Hierarchical Normalized Cuts: Unsupervised Segmentation of Vascular Biomarkers from Ovarian Cancer Tissue Microarrays

Andrew Janowczyk[1], Sharat Chandran[1], Rajendra Singh[2], Dimitra Sasaroli[3], George Coukos[3], Michael D. Feldman[4], and Anant Madabhushi[5,*]

[1] Dept of Computer Science & Engineering, Indian Institute of Technology Bombay
{andrew,sharat}@cse.iitb.ac.in
[2] Quest Diagnostics, Inc, USA
[3] Ovarian Cancer Research Center, University of Pennsylvania, USA
[4] Dept of Pathology and Lab Medicine, University of Pennsylvania, USA
[5] Dept of Biomedical Engineering, Rutgers University, USA
anantm@rci.rutgers.edu

Abstract. Research has shown that tumor vascular markers (TVMs) may serve as potential OCa biomarkers for prognosis prediction. One such TVM is ESM-1, which can be visualized by staining ovarian Tissue Microarrays (TMA) with an antibody to ESM-1. The ability to quickly and quantitatively estimate vascular stained regions may yield an image based metric linked to disease survival and outcome. Automated segmentation of the vascular stained regions on the TMAs, however, is hindered by the presence of spuriously stained false positive regions. In this paper, we present a general, robust and efficient unsupervised segmentation algorithm, termed Hierarchical Normalized Cuts (HNCut), and show its application in precisely quantifying the presence and extent of a TVM on OCa TMAs. The strength of HNCut is in the use of a hierarchically represented data structure that bridges the mean shift (MS) and the normalized cuts (NCut) algorithms. This allows HNCut to efficiently traverse a pyramid of the input image at various color resolutions, efficiently and accurately segmenting the object class of interest (in this case ESM-1 vascular stained regions) by simply annotating half a dozen pixels belonging to the target class. Quantitative and qualitative analysis of our results, using 100 pathologist annotated samples across multiple studies, prove the superiority of our method (sensitivity 81%, Positive predictive value (PPV), 80%) versus a popular supervised learning technique, Probabilistic Boosting Trees (sensitivity, PPV of 76% and 66%).

1   Introduction

It is estimated [1] that 21,650 women will be diagnosed with and 15,520 women will die of cancer of the ovary (OCa) in 2008. The 5-year survival rates of these women are highly correlated to the early detection of OCa. Recent work [2] suggests that specific tumor vascular biomarkers (TVMs) may be identifiable on OCa TMAs that could have prognostic significance, helping to not only predict this survival rate but also help determine a more specific course of treatment. It has also been suggested that genes expressed uniquely by the vasculature of tumors may provide important therapeutic targets. Biomarkers are typically discovered by staining explicitly for TVMs of interest on OCa TMAs. ESM-1 is one such TVM of interest in OCa, which can be visualized by staining the anti-body to ESM-1 in TMAs with Diaminobenzidine. Precise quantification of the extent and intensity of the stain could serve as a prognostic metric reflecting risk of disease recurrence and patient survival. However, it is currently infeasible in terms of both time and effort for an expert pathologist to perform this segmentation manually.

One of the issues to consider in the choice of a segmentation method is that it should accurately and reliably distinguish true positive (TP) stains from other structures in the tissue that pick up the D-benzidine stain, but are not part of the vasculature. This produces a dataset that suffers greatly when operated on by algorithms that are highly dependent upon initial conditions, such as k-means. Since there is a significant amount of inter-sample and inter-study variability, a more sophisticated algorithm must be employed. In addition, the need to process several hundred cylinders quickly, with high accuracy, is another major requirement. Probabilistic boosting trees (PBT) [3] are popularly employed for classification and segmentation because they have the attractive property that the posterior probability can be used as a threshold to balance between sensitivity and specificity. Additionally, the method is relatively fast. Unfortunately, one of the challenges in constructing a supervised classifier (such as PBTs) is the difficulty in obtaining ground truth segmentations for classifier training for the object, or region, of interest from experts. Creating enough acceptable training samples is both costly and time consuming.

The mean shift (MS) algorithm was originally presented in [4] and revised in [5] as an unsupervised technique aimed at mode discovery for use in place of k-means. MS attempts to overcome its predecessors faults by searching for the mean data point within a certain bandwidth (σ) setting. The improved Fast Gauss Transform [6] implementation of MS allows computation times for large images to become reasonable for use in clinical settings.

