



A STUDY OF SHAPE DISTRIBUTIONS FOR ESTIMATING HISTOLOGIC GRADE

Jasper Z. Zhang*, Sokol Petushi**,
William C. Regli*, Fernando U. Garcia** and David E. Breen*
* Drexel University ** Drexel University College of Medicine

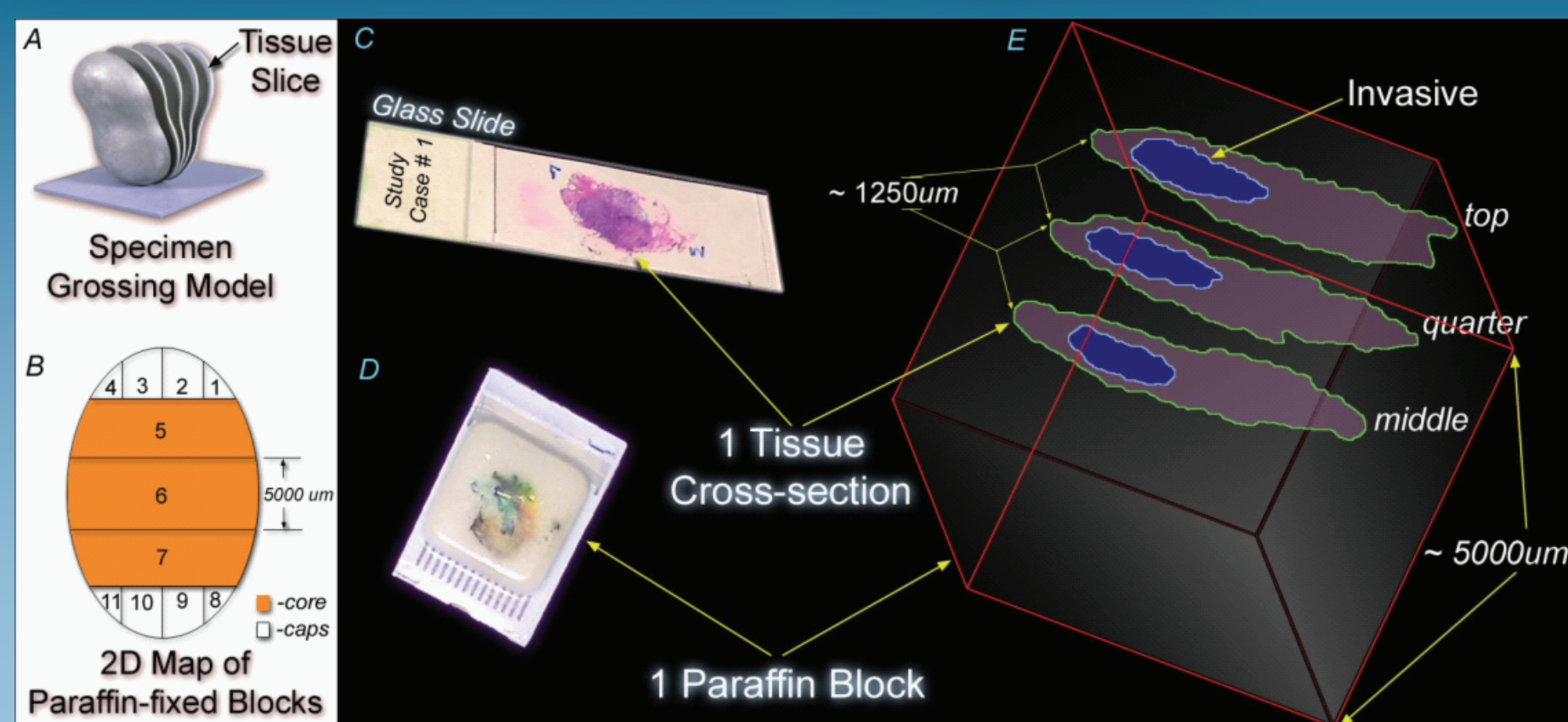
Abstract

Breast cancers can be histologically categorized (graded) based upon their architectural patterns and cellular types. Inaccurate histologic grading can result in inappropriate treatment for a given patient. Computational analysis of breast cancers offers an operator-independent method for histologic grading that should enhance grading reliability. We present the initial efforts to develop computational technologies that may be used to automatically and objectively estimate the histologic grade of breast cancer tumors. The approach utilizes image processing and shape analysis of imaged histologic sections. Our work is based on the hypothesis that cellular structures found in breast cancer tumors can be transformed into distinct high-resolution shape distributions using geometric measures from stochastic geometry. The resulting shape distributions define well-populated regions of the associated high-dimensional space. Mapping an unknown breast cancer sample into this high-D space and determining to which region it belongs will allow for the automatic estimation of its histologic grade.

