MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Image Segmentation Using Region Merging Combined with a Multi-class Spectral Method

Fernando C. Monteiro
INEB - Instituto de Engenharia Biomédica
IPB - Instituto Politécnico de Bragança
Apartado 1134, 5301-857 Bragança, Portugal
monteiro@ipb.pt

Aurélio Campilho
INEB - Instituto de Engenharia Biomédica
FEUP - Fac. de Engenharia, Univ do Porto
Dr. Roberto Frias 4200-465 Porto, Portugal
campilho@fe.up.pt

Abstract

In this paper we propose an image segmentation algorithm that combines region merging with spectral-based techniques. An initial partitioning of the image into primitive regions is produced by applying a region merging approach which produces a chunk graph that takes in attention the image gradient magnitude. This initial partition is the input to a computationally efficient region segmentation process that produces the final segmentation. The latter process uses a multi-class partition that minimizes the normalized cut value for the region graph. We have efficiently applied the proposed approach with good visual and objective segmentation quality results.

1. Introduction

In a conventional sense, image segmentation is the partitioning of an image into regions, in a manner consistent with human perception, where parts within a region are similar according to some uniformity property and dissimilar between neighbouring regions. Although being a key research field in various domains, image segmentation is still on its research stage.

Spectral-based segmentation treats image segmentation as a graph partitioning problem. These methods use the eigenvectors of a matrix representation of a graph to partition image into disjoint regions with pixels in the same region having high similarity and pixels in different regions having low similarity. A common characteristic among these techniques is the idea of clustering/separating pixels or other image elements using the dominant eigenvectors of a $n \times n$ matrix derived from the pair-wise similarities, as measure by one or more cues, between pixels where n denotes the number of pixels in the image. It thus segments an image from a global point of view.

One major issue for segmentation methods based on graph representations is the size of the corresponding similarity matrix. If the node set V contains the pixels of an image, the size of the similarity matrix is equal to the squared number of pixels, and therefore generally too large to fit into computer memory completely (e.g. for an image of 481 × 321 pixels — the size of the images from the Berkeley Segmentation Dataset [3] — the similarity matrix contains $\approx 23.8 \times 10^9$ cells).

In this paper we propose a method that immensely reduce the problem size by replacing the individual pixels with micro regions in order to reduce the number of nodes in the graph. However, it is very important that the atomic regions will already yield a meaningful segmentation, i.e. the atomic regions must be homogeneous and the edges contained in the image must correspond to segment boundaries. The basic idea of the method resembles the perceptual grouping task: abandoning pixels as the basic image elements, we instead use small image regions (atomic regions) of coherent structure to define the corresponding graph representation. By treating regions as the elementary unit for image processing, we can reduce the computational complexity without a corresponding loss of accuracy. In fact, it can be argued that this is even a more natural image representation than pixels as those are merely the result of the digital image discretization. *The real world does not consist of pixels!*

This paper proposes a combined two-stage approach for image segmentation considering both local homogeneity and global information. The first stage is a gradient-based region merging algorithm and provides over segmented, but homogeneous regions. It produces a chunk graph where each chunk (or subgraph) corresponds to a micro region. In the second stage, these micro regions are used to construct a graph representation of the image, which is processed by a discrete multi-class normalized cut algorithm. The new criterion takes joint advantage of these two methods, aiming to combine the best qualities of both segmentation approaches, giving a final segmentation that is more visually appropriate.

2. Overview of the Method

In this paper, we present a new method that significantly improves the normalized cut performance by allowing the introduction of small regions (atomic regions) instead of pixels on the calculation of similarity matrix. This approach integrates region merging with spectral-based clustering as follows:

1. Initialize the graph where each node correspond to a pixel in the image;
2. Image is pre-segmented into sets (*chunks*) of connected pixels, in no particular order, using a gradient based region merging algorithm;

3. Construct the final similarity graph where each node corresponds to a chunk (atomic region) in the chunk graph;
4. Calculate the statistics of all the atomic regions;
5. Use a multi-class normalized cut process in order to obtain the final region-based segmentation.

