IMAGE-BASED HISTOLOGIC GRADE ESTIMATION USING STOCHASTIC GEOMETRY ANALYSIS

Sokol Petushi, Jasper Zhang**, Aladin Milutinovic*, David E. Breen** and Fernando U. Garcia**

*** Drexel University College of Medicine   ** Drexel University**

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.

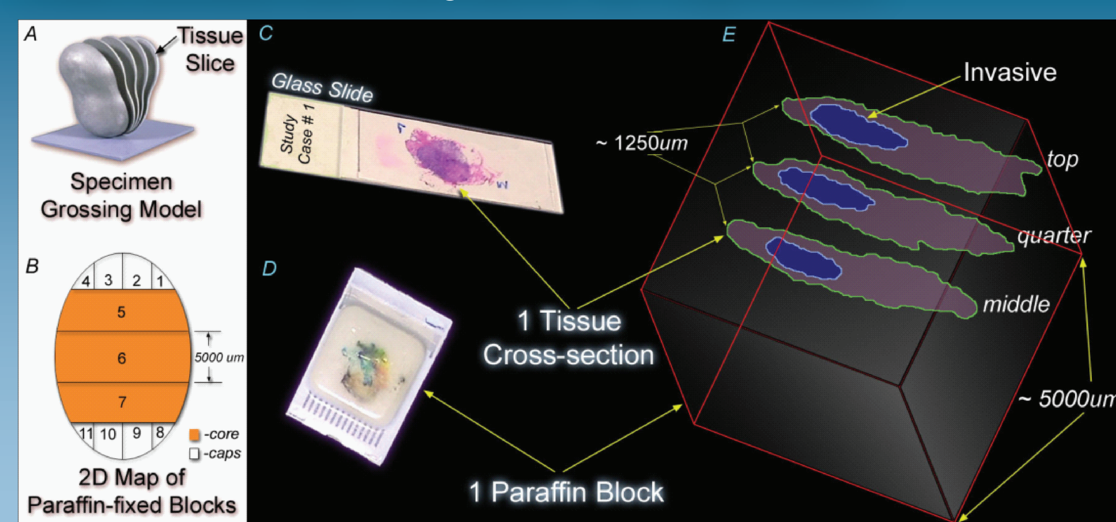
Overview

GOAL: Produce an automated image-based technique that estimates the histologic grade of a breast carcinoma specimen

- Correlate spatial patterns of invasive cancer cells in tumor with histologic grade
- Process H & E stained scans of specimens to identify invasive cancer cells
- Utilize Stochastic Geometry to generate spatial statistics
- Employ Earth Mover's Distance to quantify differences between distributions
- Use k-nearest-neighbors (KNN) algorithm to classify specimens
- Initial results are encouraging with ~75% success rate

