Multidimensional Shape/Color Distribution for the Prediction of Lymph Node Metastasis Status

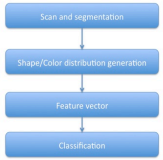
Md. Alimoor Reza[1], Aladin Milutinovic[2], Robi Polikar[3], Fernando U. Garcia[2], David E. Breen[1].
1. Department of Computer Science, College of Engineering, Drexel University, Philadelphia, PA, USA. 2. Pathology Department, Drexel College of Medicine, Drexel University, Philadelphia, PA, USA.
3. Department of Electrical and Computer Engineering, Rowan University

Motivation

The goal of this project is to develop computational techniques for analyzing histology images of breast cancer tumors. The computational techniques derive shape and color information from the images and will enable automated evaluation of the tumor. Specifically, the techniques will be used to determine if a patient's breast cancer has spread to nearby lymph nodes by examining a primary tumor that has been excised from the patient. The image analysis capability will obviate the need for exploratory surgical removal of lymph nodes; thus eliminating the associated side effects, e.g. pain, swelling and morbidity, and cost.

Flow of Work

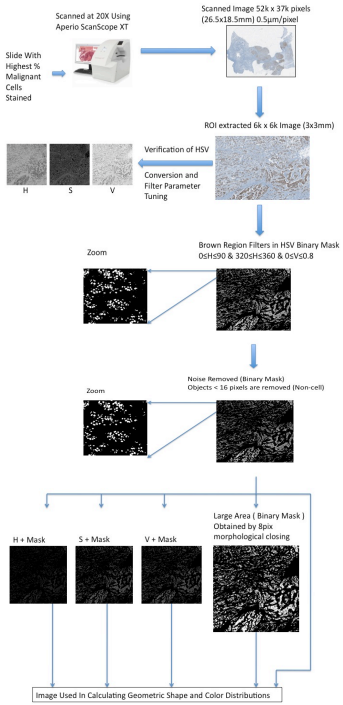
The whole process can be broken into four steps of activities. Raw images are segmented first. Then raw distributions are generated from the segmented images using various geometric and color metrics. Some statistical information are extracted from the raw distribution that form the feature vector. And finally classification technique is applied on the feature vectors.



Step 1: Scan and Segmentation

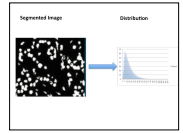
The segmentation method is based on computational examination of a routinely applied prognostic panel that uses immunohistochemistry and heatoxylin & eosin (H&E) staining of 100 invasive breast cancer carcinomas with known lymph node metastasis status. This panel includes: estrogen and progesterore receptors, MIB-1(proliferative activity), mutated p53 and HER2/neu. The highest staining marker from the panel was selected for further digital image analysis.

Segmentation: Identifying Brown Stained Nuclei

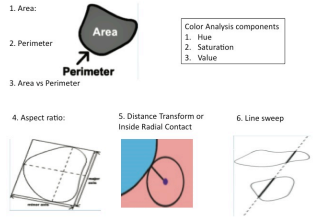


Step 2: Shape/Color Distribution generation

Several geometric measures are applied to each segmented image to generate distributions for each sample.



Various Geometric and Color Measures



Step 3: Feature Vector

Ten Distributions using geometric measures

1. Area
2. Perimeter
3. Area vs Perimeter
4. Aspect Ratio
5. Distance transform log
6. Distance transform
7. Line sweep log
8. Line sweep
9. Large area distance transform log
10. Large area line sweep log

Three Color distributions

1. Hue
2. Saturation
3. Value

These 13 normalized distributions are generated for each sample. 10 statistics can be extracted from each of these distributions. Thus feature vectors are generated for each sample and each feature is a statistic from its original normalized distribution. 13 statistical feature vector for each sample.

Ten statistics

1. Mean
2. Median
3. Mode
4. Standard Deviation
5. Skew
6. Kurtosis
7. Maximum bin size
8. Maximum bin value
9. Mean of bin value
10. Standard Deviation of bin value

Feature Vector: Continued

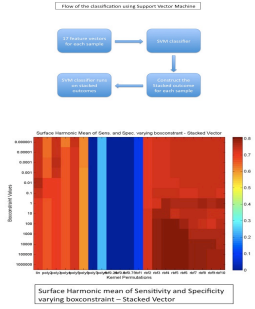
4 Average distribution

1. Hue average
2. Saturation average
3. Value average
4. Aspect ratio average

4 more feature vectors are derived by averaging Hue, Saturation, Value and Aspect ratio distribution.
Thus total number of feature vectors are 13 statistical + 4 average = 17.

Step 4 : Classification

Once the 17 feature vectors for each sample are derived we feed those feature vectors into a supervised classifier, Support Vector Machine (SVM). We have 100 samples. We used the Leave-One-Out approach for classification. In an iterative way each sample is selected for classification and the remaining 99 samples are used for training the SVM classifier. An SVM classification is done for each vector and for each sample. Thus for each sample we have 17 binary outcomes, the computed classification based on each feature vector. These 17 stacked outcomes per sample are again fed into the SVM classifier to find an improved classification output. Leave-One-Out is employed again for the final classification of the samples.



Classification: Continued

$$Sensitivity = \frac{TP}{TP + FP}$$

$$Specificity = \frac{TN}{TN + FN}$$

TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

$$HarmonicMean = \frac{2 \times Sensitivity \times Specificity}{Sensitivity + Specificity}$$

N1+ lymph node is considered positive and N0 lymph node is negative

Result

We analyzed 100 samples, 45 samples are lymph node positive (N1+) and 55 samples are lymph node negative (N0). The stacked SVM classification gives us a sensitivity of 0.80 and specificity of 0.82. Out of the 45 positive lymph nodes 36 are correctly classified. Out of 55 negative samples 45 are correctly classified.

	Positive lymph node status (N1+)	Negative lymph node status (N0)
Classified lymph node positive (N1+)	36	10
Classified lymph node negative (N0)	9	45

Conclusion

1. The current study demonstrates that the metastasis status of a breast carcinoma may be determined via geometric and color analysis of the primary tumor.
2. Future work involves improving sensitivity and specificity by including additional feature vectors and cases.
3. Additionally work has begun to calculate a quantitative confidence measure of the estimated lymph node metastasis status

References

1. Zhang J. An Approach to Analyzing Histology Segmentations Using Shape Distributions. M.S. Thesis, Drexel University, Philadelphia, PA 2008.
2. Multidimensional Shape and Color Distributions as a Computational Biomarker for Cancer Pathology - Predicting lymph node status. United States and Canadian Academy of Pathology 2010 Annual Meeting.