The U.S. Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick, MD 21702-5014 is the awarding and administering acquisition office. This investigation was partially funded under a U.S. Army Medical Research Acquisition Activity; Cooperative Agreement W81XWH 04-1-0419. The content of the information herein does not necessarily reflect the position or the policy of the U.S. Government or the U.S. Army and no official endorsement should be inferred.

1 Scan

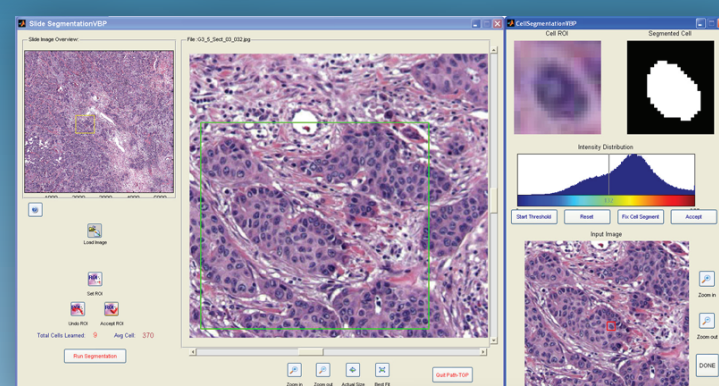
Slices from biopsies stained using the Hematoxylin and Eosin (H&E) process were scanned at 10x magnification.



2 Segmentation

The segmentation technique is a semi-automated image clustering algorithm that combines optimal color-space information with structure-oriented multi-level thresholding and filtering, while incorporating high-level domain knowledge.

We present a graphical user interface to the user that allows him/her to train the segmentation process to identify invasive cancer cells in a scanned H&E stained histology specimen. The training helps to determine the color, size and threshold values needed to segment the cells.