3. Multi-class Normalized Cut

Spectral-based methods use the eigenvectors and eigenvalues of a matrix derived from the pairwise similarities of pixels. The problem of image segmentation based on pairwise similarities can be formulated as a graph partitioning problem in the following way: consider the weighted undirected graph $G = (V, E, W)$ where each node $v_i \in V$ corresponds to a pixel, and the edges E connect pairs of nodes. A weight $w_{i,j} \in \mathbb{R}_0^+$ is associated with each edge based on some property of the connected pixels (e.g., the difference in intensity, intervening contours and location). Let $\Psi = \{V_i\}_{i=1}^k$ be a multi-class disjoint partition of V such that $V = \cup_{i=1}^k V_i$ and $V_i \cap V_j = \emptyset, i \neq j$. Image segmentation is reduced to the problem of partitioning the set V into disjoint non-empty sets $(V_1, .., V_k)$, such that similarity among nodes in V_i is high and similarity across V_i and V_j is low.

The normalized cut segmentation criterion was introduced by Shi & Malik in [6] for segmentation with $k = 2$ regions. Let V_A, V_B be two disjoint sets of the graph $V_A \cap V_B = \emptyset$. We define $links\,(V_A, V_B)$ to be the total weighted connections from V_A to V_B:

$$links\,(V_A, V_B) = \sum_{i \in V_A, j \in V_B} w_{i,j}$$

The intuition behind the normalized cut criterion is that not only we want a partition with small edge cut, but we also want the subgraphs formed between the matched nodes to be as dense as possible. This latter requirement is partially satisfied by introducing the normalizing denominators in the NCut equation. The normalized cut criterion for a bipartition of the graph is then defined as follows:

$$Ncut\,(A, B) = \frac{links\,(A, B)}{links\,(A, V)} + \frac{links\,(A, B)}{links\,(B, V)} \tag{1}$$

This formulation allows to decompose the problem into a sum of individual terms, and formulate a dynamic programming solution to the multi-class normalized cut ($kNCut$). So, the NCut problem is naturally extended to a $kNCut$, finding a partition Ψ that minimizes the objective function

$$kNCut\,(\Psi) = \frac{links\left(V_1, \overline{V_1}\right)}{links\,(V_1, V)} + ... + \frac{links\left(V_k, \overline{V_k}\right)}{links\,(V_k, V)} \tag{2}$$

where $\overline{V_i}$ represents the complement of V_i.

Let $D = diag\,(D_1, ..., D_k)$ be the $n \times n$ diagonal matrix such that D_i is given by the sum of the weights of all edges on node i: $D_i = \sum_{j=1}^k W_{ij}$. It is easy to verify that

$$links\left(V_i, \overline{V_i}\right) = D_i - W_{ii} \quad \text{and} \quad links\,(V_i, V) = D_i$$

The multi-class partitioning problem can be formulated in terms of an indicator matrix. A multi-class partition of the nodes of G is represented by an $n \times k$ indicator matrix $X = [x_1, ..., x_k]$ where $X\,(i, l) = 1$ if $i \in V_l$ and 0 otherwise. Since a node is assigned to one and only one partition there is an exclusion constraint between columns of X: $XI_k = I_n$.

It follows that

$$\begin{aligned} kNCut\,(\Psi) &= \frac{x_1^T\,(D - W)\,x_1}{x_1^T D x_1} + ... + \frac{x_k^T\,(D - W)\,x_k}{x_k^T D x_k} \\ &= k - \left(\frac{x_1^T W x_1}{x_1^T D x_1} + ... + \frac{x_k^T W x_k}{x_k^T D x_k}\right) \end{aligned} \tag{3}$$

subject to $X^T D X = I_k$.

