

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

Segmentation on Ovarian Cancer Tissue Microarrays using FWMS and NCuts Algorithm

Hemalatha G, Sundararajan.M, Arulselvi S

Assistant professor, Department of CSE, Bharath University, Chennai, India

Director, Research Center for Computing and Communication, Bharath University, Chennai, Tamil Nadu, India

Co-Director, Research Center for Computing and Communication, Bharath University, Tamil Nadu, India

ABSTRACT: Quantifying the presence and extent of staining on account of a vascular biomarker on tissue microarrays is a laborious, time consuming and error prone job. We demonstrate our flexible, robust, accurate, and high-throughput minimally supervised segmentation algorithm, termed hierarchical normalized cuts (HNCuts) for the specific problem of quantifying extent of vascular staining on ovarian cancer tissue microarrays.. In this paper we have addressed this need by using Hierarchical Normalized cuts algorithm which is driven by the use of a hierarchically represented data structure. This algorithm merges two powerful image segmentation algorithms namely, Frequency weighted mean shift for supervised clustering of the images and Normalized cuts for graph partitioning. We have designed a system for high throughput in computing and detecting staining of ovarian cancer(OCa) on a very large pathology in less time.. HNCut could be a very powerful tool in digital pathology and bioinformatics applications where it could be used to facilitate computer-assisted prognostic predictions of disease outcome.

KEYWORDS: Biomarker, Image Segmentation, Supervised clustering, Graph partitioning, Tissue microarray(TMA), Hierarchical normalized cuts(HNCuts).

I.INTRODUCTION

In the field of medical imaging, image segmentation algorithms are widely used in the creation and segmentation of images of the human body such as tissues and organs.This helps in revealing,diagnosing and examining them for treatment planning and various other clinical purposes. The use of tissue microarrays(TMA) technology has made it now possible to simultaneously stain several hundred tissue sections for the presence of various biomarkers. Accurate identification of the extent and intensity of the stain could help predict risk of disease recurrence and patient survival.Most previous computerized image analysis algorithms proposed for tissue microarrays have involved thresholding based schemes. Methods involving threshold are known to be highly sensitive to light and changes in color and illumination.Clustering based approaches like k-means,active contour schemes and supervised learning methods such as probabilistic boosting trees (PBT) have been investigated for the analysis of tissue microarrays.These approaches are not particularly suited for pixel level detection and classification.In our paper we have addressed all these issues with the use of a supervised segmentation algorithm Hierarchical Normalized cuts HNCut and its specific application to a problem of automated quantification of stain extent associated with a vascular marker on OCa TMAs. HNCut only requires specifying a few representative pixels from the target class and, unlike more traditional supervised classification algorithms, does not require more detailed target object annotation.Its efficiency involves in combining a supervised technique,the frequency weighted mean shift(FWMS) and the normalized cuts (NCuts)for partitioning of the images.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

II.RELATED WORKS

The HNCuts[7][8] is not a new approach as it was already implemented with mean shift(MS) and normalized cuts(NCuts)[5][6].But normalized cuts is known for its major drawback that it involves a large computational overhead.Inorder to overcome this,the mean shift was merged with NCuts[3][4].But later it was noticed when pixels of similar values are within the neighborhood of each other, their contribution to the overall system can be merged.This will lead to a reduction in the number of computations needed per iteration.In our paper by making use of FWMS,the convergence is addressed making it a novel scheme for clustering.Also this achieves an improvement in the speed of the NCuts algorithm.The various schemes which are in use are discussed to analyse the drawbacks in them.

2.1 k-means Clustering

It is a method of cluster analysis[9][10] which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.The two key features of k -means that make it efficient are often regarded as its biggest drawbacks.Euclidian distance is used as a metric and variance is used as a measure of cluster scatter.The number of clusters k is an input parameter(i.e) an inappropriate choice of k may yield poor results. That is why, when performing k -means, it is important to run diagnostic checks for determining the number of clusters in the data set.Also convergence to a local minimum may produce wrong results.

