many output nodes, competitive learning network converges to a so-called perfectly stable state. In an image, each pixel is connected to its eight-neighbors only, and input patterns are quite sparse. The input patterns corresponding to an image pixel consist of 1's for eight neighbors, and 0's for rest of the pixels. The output nodes correspond to the segment labels. It is well-known that competitive learning mechanism solves the graph partitioning problem analogous to pixel-to-pixel adjacency so that the number of segment boundaries is minimal [6]. A sparse implementation is developed for interactive segmentation. Because of space constraint, its description is omitted.

Hopfield network-based solution to shape regularization offers more flexibility over competitive learning, but only at the expense of careful parameter tuning.

3 Results

The proposed integrated technique to image segmentation has been applied to a wide variety of sample stock photography images. One of these is shown below. Figure 1(a) shows the original image, taken from the "sunset" category of Corel stock pictures. The region-based segmentation is shown in Figure 1(b). The sky region is segmented into two because of gradual shading. These segments are merged by the edge-based postprocessing, as shown in Figure 1(c). The shape regularized segments are generated in Figure 1(d) using the Hopfield network. Region-based segmentation for a 128×192 image can be obtained in around 5 sec. on a standard RS6000 workstation. Edge-based postprocessing and shape regularization are dependent on the number of segment boundaries in the region-based segmentation. For a wide variety of stock photography images, it takes less than 5 sec. for edge-based postprocessing, and around 1 sec. for 20 iterations of the Hopfield network. We like to reiterate that without the initial segmentation and the proposed sparse implementation of the Hopfield network, it takes minutes of execution time for the Hopfield network to generate acceptable segmentation. The time complexity of competitive learning and Hopfield network implementations are similar. In general, better results are obtained using the Hopfield network with proper choice of parameters.



(a) Original	(b) After Region-based segmentation	(c) Edge-based postprocessing	(d) Shape regularization using the Hopfield network

Figure 1. Segmentation results using the proposed integrated approach

REFERENCES

1. C. Carson, S. Belongie, and H. Greenspan, et al, "Region-based image querying," *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, 1997.
2. W.Y. Ma, B.S. Manjunath, "NeTra: A toolbox for navigating large image databases," *Multimedia Systems*, vol. 7, no. 3, pp. 184-198, 1999.
3. G. Aggarwal, P.K. Dubey, S. Ghosal, A. Kulshreshtha, A. Sarkar, "iPURE: Perceptual and User-friendly REtrieval of images," *Proc. IEEE Conf. Multimedia and Expo*, 2000, New York, to appear.
4. G. Aggarwal, S. Ghosal, P.K. Dubey, "Efficient query modification for image retrieval," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2000, Hilton Head, to appear.
5. D. Comaniciu, D., P. Meer, "Robust Analysis of feature spaces: Color image segmentation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 1997.
6. D.E. Rumelhart, D. Zipser, "Feature discovery by competitive learning," *Cognitive Science*, vol. 9, pp. 95-112, 1985.
7. J. Shi, J. Malik, "Normalized cuts and image segmentation," ," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 1997.