The Normalized Cuts (NCut) algorithm is among the final mature descendants of a series of graph cutting techniques ranging from max cut to min cut [7]. It is a popular scheme in spite of its main drawback: the large number of calculations needed for determining the affinity matrix and the time consuming eigenvalue computation. In large sized images the computation and overhead of these border on the infeasible [8], leading to a significant amount of research in avoiding

---

[1] Cancer Facts and Figures, American Cancer Society (ACS), Atlanta, Georgia, 2008.

their direct calculations [9]. A standard approach in the NCut method is to make the assumption that pixels in the same cluster are spatially located near each other, resulting in a sparse matrix. In our specific case, this distance constraint cannot be applied, thereby making a direct implementation of NCut essentially intractable.
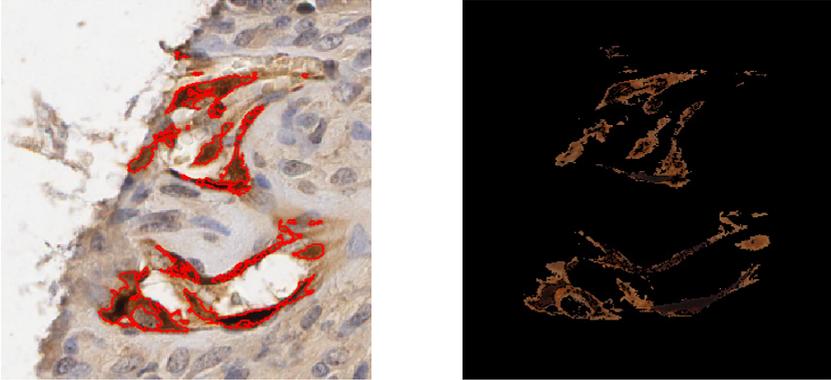

Fig. 1. The input (left) with annotated ground truth in red. Notice the difficulty in spotting stained cells with the untrained naked eye. Segmentation results (right) from HNCut.

2   Novel Contributions of This Work

The major contribution of this work is a novel, hierarchical unsupervised segmentation method termed HNCut. Additionally, to the best of our knowledge the application of HNCut to precisely quantify a vascular marker on OCa TMAs represents the first attempt at creating a quantitative, image based metric for OCa prognosis and survival. The novelty of HNCut is the use of a hierarchically represented data structure that bridges MS and NCut. The first component of HNCut, namely MS, efficiently prunes down the total number of clusters for the second component NCut. Unlike traditional clustering algorithms, our final aim is to extract a single cluster pertaining to the stained region (Fig. 1), while ignoring pixels in all other clusters. In traditional algorithms the pixels that normally should be removed are usually allocated to the least dissimilar cluster. In contrast, our setup encourages cuts that confidently trim away these undesired pixels. The unsupervised aspect of HNCut is particularly desirable for image analysis applications in histopathology where obtaining annotated samples for training a supervised classifier depends on the annotations provided by an expert and hence difficult to obtain. To summarize, the primary methodological contributions of this work are,

- A novel hierarchical segmentation approach (HNCut) that marries MS and NCut. The affinity matrix can now take advantage of multiple features, and multiple color spaces efficiently across large windows sizes.

■ Parameter insensitive segmentation for large images and the ability of HN-Cut to discriminate between regions with similar color values. The parameter for the Gaussian kernel in the affinity matrix of NCut is automatically computed. The parameters for MS are automatically adjusted based on the variance of the output.

3   Hierarchical Normalized Cut

As mentioned earlier, our method has two broad components consisting of MS and NCut. The intuition for these are first presented, followed by the complete details in Section 3.4.

3.1   Mean Shift for Reducing Color Space Cardinality

MS is used to detect modes in data using density gradient estimation. With an image defined as $\mathcal{C} = (C, f)$ where C is a 2D Cartesian grid of N pixels, $c \in C = (x, y)$ representing the Cartesian coordinates of a pixel, and f a color intensity function associated with c, we have the fixed point iteration update $\forall c \in C$ in MS as

$$f_{k+1}(c) \leftarrow \frac{\sum_{i=1}^{N} G_{\sigma_{\mathrm{MS}}}(f_k(c) - f_k(c_i)) f_k(c_i)}{\sum_{i=1}^{N} G_{\sigma_{\mathrm{MS}}}(f_k(c) - f_k(c_i))} \tag{1}$$

where a Gaussian function, G, with a bandwidth parameter σ_{MS}, is used to compute the kernel density estimate at data point c. $k \in \{1, \ldots, K\}$ represents various levels of color resolution produced at each iteration. The overall computation time for Eq. 1 is $O(N^2)$. By employing the improved Fast Gauss Transform [6], we can reduce the complexity to $O(N)$ with minimal precision loss.