2.2 Mean Shift Clustering

The mean shift algorithm[5] has been employed and modified as an unsupervised technique for mode discovery instead of k -means.It attempts to identify cluster mean within a predefined bandwidth. This paper attempted to provide an appropriate generalization to the mean shift algorithm, so that many interesting and useful properties would be preserved. One property is that the process is either a gradient mapping, or a similar one that seeks modes of a real function. There may be computational obstacles,and they may not be suitable for problems with prohibitive sizes and dimensionalities. The computational cost of an iteration of mean shift is $O(n^2)$ where n is the size of S , the data set.

2.3 Contour Schemes

A scheme for object boundary detection was proposed based on geodesic or minimal path computations called geodesic active contour model[1][2]. This approach also gives possible connections between classical energy based deformable contours and geometric curve evolution ones. Such active contour schemes while suitable for cell and nuclear segmentation in digital pathology are not ideally suited to the problem of pixel level classification.Additionally they are typically infeasible for problems where hundreds of objects need to be concurrently segmented on very large images.

2.4 Probabilistic Boosting Trees

The probabilistic boosting-tree (PBT), was proposed for learning two-class and multi-class discriminative models. In the learning stage, the probabilistic boosting-tree automatically constructs a tree in which each node combines a number of weak classifier into a strong classifier (a conditional posterior probability). Supervised learning methods like PBT[2][3] are constrained by the difficulty in obtaining ground truth segmentations from experts for classifier training of the object of interest.Manual annotation of the data,apart from being time consuming and laborious can also be expensive if only a medical partitioner can provide accurate annotations. Additionally if the target of interest changes, considerable effort might be required to generate new annotations.

2.5 Normalized Cuts

Normalized cut[4][5] is an unbiased measure of disassociation between sub groups of a graph and it has the nice property that minimizing normalized cut leads directly to maximizing the normalized association,which is an unbiased

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

measure for total association within the sub groups. It is a popular scheme in spite of its main drawbacks, the large number of calculations needed for determining the affinity matrix and the time-consuming eigenvalue computations.

III. PROPOSED WORK

The proposed algorithm combines frequency weighted mean shift (FWMS) and NCuts algorithm and is specifically designed for the rapid extraction of pixels of interest which is insensitive to the user's domain knowledge. The FWMS unlike the mean shift algorithm works directly in the color space merging points of similar values that lie within the neighbourhood of each other and exhibits novelty in convergence allowing faster computation with few steps.

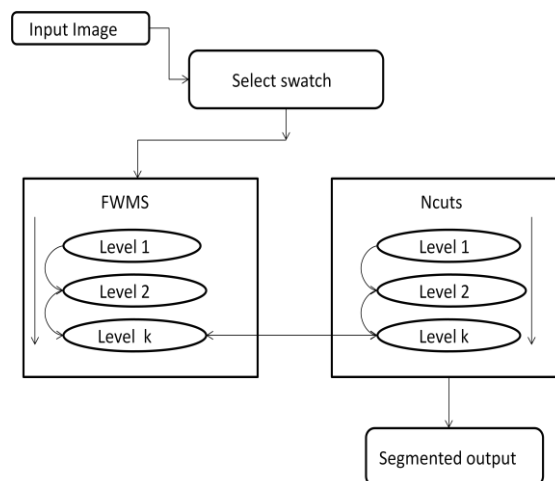


Fig.1 Flow Chart of HNCut Process

The user selects the domain swatch. This is followed by the FWMS of the image. This results in the original image being decomposed into multiple levels of color resolution, which is then followed by the application of the NCut at each of the color resolutions generated. At each pyramid level, colors that are not a part of the swatch are eliminated. Following the application of NCut on the color pyramid from the lowest to the highest color resolution, the color values that have not been eliminated are mapped back to the spatial domain via their original pixel locations, and the final segmentation is obtained.