The optimal solution for the generalized Rayleigh quotients that compose Eq. (3) is the set of eigenvectors X associated with the set of the smallest eigenvalues $\Theta = \{0 = \nu_1 \leq ... \leq \nu_k\}$ of the system

$$(D - W)\,X = \Theta D X \tag{4}$$

Unfortunately, this problem is NP-hard [6] and therefore generally intractable. If we ignore the fact that the elements of x_i are either zero or one, and allow them to take continuous values, by using the method of Lagrange multipliers Eq. (4) can be expressed by the standard eigenvalue problem. Let $y_i = D^{1/2} x_i$ and $Y = [y_1, y_2, ..., y_k]$. If Y is formed with any k eigenvectors of the normalized graph Laplacian matrix[1] $\widetilde{W} = D^{-1/2} W D^{-1/2}$, then

$$\widetilde{W} Y = Y \Lambda \tag{5}$$

subject to $Y^T Y = I_k$, where Λ is the $k \times k$ diagonal matrix formed with the eigenvalues corresponding to the k eigenvectors in Y. $\Lambda = \{1 = \lambda_1 \geq ... \geq \lambda_k\}$ with $\lambda_i = 1 - \nu_i$. These k eigenvectors must be distinct to satisfy $Y^T Y = I_k$. This means that

$$Y^T \widetilde{W} Y = Y^T Y \Lambda = I_k \Lambda = \Lambda \tag{6}$$

and the *trace* of $Y^T \widetilde{W} Y$ is the sum of the eigenvalues corresponding to the k eigenvectors in Y. It follows from Fan's Theorem [1] that this sum is maximized when Y is taken to by any orthonormal basis for the subspace spanned by the eigenvectors corresponding to the k largest eigenvalues of $\widetilde{W}$. From this we arrive at the following relaxed optimization problem

$$\min_{X^T D X = I_k} kNCut\,(\Psi) = k - \max_{Y^T Y = I_k} trace\left(Y^T \widetilde{W} Y\right) \tag{7}$$

The relationship between the Laplacian matrix $\widetilde{W}$ and the Markov random walk transition matrix P was

[1] Although the Laplacian matrix is usually represented by $I - \widetilde{W}$, replacing $\widetilde{W}$ with $I - \widetilde{W}$ only changes the eigenvalues (from λ to $1 - \lambda$) and not the eigenvectors.

presented by Meila & Shi [5]. The stochastic transition matrix P is obtained by normalizing the similarity matrix in order to the rows sums are all 1 (the degree matrix of $P = D^{-1}W$ is the identity matrix).

Equation (5) can be transformed into a standard eigenvalue problem of,

$$PZ = \Lambda Z \tag{8}$$

where the eigenvectors of P are related with the eigenvectors of $\widetilde{W}$ by $Z = D^{-1/2}Y$.

$Z = [z_1, ..., z_k]$ is an $n \times k$ matrix formed by stacking the k largest eigenvectors of the eigensystem from Eq. (8) in columns. The continuous solution $\widetilde{X}$ is obtained from Z by renormalizing each of Z's rows to have unit norm.

$$\widetilde{X} = Z\left(Z^T Z\right)^{-1/2} \tag{9}$$

Recovering a discrete solution X from the continuous solution $\widetilde{X}$ is however a complex task. To overcome this problem, a majority of the theoretical work on spectral methods have dealt with successive bipartitioning generating 2^k partitions [6]. To obtain a discrete solution we follow the approach presented by Yu & Shi in [7].

4. Building the Chunk Graph

The normalized cut criterion considers global similarity relationships between nodes of a graph. This effect is achieved by constructing a fully connected graph. However, considering all pairwise pixel relations in an image may be too computational expensive. Unlike other famous clustering methods [6, 7] which use all pixels to construct a graph, our method is based on selecting edges from a chunk graph where each node corresponds to a set of homogeneous pixels.