MS produces a feature-based pyramidal scene representation $\mathcal{C}_u = (C, f_u)$, where $u \in \{1, \ldots, K\}$ represent levels of the pyramid of height K. This results in a series of scenes $\mathcal{C}_u$, all mutually aligned, but with a fewer number of colors in the lower levels of the pyramid compared to the top. It is the fewer colors at the lower levels that enable NCut to be tractable, however, the upper levels of the pyramid are needed for high quality segmentation.

3.2   Normalized Cuts on Mean Shift Reduced Color Space

By setting the vertices (V) to the basins of attraction from the MS (i.e., unique color values), NCut can segment the data by representing it as a connected graph $(\mathbf{G}=(E, V))$, with edges (E) representing affinity or strength of connectivity. A cut is the processes by which the removal of edges leads to two partitions. A value of a normalized cut between two disjoint sets A and B is computed using

$$\mathrm{NCut}(A, B) = \frac{\mathrm{cut}(A, B)}{\mathrm{assoc}(A, V)} + \frac{\mathrm{cut}(A, B)}{\mathrm{assoc}(B, V)} \tag{2}$$

where $\text{cut}(A, B) = \sum_{c_i \in A, c_j \in B} w(f(c_i), f(c_j))$, $i, j \in \{1, \ldots, N\}$. Additionally, $\text{assoc}(A, V) = \sum_{c_i \in A, c_k \in V} w(f(c_i), f(c_k))$, where $i, k \in \{1, \ldots, N\}$. Also, $A \cup B = V$ and $w(f(c_i), f(c_j))$ is a function used to compute an affinity measure between c_i and c_j. Normally, in NCut, an additional spatial constraint is introduced such that if $||c_i - c_j|| \geq \theta$, where θ is a spatial radius threshold, $w(f(c_i), f(c_j))$ is set to 0. Because of this spatial constraint, the affinity matrix $\mathbf{W}$ is normally sparse, making the method tractable. The goal of NCut is to identify partitions A, B such that the they have the most intragroup similarity, and the most intergroup dissimilarity. This process can be recast [8] and solved as a generalized eigenvalue system.

In our implementation of NCut we are concerned with partitioning the color space and not the image pixels per se. Hence if the process of MS in Section 3.1 on $\mathcal{C}$ eventually results in M unique colors, V is the set of vertices corresponding to these colors. Thus $\mathbf{W} \in \mathbb{R}^{\mathbf{M} \times \mathbf{M}}$ in the first application of NCut.

3.3   Integrating Domain Knowledge to Guide Normalized Cuts

Our goal with NCut is to discard A or B. We therefore define a swatch (color template) reflecting the attributes of the object of interest in the scene. We define $Q = \{f_{\alpha_1}, f_{\alpha_2}, \ldots, f_{\alpha_t}\}$ to represent the colors of the objects we seek. Note that Q is trivially determined by annotating (manually) a few pixels from the object of interest on a representative image and may be easily changed based on the application. Hence HNCut is still an unsupervised algorithm.

3.4   The HNCut Algorithm

Reducing color space cardinality with MS enables NCut to be tractable (irrespective of the distance constraint). However, the first application of NCut resulting in the sets A and B leaves us with too many false positives. We can overcome this draw back by maintaining a pyramid as we perform MS, as detailed below.
Step 1: <u>Generate a hierarchical color pyramid via MS:</u> For any scene $\mathcal{C}$, corresponding scenes in the pyramid $\{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$ of monotonically increasing resolutions $\{M_1, M_2, \ldots, M_K\}$ are generated.
Step 2: <u>Apply NCut on lowest level of pyramid:</u> Apply NCut on $\mathcal{C}_K$ to partition the scene into two color sets A_K and B_K. Use in the affinity matrix the values $\mathbf{W}_K(\beta_1, \beta_2) = \exp(-\frac{||\beta_1 - \beta_2||^2}{\sigma})$ where σ is a scaling parameter.
Step 3: <u>Use color swatch to identify unique color partition:</u>   Identify which of A_K and B_K uniquely contains colors in the swatch Q. If $Q \subset A_K$ and $Q \cap B_K = \emptyset$ then eliminate B_K. If $Q \subset B_K$ and $Q \cap A_K = \emptyset$, similarly eliminate A_K. However if $Q \cap A_K \neq \emptyset$ and $Q \cap B_K \neq \emptyset$ then set $\sigma = \sigma + \Delta\sigma$ and repeat NCut on V_K to obtain a new partition. Keep incrementing σ until Q is uniquely contained within either of A_K or B_K.
Step 4: <u>Repeat color partitioning at a fixed scale using NCut:</u> Assuming for some value of σ, Q is uniquely contained in A_K, set $V = A_K$ and repeat

NCut on V. Repeat steps 2 and 3 until no further partitioning of color space at scale $\ell = K$ is possible.