IV. FUNCTIONAL MODULES

4.1 High Level Data Flow

The high level data flow is depicted in the diagram below. From the database containing the target images the image needed for the study is selected by the user. This image is now identified as the target image and is given as input to the system presented.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

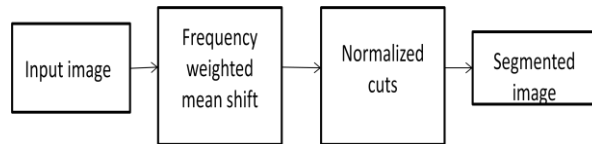


Fig.2 High level Flow Diagram

The user initially selects a few sample pixels from the target class of an image. These pixels are used to guide the subsequent pixel classification process across all images. Then the Mean shift algorithm is employed on the color values in the image to form a hierarchical data structure. Now the FWMS algorithm allows for identification of color values within some specified tolerance of each other and assigns them to the same mode. Employing the NCuts operation only on the unique values at each level of the pyramid, as opposed to all color values, allows for a factorization resulting in fewer computations.

4.2 Frequency Weighted Mean Shift

Similar color values are identified and are assigned to the same mode. Then the weight for each unique node is calculated. This reflects the actual frequency of the number of pixels associated with it. After applying the FWMS algorithm the original image gets decomposed into multiple levels of color resolution. The convergence requirement may be specified by three possible criteria. The first is the maximum number of iterations, a number specified by the user. The second approach is to stop the algorithm when the difference between any two iterations falls below a predefined threshold. Finally, convergence can be reached when the number of elements in the color vector becomes small enough that additional clustering provides no efficiency benefit as the overhead in the NCuts starts to outweigh the computation time. Thus the overall color resolution is significantly reduced as the algorithm proceeds from level 1 to level k.

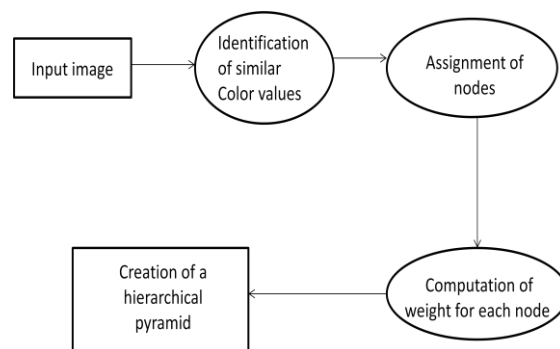


Fig.3 Frequency Weighted Mean Shift

A. Symbols Used

c	2D Cartesian grid of pixels
C	Image Scene
F	Color Intensity Function
F ₁	Color Vector at Level 1
F _k	Color Vector at Level k
W _k	Weight vector at Level k
N	Number of pixels in C

Table.1 Description of symbols in FWMS

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

B. Algorithm of FWMS

An image scene is defined as $C=(c,f)$ where $c \in C$ at full color resolution. Elements of F_1 namely $f_{1,i}$ are derived in such a way that $c_i, f_{1,i}=f(c_i)$.

*Step1:*Select swatch (few pixels) S_k at $k=1$ where $k \in \{1, \dots, K\}$, various levels of color resolution.
*Step2:*Determine unique values in F_k . *Step3:*Construct color vector \hat{F}_k where $\hat{F}_k \in F_k$ where \hat{F}_k is the unique set of values in F_k . *Step4:*Compute the weight vector W_k for F_k , summing the weights from the previous level into new values that resulted from the next iteration. *Step5:*FWMS creates a pyramidal scene with $C_k=(C, F_k)$, series of scenes C_k with smaller number of colors compared to $\{C_1, C_2, \dots, C_k\}$. The result of FWMS is the color vectors $(\hat{F}_1, \hat{F}_2, \dots, \hat{F}_k)$. Therefore from level 1 to level k the overall color resolution reduces significantly.

C. Convergence

Consider two points C_{β_1}, C_{β_2} where β_1 and $\beta_2 \in \{1, \dots, N\}$ and $|f_{k,\beta_1} - f_{k,\beta_2}| \leq e$ where e is the predefined threshold value assigned by the user. The requirements for convergence are:

- The maximum number of iterations is the number specified by the user.
- Stop the algorithm when the difference between any 2 iterations falls below a predefined threshold.
- Convergence is said to have reached when the number of elements in \hat{F} becomes small enough that additional clustering provides no benefit.