Definition 1 (*Chunk graph*): *A chunk graph $G' = (V', E')$ for a graph G is as follows: Each node of G' represents a chunk, which is a subset of G; each chunk corresponds to a set of homogeneous pixels; chunks on G' are disjoint and their union is G.*

Therefore, we transform graph $G = (V, E)$ into a new graph $G' = (V', E')$, where $E' \subseteq E$. Graph G' is composed by a set of subgraphs (*chunks*) that follow the normalized cut criterion in their construction. This means that edges between two nodes in the same chunk should have relatively high similarity weights, and edges between nodes in different chunks should have lower similarity weights.

In the following discussion, we denote nodes of graph G' using v_i and v_j, and use e_{ij} to represent the edge connecting nodes v_i and v_j. An edge e_{ij} is labelled according to the absolute difference of the mean intensities of nodes v_i and v_j. A merge, $M(i,j)$, is a graph transformation operation that merges the nodes v_i and v_j. The procedure of node merging is actually to integrate two or more chunks into a bigger one. It is also called an *edge contraction* as the edge e_{ij} is removed. The graph G is transformed in a new graph G' that has node v_i and all other nodes of G except node v_j.

Graph G is initially set to represent the 8-neighbour of pixels in the image. Since we want to find sets of homogeneous nodes the processing order of the nodes is not important. The edges corresponding to connections between homogeneous nodes are removed. The resulting graph G' contains nodes that represent homogeneous atomic regions in the image.

By the above definition, a merge always reduces the total number of regions. This merge process is guaranteed to converge. A decision function, called the *merge criterion* determines whether two chunks should be merged. Basically, this merge criterion measures the strength of the boundary between two regions by comparing two quantities: one based on intensity differences across the boundary, and the other based on intensity differences between neighbouring pixels within each region. We define two measures

$$In_w(A) = \max_{e_{ij} \in N_8(A,E)} w_{ij}$$

$$Out_w(A,B) = \min_{v_i \in A, v_j \in B, (v_i,v_j) \in E} w_{ij}$$

where A and B are regions, $In_w(A)$ is the internal variation within the region, $N_8(A,E)$ are the 8-neighbours of A, and $Out_w(A,B)$ is the external variation between regions A and B.

We merge together regions when the external variation between them is small regard to their respective internal variations

$$Out_w(A,B) \leq MIn_w(A,B)$$

with

$$MIn_w(A,B) = \min\left(In_w(A) + \tau(A), In_w(B) + \tau(B)\right)$$

where the threshold value $\tau(A) = \alpha/|A|$ determines how large the external variation can be, with regards to the internal variation, to still be considered similar, α is some constant parameter, and $|A|$ is the size of A.

5. Implementation Issues

Images are first convolved with Gaussian oriented filter pairs to extract the magnitude of orientation energy (OE) of edge responses, as used by Malik *et al.* in [2]. At each pixel i, we can define the dominant orientation as $\theta^* = \arg\max OE_\theta$ and OE^* as the corresponding energy. The value OE^* is kept at the location of i only if it is greater than or equal to the neighbouring values. Otherwise it is replaced with a value of zero.

For computational consideration, it is important to sort and label all the regions created by the initial segmentation: 1) For each region r_i, spatial location x_i is computed as centroids of their pixels. If the region is convex, the centroid is inside of it, but if the region is concave, the centroid is situated in the corresponding location of the nearest boundary pixel of that region. 2) For each region r_i, mean intensity μ_i is the arithmetic sum of the intensity of each pixel divided by the amount of pixels of that region. 3) For each pair of nodes, similarity is inversely correlated with the maximum contour energy encountered along the path connecting the centroids of the regions. A large magnitude indicates the presence of an *intervening contour* [2] and suggests that the regions do not belong to the same group.

The core computational technique of the normalized cut algorithm is the eigenvalue problem of Eq. (8). It requires the solution to a large sparse system of symmetric equations. The LANCZOS algorithm provides of an excellent method for approximating the eigenvectors corresponding to the smallest or largest eigenvalues of a sparse matrix.