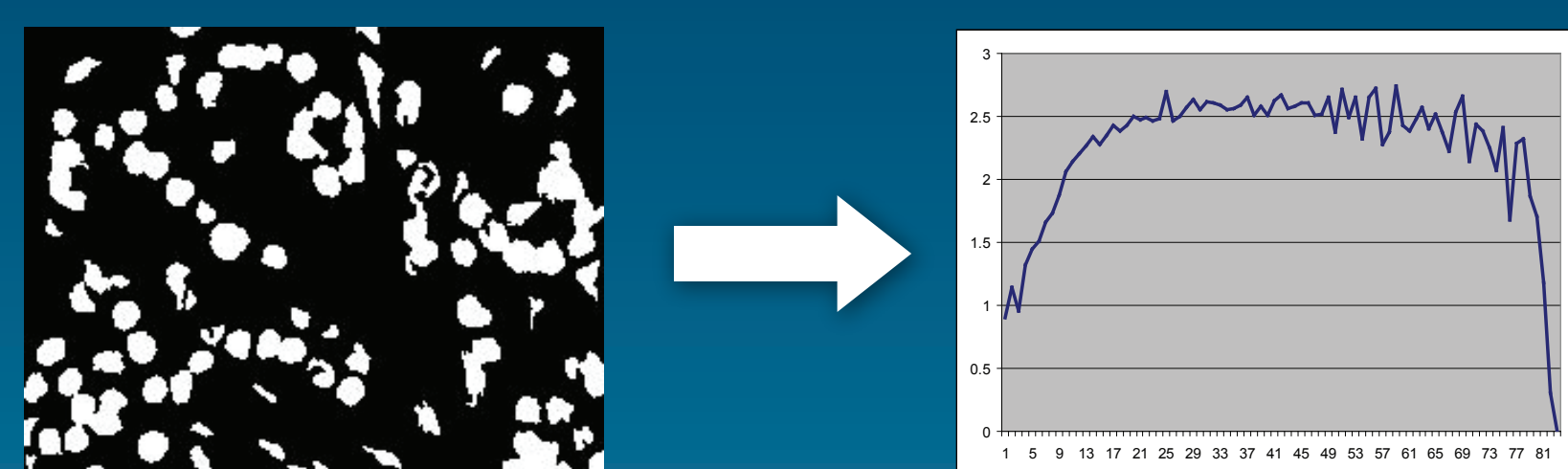
1 Scan

10x

2 Segmentation

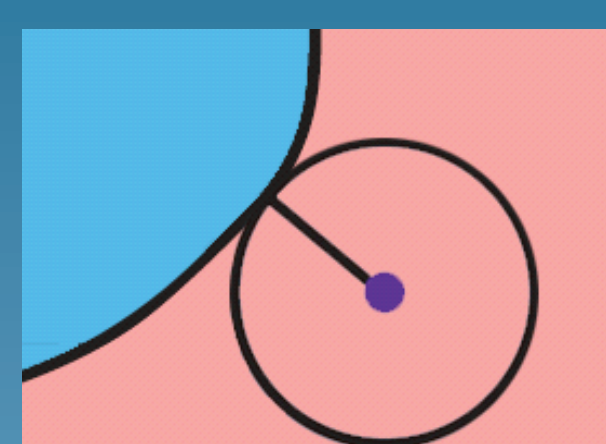
3 Shape Distribution Generation

Each segmented image is transformed into a set of shape distributions by applying a variety of geometric measures.