Step 5: Repeat NCut and color partitioning at higher image scales: Migrate to the next higher image resolution, First, identify V as the set of colors at resolution $\ell = K - 1$ but derived from the non-decomposable partition $\ell = K$. Repeat steps 2-4 to identify a unique non-reducible color partition at $\ell = K - 1$ that contains the swatch. Also, do this step for subsequent lower values of $\ell \in \{1, \ldots, K - 2\}$

4   Experiment Setup and Methodology

Our database comprises of a total of seven digitized TMAs of OCa, in turn comprising over 500 tissue cylinders. The TMAs were obtained by sampling OCa tissue from over 100 patients and were stained for the presence of the TVM ESM-1, resulting in vascular regions with the antibody to ESM-1 staining brown. The TMA files are down-sampled and stored at 20x magnification, producing images that are approximately 1300 x 1400. An expert pathologist annotated 100 cylinders. The exact regions highlighted by him were extracted and stored separately, creating a binary map used as the ground truth.

Performance Metrics: Two different metrics were used to grade the performance of the algorithms. A macro metric (region) matched minimum surface area ellipses fit to both the test method and the ground truth to assess the correctness of region matching. A micro metric (pixel) was used to determine the amount of overall correctness of pixel matching.

4.1   Comparison of HNCut to k-Means, PBT

All algorithms loaded the same image and converted it to YCbCr space. To ensure fair comparisons, each algorithm operated only with chromatic information.

k-means. Ten clusters were employed. The initial cluster centers were manually chosen, by experimentation, to provide a good representation of the image.

Probabilistic Boosting Tree. PBT was implemented as described in [3], using suggested default values for both θ and ϵ (.45 and .4, respectively). Each strong Adaboost classifier was constructed using seven weak classifiers. The PBT performed seven levels of dataset separation, in the prescribed tree like manner. The training set was created by taking a 3×3 window around every $c \in C$, across all 3 color channels in YCbCr space, resulting in a 27 dimensional vector. 1000 random positive (stained) samples and 1000 random negative (unstained and spuriously stained) samples were selected from 25 randomly selected images, resulting in a total training vector of size 27x50,000. 50 cross validation iterations took place resulting in an average AUC of .9296 with a standard deviation of .0039. The probability returned by the PBT was converted into a strong classifier by taking the upper Otsu threshold.

HNCut. MS was performed using a $\sigma_{\mathrm{MS}} = .05$, with the number of clusters for the improved fast Gauss transform set to 350. NCut uses the Silverman function [10] to determine its initial σ and then increases by a factor of 10 as prescribed above. The domain knowledge is six pixels of shades of brown that we find acceptable as stain.

5   Results and Discussion

The first column in Fig. 2 represents the original input image, with the boundary of the ground truth highlighted by the pathologist labeled in red. The first row illustrates a case where all of the algorithms performed as expected. The second row illustrates an example where the HNCut algorithm performs optimally, while both the PBT and k-means extract many miscellaneous false positives. The hierarchical set of operations in HNCut is essentially a feedback loop that enables the algorithm to operate extremely efficiently and accurately. The final row is used to illustrate a scenario where false negatives (FN) occur for all three methods. The middle region is correctly segmented in all algorithms, while the three other regions are incorrectly rejected. The stain in those regions
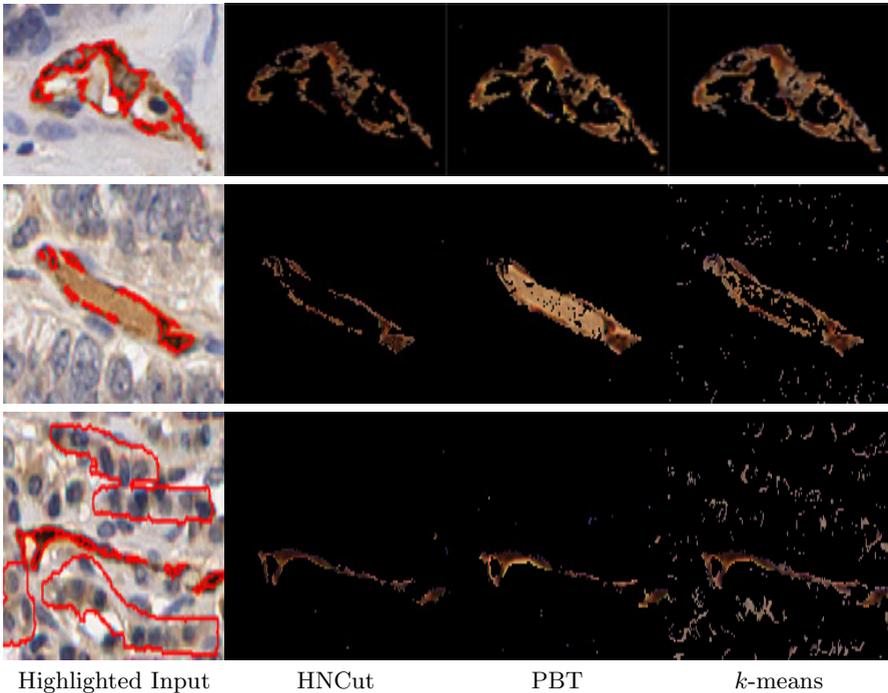