6. Results and Evaluation

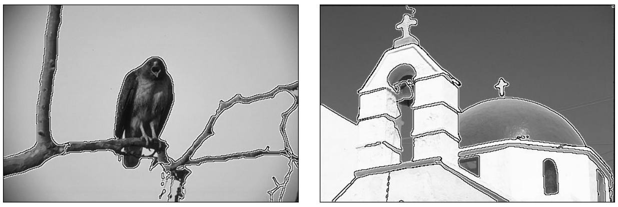
To provide a numerical evaluation measure and, thus, allow comparison with other methods, the algorithm was evaluated against the Berkeley Segmentation Dataset [3]. This database comprises a ground truth of 100 hand-segmented images to compare the segmentation outputs. The task is cast as a boundary detection problem, with results presented in terms of Precision (P) and Recall (R) measures.

In probabilistic terms, precision is the probability that the result is valid, and recall is the probability that the ground truth data was detected. The two statistics may be distilled into a single figure of merit:

$$F = \frac{PR}{\beta R + (1-\beta)\,P} \ , \tag{10}$$

where β determines the relative importance of each term. Following [4], β is selected as 0.5, expressing no preference for either.

A selection of typical results is presented in Fig. 1.



(a) Image 42049

(b) Image 118035

Figure 1: Results overlapped on original images.

The algorithm provides a binary boundary map which is scored against each one of the hand-segmented results of Berkeley Dataset, producing a (R, P, F) value. The final score is given by the average of those comparisons. These quantitative results are summarized in Table 1 for a set of tested images. We identify each image with the id number presented in [3].

It is also interesting to note that from comparison with image 189080 with human results we obtain for one of them $F = 0.77$. The difference in final score come from the disagreement among different human subjects about the correct image segmentation.

Table 1: Results of quantitative experiment in terms of Recall(R), Precision(P) and F-measure. F_H represents the F-measure between hand-segmented results.

images	R	P	F	F_H
24063	0.73	0.72	0.73	0.83
42049	0.85	0.93	0.89	0.92
118035	0.57	0.84	0.68	0.83
135069	0.93	0.87	0.90	0.97
189080	0.74	0.59	0.68	0.85
241004	0.75	0.78	0.77	0.95
296059	0.57	0.85	0.68	0.95

In a set of 10 tested images, the reduction obtained on the number of nodes using the chunk graph G' is about 87% of the nodes in graph G.

7. Conclusion

This paper shows that good segmentation results can be achieved when using a combined approach between region merging and graph-based methods. Using small atomic regions instead of pixels leads to a more natural image representation - the pixels are merely the result of the digital image discretization process, and do not occur in the real world. In comparison to pixel-based methods, the reduction obtained on the number of nodes using chunk graphs is about 87%.

As a future work, we plan to investigate other strategies for generating atomic regions and we will explore how the interaction between this generation and the segmentation components can improve the performance of such a system as a whole.

References

[1] K. Fan, On a Theorem of Weyl Concerning Eigenvalues of Linear Transformations (I), In *Int. Proc. of National Academy of Sciences*, Vol. 35, pp. 652-655, 1949.

[2] J. Malik, S. Belongie, T. Leung, J. Shi, Contour and Texture Analysis for Image Segmentation, *Int. J. of Computer Vision*, Vol. 43, No. 1, pp. 7-27, June 2001.

[3] D. Martin and C. Fowlkes, The Berkeley Segmentation Database and Benchmark, available online at: *http://www.cs.berkeley.edu/projects/vision/grouping/segbench/*

[4] D. Martin, *An empirical Approach to Grouping and Segmentation*, PhD dissertation, University of California, Berkeley, 2002.

[5] M. Meila, The Multicut Lemma, *Tech. Report 417*, Dep. of Statistics, University of Washington, 2001.

[6] J. Shi and J. Malik, Normalized Cuts and Image Segmentation, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 888-905, 2000.

[7] S. Yu and J. Shi, Multiclass Spectral Clustering, *Int. Conf. on Computer Vision*, pp. 313-319, Oct. 2003.