4.3 Normalized Cuts

By starting at the bottom of the pyramid, the unique values are partitioned. The Ncut is applied to partition the scene into two disjoint color sets A and B. An affinity matrix is calculated to perform the partitions. All values selected by the user in the first step are assigned to the first partition. Subsequently the second partition is eliminated and the colors in the first partition are mapped to an immediately higher color resolution level in the pyramid. This process is continued until the entire pyramid is traversed. The last step involves mapping the color values not eliminated back into the spatial domain.

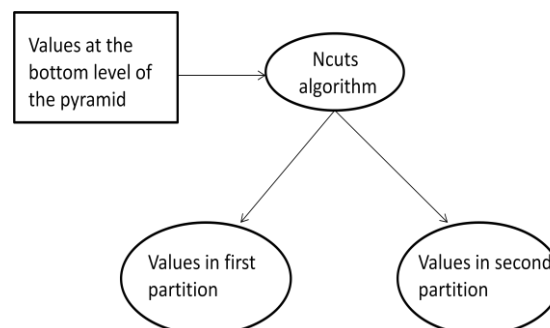


Fig.4 Normalized Cuts

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

A. Symbols Used

Ψ	Symmetric matrix comprised of $\psi(c_i, c_j)$
$\psi(c_i, c_j)$	Affinity measure between c_i, c_j
V	Vertices
E	Edges
W	Weight

Table.2 Description of symbols in NCuts

B. Algorithm of NCuts

Step 1: The input to this module is the color vectors $(\hat{F}_1, \hat{F}_2, \dots, \hat{F}_k)$.

Step 2: It takes a connected graph $G=(E,V,W)$ and separates vertices into disjoint groups. Of similar color values. The edges describe the similarity between the color values. The $\Psi (\hat{f}_{k,i}, \hat{f}_{k,j})$ function is used to define the affinity between two points. *Step 3:* Starting at the lowest level of the pyramid, it removes the edges which leads to two partitions. These two color sets can be named A and B.

Step 4: Identify if either A or B contains uniquely all colors in the swatch defined S_k .

- a) If S_k belongs to A and $S_k \cap B = \emptyset$ then eliminate all colors in B setting $V_k=A$.
- b) If S_k belongs to B and $S_k \cap A = \emptyset$ then eliminate all colors in A setting $V_k=B$.
- c) If S_k not uniquely contained in both A and B increase the bandwidth parameter σ_{NCut} .

Step 5: Increment σ_{NCut} until S_k is uniquely contained in A or B, eg. by a factor of 10 and set V_k to that partition.

Step 6: Begin the process again until no further partitioning is possible at level k.

Step 7: Similarly traverse the entire pyramid until k-1 and repeat the process again from step 1 until k=1. At level 1, V_1 contains a subset of values from \hat{F}_1 .

Step 8: The final image is computed by retaining all pixels $j \in \{1, \dots, N\}$ such that $f_{1,j} \in V_1$ and eliminate all others.

$$\text{Cut}(A,B) = \sum_{u \in A, v \in B} w(u, v) \quad (1)$$

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

Where $\text{aasoc}(A,V)$ defines the total connection from nodes in A to all nodes in the graph and $\text{assoc}(B,V)$ defines the total connection from nodes in B to all nodes in the graph.

V.CONCLUSION AND FUTURE WORK

In this paper we have presented a minimally supervised segmentation scheme for quantifying the extent of staining in OCa TMAs. By using our newly presented combination of FWMS and NCuts, different target classes can be extracted based on the user input. Faster extraction of pixels and handling larger images are the two major advantages of the proposed scheme.

In future work, the applicability of HNCut to other biomarker quantification is to be studied. Also this method is easily extensible to incorporate image features such as texture other than just the chromatic information.