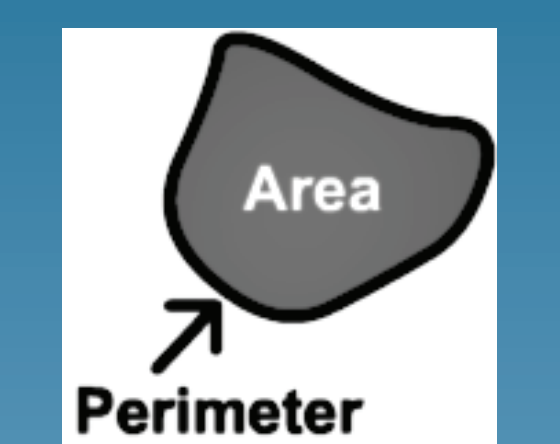


Shape Functions

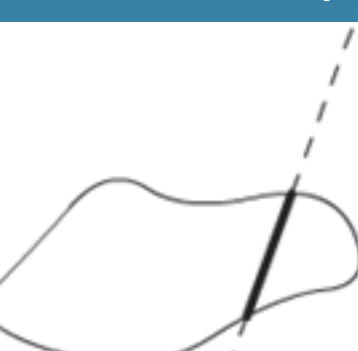
Inside Radial Contact



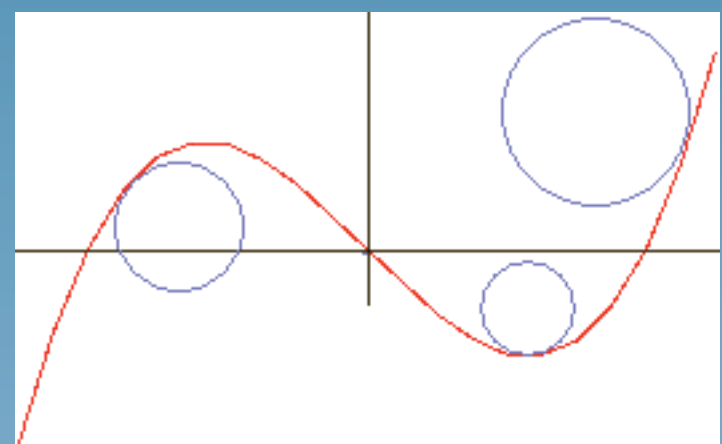
Area, Perimeter, Area vs. Perimeter



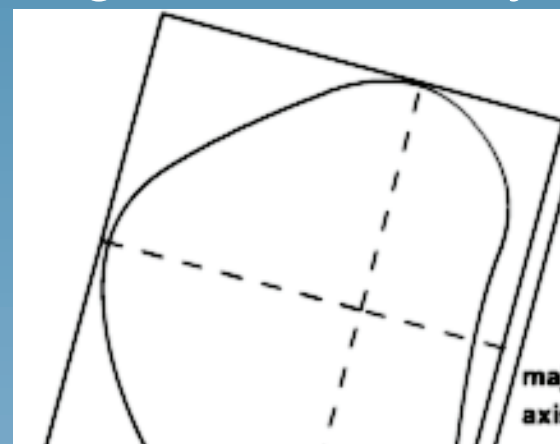
Line Sweep



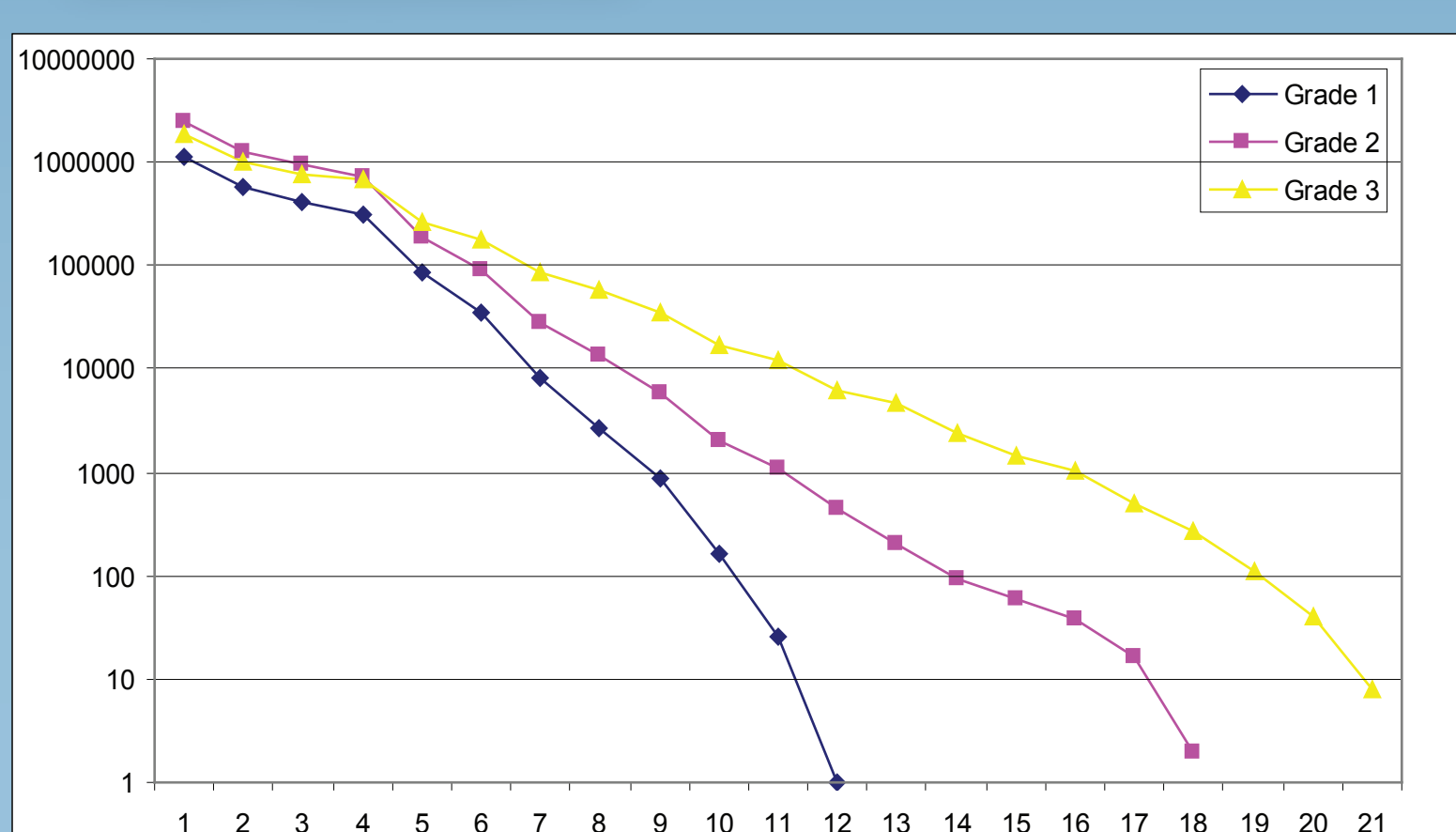
Curvature



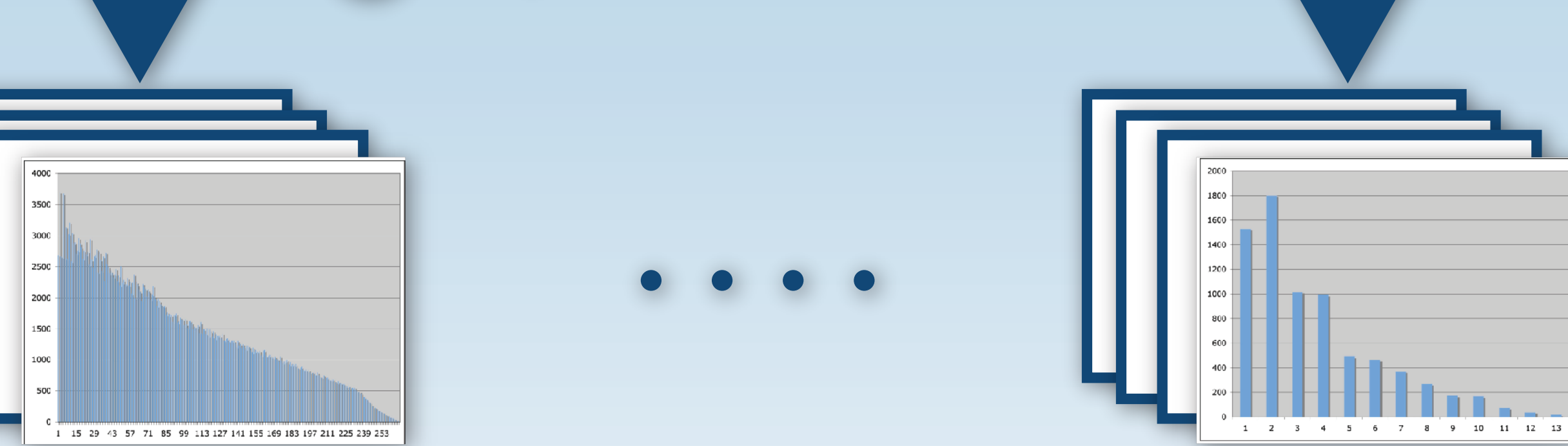
Aspect Ratio, Eigenvector Analysis



Distributions



3 Shape Distribution Generation