| Highlighted Input | HNCut | PBT | k-means |

Fig. 2. Selected examples are presented. The original input, with the annotated ground truth in red, is presented on the left, followed by HNCut, PBT and k-means.

Table 1. Performance measures for HNCut, PBT and k-means

	Region					Pixel	
	FN	TP	FP	Sensitivity	PPV	Sensitivity	Specificity
HNCut	217	920	224	80.9%	80.4%	73.8%	99.7%
PBT	265	869	456	76.6%	65.6%	67.4%	99.6%
k-means	264	873	1646	76.8%	34.7%	69.1%	98.3%

is only barely visible to an expert. Lastly, we can see that although k-means does the best in attempting to detect those regions, the same setup results in many false positives in other images. This is a result of the k-means requiring all pixels to be assigned to a cluster, filling the stain cluster as it is the most similar out of the candidates. Both PBT and HNCut determine, in most instances, that these same values are simply too different to be considered as stain.

Table 1 quantitatively illustrates that HNCut provides the best balance between sensitivity and specificity, providing 50% less false positives (FP) than PBT and 86% less FP than k-means. All of the algorithms seem to do very well in the specificity category, but this is a result of the vast number of true negative pixels (TN) associated with the majority of the sample.

Using HNCut on 500 discs, about 10 of them failed to converge properly, resulting in very poor segmentations. Interestingly, these 10 images all had little to no stain present. This failure of convergence is due to the σ_{MS} for the MS being inappropriately selected, in all experienced cases being too wide. The automatic adjustment of the σ_{MS} took place, as described above, to a narrower setting resulting in appropriate results in all of the experiments we ran.

6   Concluding Remarks

In this paper we have presented a novel unsupervised segmentation scheme called Hierarchical Normalized Cuts (HNCut). The strength of HNCut is derived from the fact that it integrates the best of both the Mean Shift and the Normalized Cut algorithms. The presented algorithm is not only efficient and accurate but not encumbered by the need for precisely annotated training data. While HNCut could be applied to any general segmentation task, in this paper we evaluate it in the context of identifying a vascular biomarker in OCa TMAs. Results on 100 images reveal that HNCut significantly outperforms both Probabilistic Boosting Trees and k-means. The algorithm does use a color swatch but this involves manual identification of half a dozen pixels to represent features of interest. The swatch can be changed to suit similar biomedical segmentation problems swiftly. The swatch does not require laborious, careful annotation by an expert, but is a small sampling of the target class and may be provided by any user who is able to partially identify the target class.

References

1. Doyle, S., Madabhushi, A., Feldman, M.D., Tomaszeweski, J.E.: A boosting cascade for automated detection of prostate cancer from digitized histology. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 504–511. Springer, Heidelberg (2006)
2. Buckanovich, R., Sasaroli, D., et al.: Tumor vascular proteins as biomarkers in ovarian cancer. Journal Of Clinical Oncology, 852–861 (March 2007)
3. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: ICCV, pp. 1589–1596 (2005)
4. Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Information Theory 21(1), 32–40 (1975)
5. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. PAMI 24(5), 603–619 (2002)
6. Yang, C., Duraiswami, R., Gumerov, N., Davis, L.: Improved Fast Gauss Transform and efficient kernel density estimation. In: ICCV, pp. 664–671 (2003)
7. Wu, Z., Leahy, R.: An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. IEEE Trans. PAMI 15(11), 1101–1113 (1993)
8. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. PAMI 22(8), 888–905 (2000)
9. Dhillon, I.S., Guan, Y., Kulis, B.: Weighted graph cuts without eigenvectors a multilevel approach. IEEE Trans. PAMI 29(11), 1944–1957 (2007)
10. Silverman, B.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, Boca Raton (1986)