Name:

Matthew Tang

Netid:

mhtang2

CS 441 - HW3: PDFs and Outliers

Complete the sections below. You do not need to fill out the checklist.

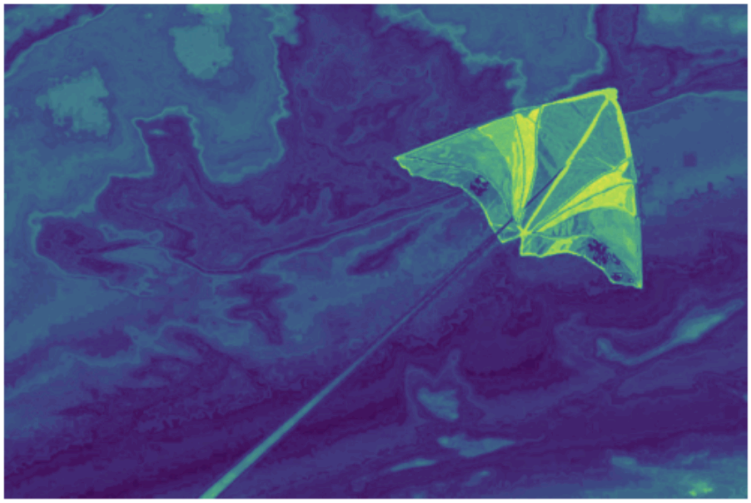
Total Points Available	**[ ] / 160**
1. Estimating PDFs	
a. Segmentation with per-channnel PDFs	[ ] / 15
b. Segmentation with clustered value PDFs	[ ] / 15
c. Segmentation with GMMs	[ ] / 20
2. Robust Estimation	
a. Assume no noise	[ ] / 10
b. Robust estimation with percentiles	[ ] / 15
c. Robust estimation with EM	[ ] / 25
3. Stretch Goals	
a. Impact of school on salary	[ ] / 20
b. Impact of experience on salary	[ ] / 20
c. Mutual information: discrete pdf	[ ] / 10
d. Mutual information: GMM	[ ] / 10

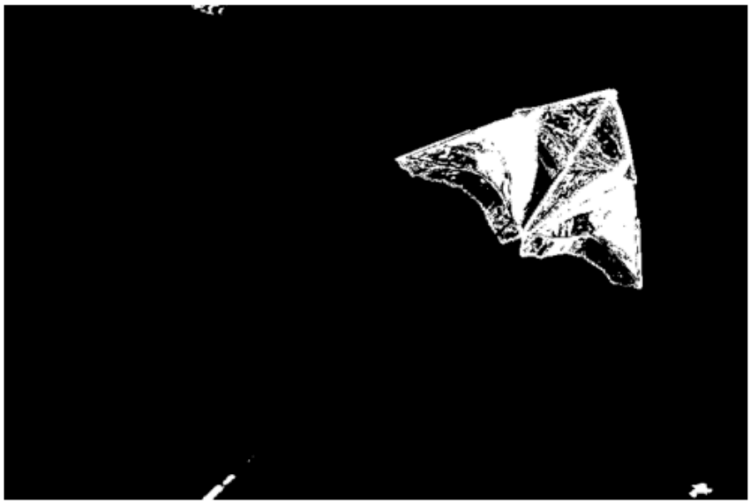
1. Estimating PDFs

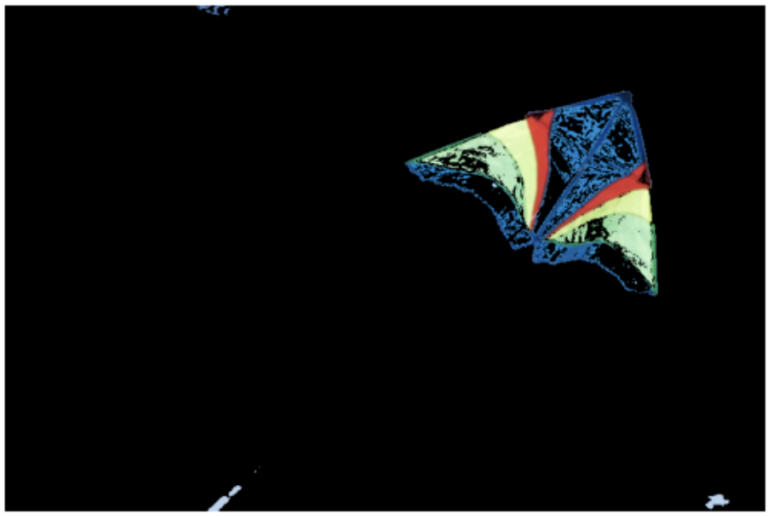
Include the generated images (score map and thresholded RGB) from the display code. List any parameters.