4 Filtering

4 Filtering

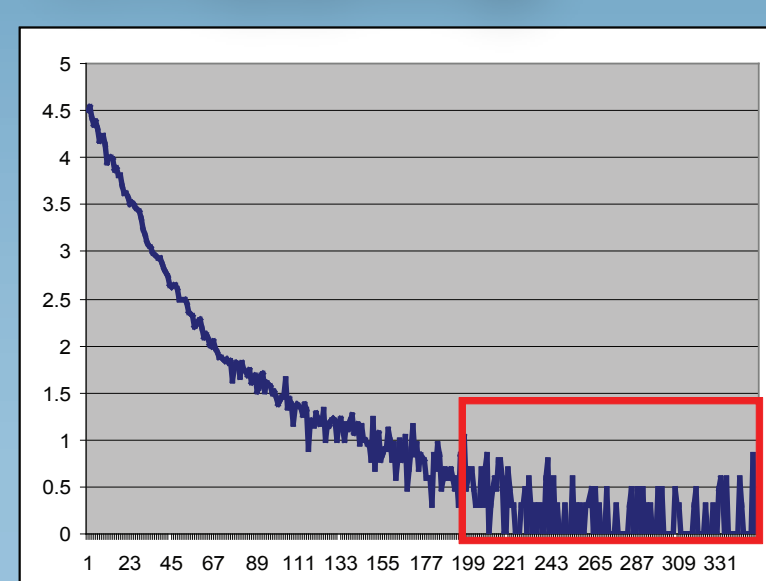
Post-filtering is applied to the shape distributions before the analysis stage to ensure consistency and "cleanness" of the data.

Multipliers

Multipliers are applied to the distributions to produce consistent and comparable histograms for the Comparison stage.

	Multiplier
Inside radial contact	1x
Line sweep	1x
Area	0.10x
Perimeter	1x
Area vs. Perimeter	10x