REFERENCES

[1] Cohen L.D and Kimmel.R,(1997) "Global minimum for active contour models", IEEE International Conference pp 57-78.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

- [2] Sundar Raj M., Arkin V.H., Adalarasu, Jagannath M., "Nanocomposites based on polymer and hydroxyapatite for drug delivery application", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S5) (2013) pp.4653-4658.
- [3] Xu.J,Janowcyk,Chandran.S,(2010) "A weighted mean shift,normalized cuts initialized color gradient based geodesic active contour model",IEEE Transactions pp 136-166.
- [4] Vijayaprakash S., Langeswaran K., Gowtham Kumar S., Revathy R., Balasubramanian M.P., "Nephro-protective significance of kaempferol on mercuric chloride induced toxicity in Wistar albino rats", Biomedicine and Aging Pathology, ISSN : 2210-5220, 3(3) (2013) pp.119-124.
- [5] Tiwari.P,Rosen and Madhubushi.A,(2009) "Probabilistic boosting tress",IEEE International Conference pp 844-851.
- [6] Khanaa V., Thooyamani K.P., Saravanan T., "Simulation of an all optical full adder using optical switch", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S6)(2013) pp.4733-4736.
- [7] Shi.J,Malik.J,(2000) "Normalized cuts and image segmentation",IEEE Transactions pp 888-905.
- [8] Kumar S., Das M.P., Jeyanthi Rebecca L., Sharmila S., "Isolation and identification of LDPE degrading fungi from municipal solid waste", Journal of Chemical and Pharmaceutical Research, ISSN : 0975 – 7384 5(3) (2013) pp.78-81.
- [9] Cheng.Y,(1995) "Mean shift,mode seeking and clustering",IEEE Transactions pp 790-799.
- [10] Laljee R.P., Muddaiah S., Salagundi B., Cariappa P.M., Indra A.S., Sanjay V., Ramanathan A., "Interferon stimulated gene - ISG15 is a potential diagnostic biomarker in oral squamous cell carcinomas", Asian Pacific Journal of Cancer Prevention, ISSN : 1513-7368, 14(2) (2013) pp.1147-1150.
- [11] Tao.W,Jin.H,(2007) "Color image segmentation based on mean shift and normalized cuts"IEEE Transactions pp 1382-1389.
- [12] Chen.W,Reiss.M and Foran.D.J,(2004) "A prototype for unsupervised analysis of tissue microarrays for cancer research and diagnostics"IEEE Transactions pp 89-96.
- [13] Chandran.S,Singh.R and Madhubushi.A ,(2009) "Hierarchical Normalized cuts: Unsupervised segmentation of vascular biomarkers",IEEE Transactions pp 230-238.
- [14] Jennsen.R,Erdogmus.D,(2008) "Mean shift spectral clustering",IEEE Transactions pp 1924-1938.
- [15] Zhang.J,Zhang.Z,(2008) "Improved k-means clustering",IEEE Transactions pp169-172.
- [16] Jemima Daniel,Language Teaching in the Digital Age,International Journal of Innovative Research in Science, Engineering and Technology,ISSN: 2319-8753,pp 11029-11031, Vol. 3, Issue 4, April 2014.
- [17] Jemima Daniel,Importance of Group Discussions,International Journal of Innovative Research in Science, Engineering and Technology,ISSN: 2319-8753,pp 9081-9084,Vol. 3, Issue 2, February 2014.
- [18] Jemima Daniel,'The Playboy of the Western World' As a Tragi-Comedy, International Journal of Innovative Research in Science,Engineering and Technology,ISSN: 2319-8753,pp 10379-10381,Vol. 3, Issue 3, March 2014.
- [19] Jemima Daniel,Techniques Used in Teaching English, International Journal of Innovative Research in Science, Engineering and Technology,ISSN: 2319-8753,pp 8791-8793,Vol. 3, Issue 1, January 2014.
- [20] M. Santhi & Dr. A. Mukunthan,A Detailed Study of Different Stages of Sleep and Its Disorders – Medical Physics,International Journal of Innovative Research in Science, Engineering and Technology ,ISSN: 2319-8753, pg 5205-5212 , Vol. 2, Issue 10, October 2013.
- [21] M.NAGESHWARI, Dr.A.MUKUNTHAN , C.RATHIKA THAYA KUMARI,A Study of Surface Ozone Measurement at Vadasery, Kanyakumari District,International Journal of Computer & Organization Trends (IJCOT),ISSN: 2319-8753, pp 160-165,Vol. 1, Issue 2, December 2012.