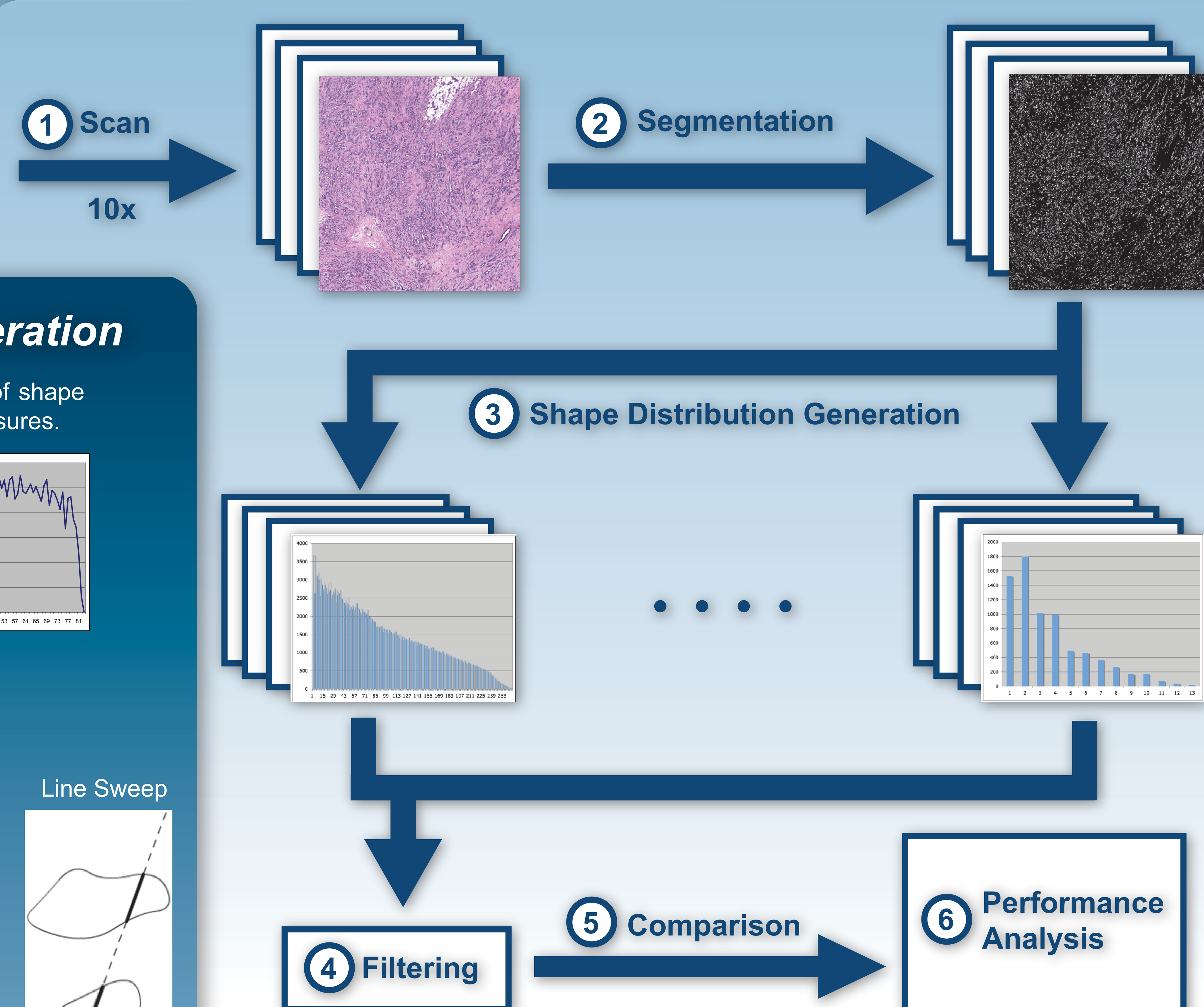
1 Scan

31 specimens stained with Hematoxylin and Eosion (H&E) were scanned at 10x magnification.



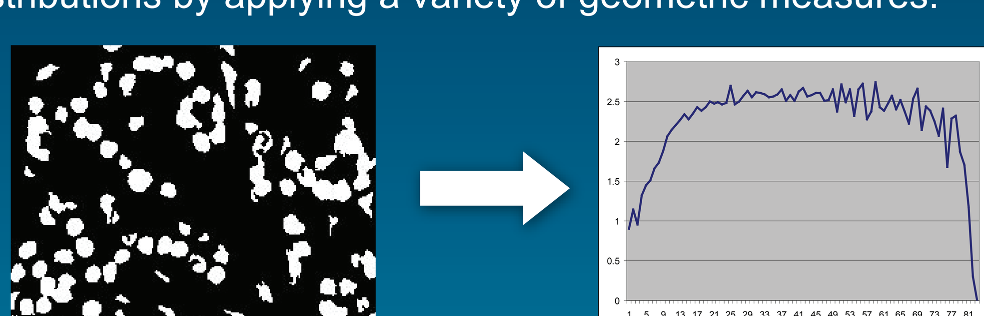
2 Segmentation

- Segment images using Otsu optimal thresholding algorithm and morphological operators
- 6,400^2 pixel region (ROI) with the highest neoplastic cell nuclei density is identified in each image
- Each ROI is resegmented by applying the Otsu algorithm to a combination of the Hue and Saturation channels
- GUI allows user to specify cell nuclei in the ROI
- Information (cell size & shape, variance of intensity) is used to remove outliers in the segmentation image