Curvature	50x
Aspect Ratio	100x
Eigenvector	100x

Cropping

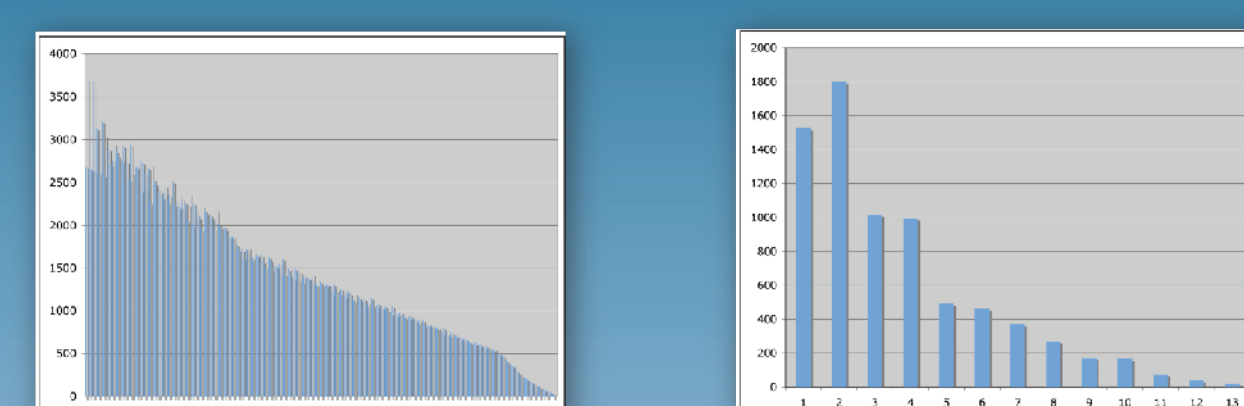


We crop the data to eliminate noise near the beginning and the end of the distributions.

5 Comparison

5 Comparison

All cases are compared by measuring the difference between their shape distributions with the Earth Mover's Distance (EMD). EMD was proposed by Rubner, Tomasi and Guibas in 1998 to measure the distance between distributions.

