

# Feature Extraction and Classification of Plankton

We done things

Dane Skinner   Nick Hockensmith   Kevin Park

Oregon State University

May 15, 2014

# Questions of Interest

- Using Histogram and Moments as a image feature extraction how accurate are the classification of plankton?

# Kratuchok's Moments

- Calculating *Kratuchok's* moments,

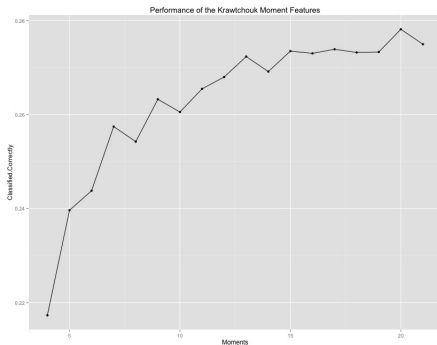
$$Q_{nm} = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \bar{K}_n(x; p_1, N-1) \bar{K}_m(y; p_2, M-1) f(x, y),$$

where  $f(x, y)$  is the pixel intensity and  $(K_n(a; p, N))$  are the weighted Krawtchouk polynomials, and  $n \in \mathbb{N}$  is order of the moment in the x- or y-direction.

- Kratuchok moments are invariant under scaling, rotation, and translation.

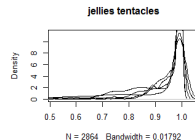
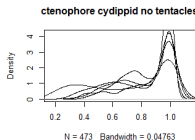
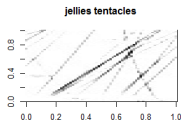
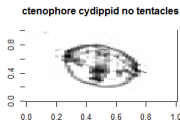
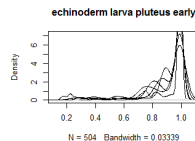
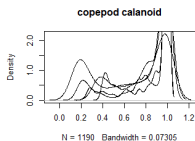
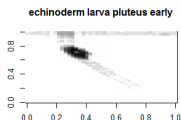
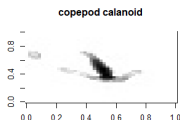
# Kratuchok's Moments

- The training set was further divided into a training and validation set to determine the order of moments that yielded the “best” classification of the validation set.
- As you can see below, it appears the 20<sup>st</sup> moment offers the best prediction of the validation set.



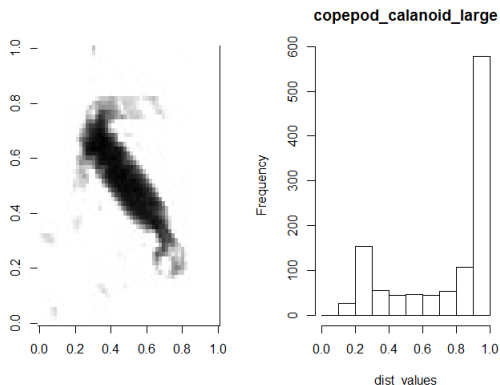
# Histogram Method

- Some of species of plankton give distinct distributions of gray scale values.



# Histogram Method

- The grayscale is on a  $[0,1]$  interval and we partition the interval into a width of 0.1.
- We have count the number of values that are between  $[0, 0.1]$ ,  $[0.1, 0.2]$ ,  $\dots$ ,  $[0.9, 1]$ .



# Indicoio Package and kNN

- This produces a sparse, 2048 digit feature vector for each image that can then be used to calculate the Euclidean distances between different feature vectors.
- The data extracted from Indicoio was too "noisy" for kNN to make any accurate classifications and the large number of classes made it difficult for computation time.
- The kaggle score was 8.1 ( $\sim 1001$  ranking).

# Kaggle Results: Histogram & Moments

- We were fairly surprised as much as you.
- Histogram method produced a score of 3.29.

660	11	Miusay	3.288775	2	Wed, 11 Mar 2015 23:58:04
-		<b>Kevark</b>	<b>3.295601</b>	-	<b>Mon, 11 May 2015 09:38:29</b> <small>Post-Deadline</small>
<b>Post-Deadline Entry</b> If you would have submitted this entry during the competition, you would have been around here on the leaderboard.					
661	1	Attila Egri	3.305361	22	Mon, 16 Mar 2015 23:45:01 (-39.7s)
662	11	Sinbad	3.326039	3	Wed, 21 Jan 2015 14:27:00

- Combination of Histogram and Moments produced a score of 2.66.

597	-	Pieter Gerrit Bosma	2.653109	1	Fri, 19 Dec 2014 09:01:26
598	-	Shical Yang	2.657105	9	Fri, 06 Feb 2015 09:31:45 (-3.9s)
-		<b>Black Heart</b>	<b>2.668862</b>	-	<b>Tue, 12 May 2015 16:31:39</b> <small>Post-Deadline</small>
<b>Post-Deadline Entry</b> If you would have submitted this entry during the competition, you would have been around here on the leaderboard.					
599	12	Dan Tybor	2.668967	3	Wed, 14 Jan 2015 04:13:31




# Kaggle Results:

- Krawtchouk moments (20<sup>th</sup>-order) produced a score of 2.18.

522	<span>↑2</span>	BRES 90	2.173581	6	Mon, 16 Mar 2015 23:35:47 (-0.1h)
-		<b>Black Heart</b>	<b>2.183565</b>	.	<b>Thu, 14 May 2015 01:21:01</b> Post-Deadline
<b>Post-Deadline Entry</b> If you would have submitted this entry during the competition, you would have been around here on the leaderboard.					
523	<span>↓2</span>	39rus	2.185156	6	Tue, 23 Dec 2014 23:13:32

- Histogram and Moments combo produced a score of 2.24.

528	<span>↑3</span>	Eyrun	2.235007	1	Fri, 27 Feb 2015 01:29:09
-		<b>Black Heart</b>	<b>2.239828</b>	.	<b>Thu, 14 May 2015 08:07:42</b> Post-Deadline
<b>Post-Deadline Entry</b> If you would have submitted this entry during the competition, you would have been around here on the leaderboard.					
529	<span>↑1</span>	ML Enthusiasts 	2.245985	14	Thu, 19 Feb 2015 03:57:38 (-19.4d)

# Conclusions and Final Thoughts

- Krawtchouk moments required 400 features to achieve it's best Kaggle rank of 523/1049.
- The Histogram method required only 10 features to achieve it's best Kaggle rank of 661/1049.

As future work,

- Perform variable selection for dimension reduction on the Krawtchouk moments.
- Increase the number of bins measured in the Histogram method.
- Look towards 2-D filter for additional features.