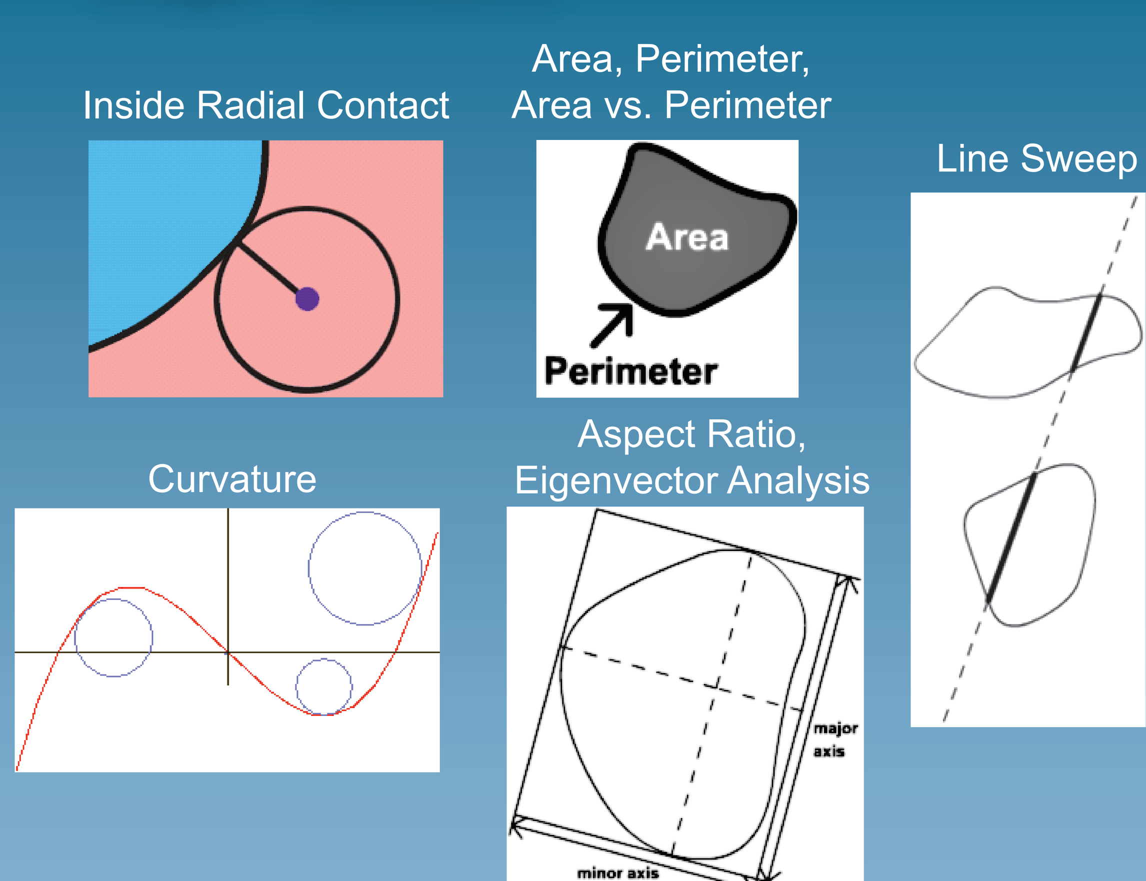


3 Shape Distribution Generation

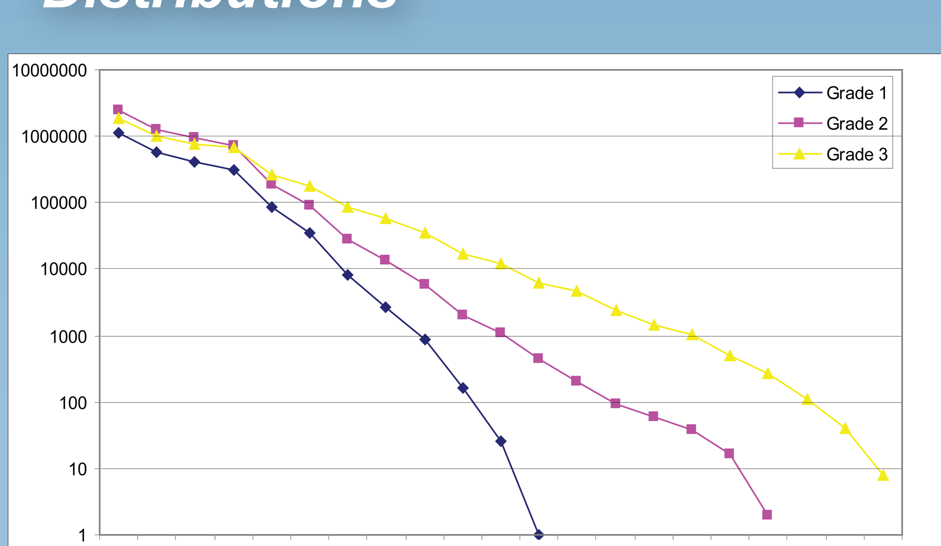
Each segmented image is transformed into a set of shape distributions by applying a variety of geometric measures.



Shape Functions



Distributions



4 Filtering

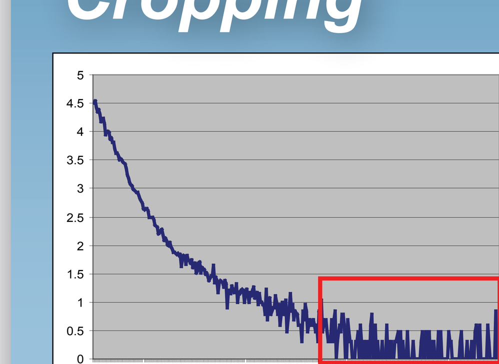
Post-filtering is applied to the shape distributions before the analysis stage to ensure consistency and "cleanness" of the data.