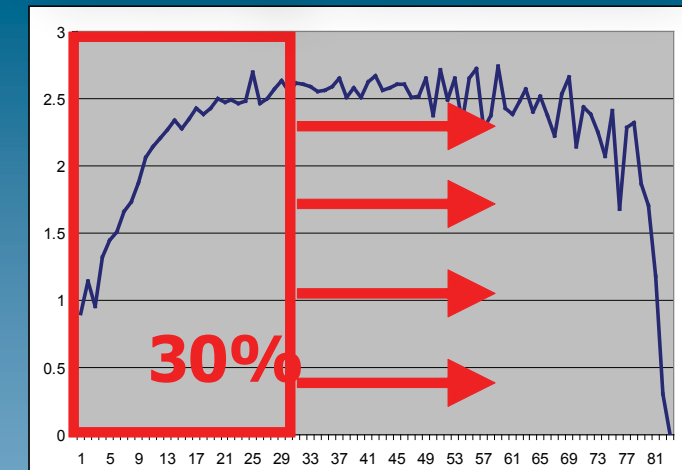


EMD

Difference

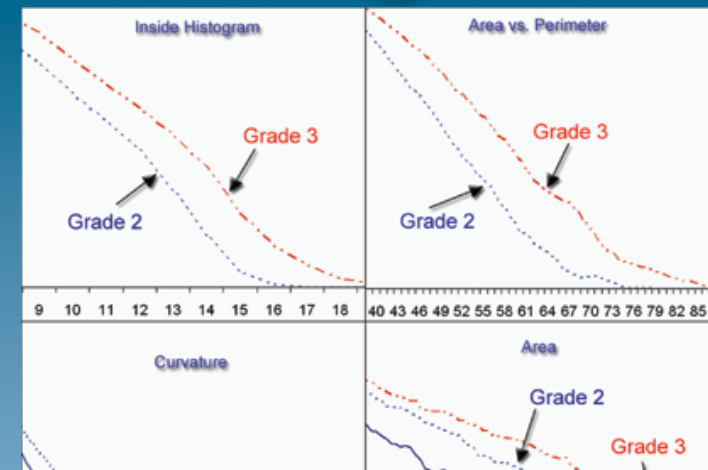
6 Performance Analysis

Sliding Windows



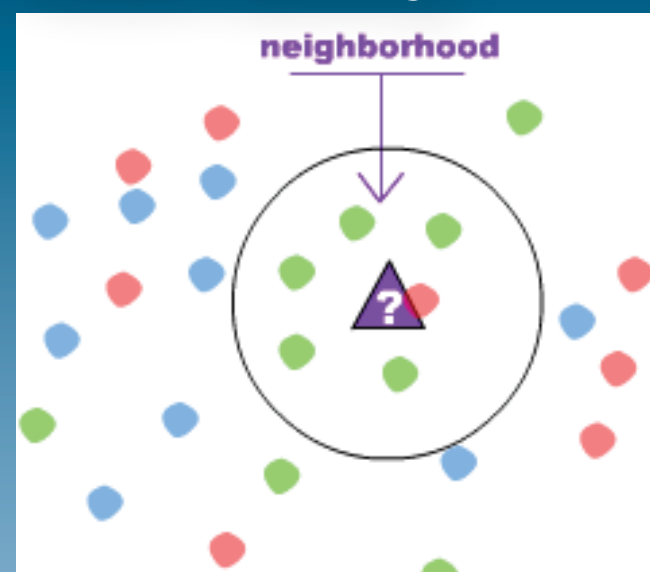
The sub-region of best separation between grades were discovered using a sliding window technique on the filtered shape distributions.

Sub-regions



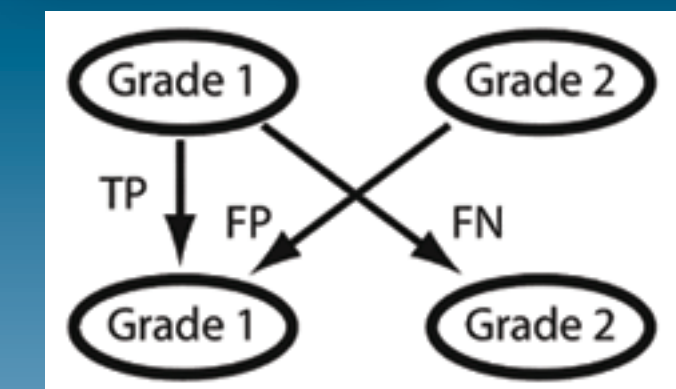
The best sub-regions of each shape distribution were used for the final analysis using EMD.

KNN Query



A KNN query is performed on each case to classify its grade.

Information Retrieval Analysis



$$precision(C_j) = \frac{TP}{TP + FP}$$

$$recall(C_j) = \frac{TP}{TP + FN}$$

$$F = \frac{GeometricMean^2}{ArithmeticMean} = \frac{2 \times (precision \times recall)}{precision + recall}$$

Results