Multipliers

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

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

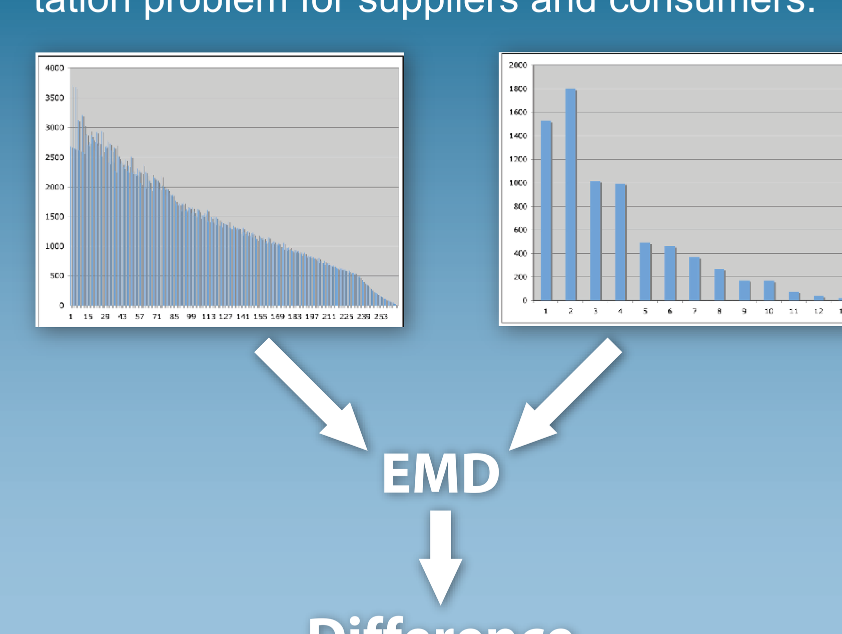
Cropping



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

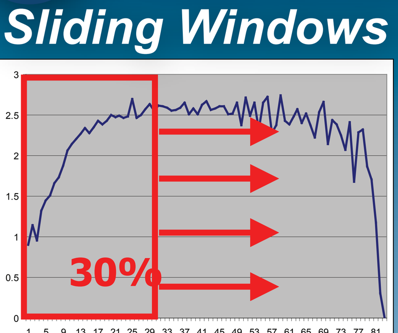
5 Comparison

The shape distributions are compared by computing the Earth Mover's Distance, an optimal similarity metric based on solving the transportation problem for suppliers and consumers.



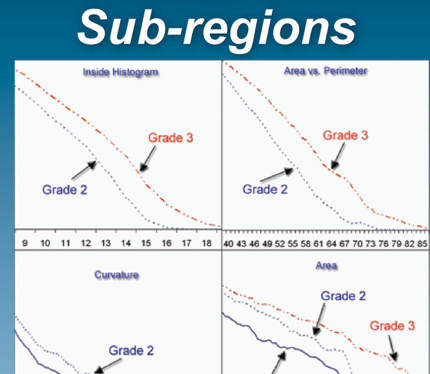
6 Performance Analysis

Sliding Windows



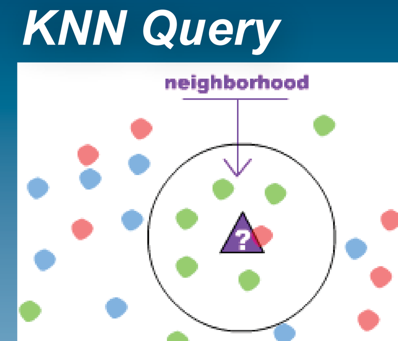
The subregions of best separation between grades were discovered using a sliding window technique on the filtered shape distributions.

Sub-regions



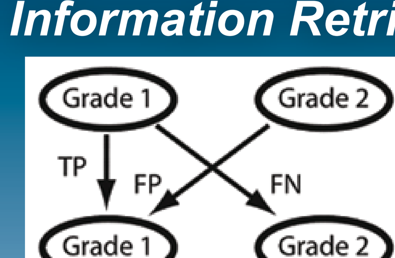
The best sub-regions of each shape distribution were used for the final analysis using EMD.

KNN Query



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

Information Retrieval Analysis



True Positive (TP), False Positive (FP), False Negative (FN)

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

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

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

The performance of each measure to properly classify the cases is evaluated with the F-measure. F-measure is the harmonic mean of the precision and recall of the KNN query results.

Results