Note: Not sure exactly which images I should include, so I included all the displayed images by the helper function.

a. Method 1 (Per-channel discrete):



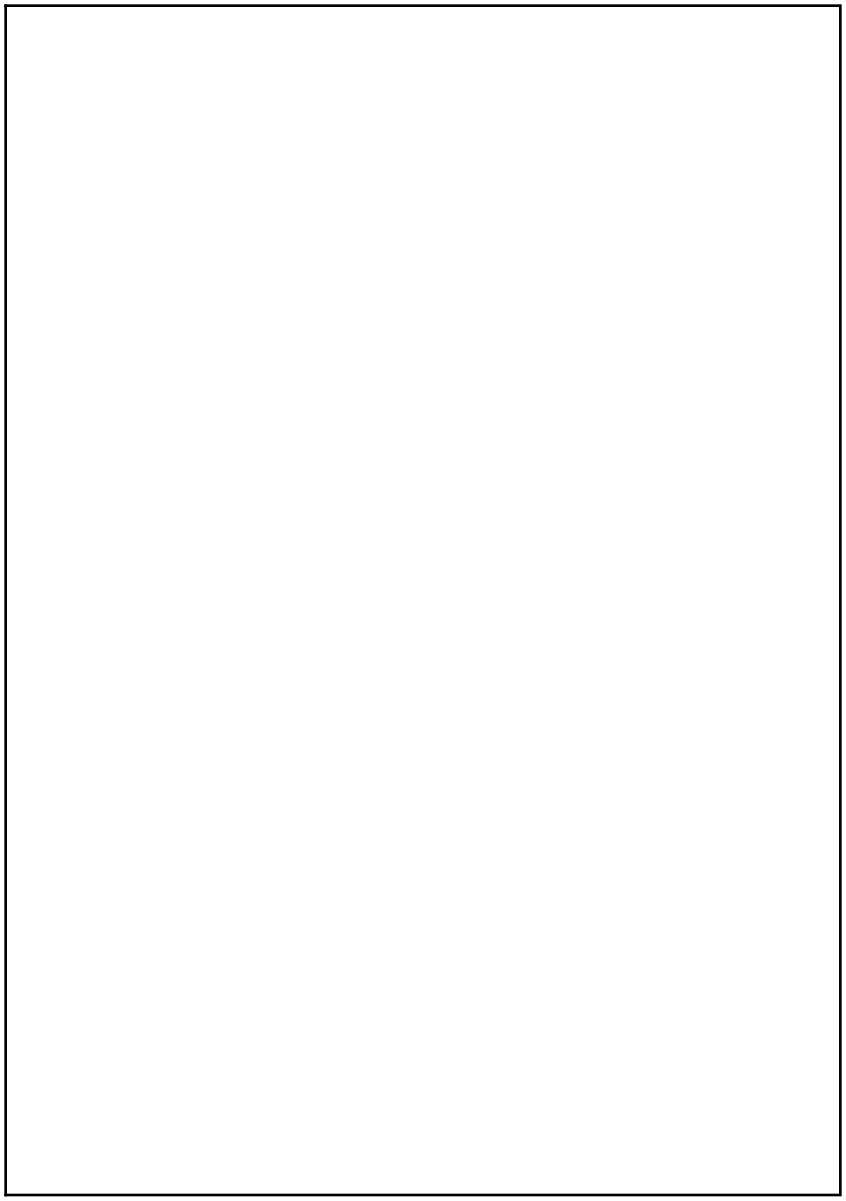


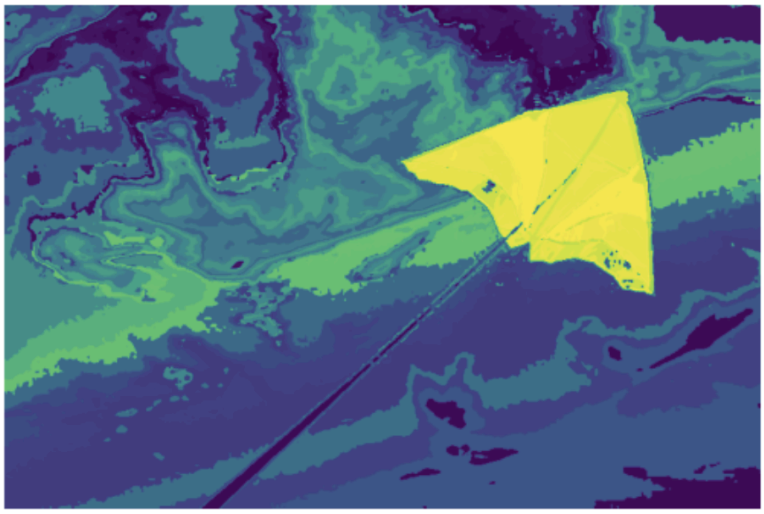


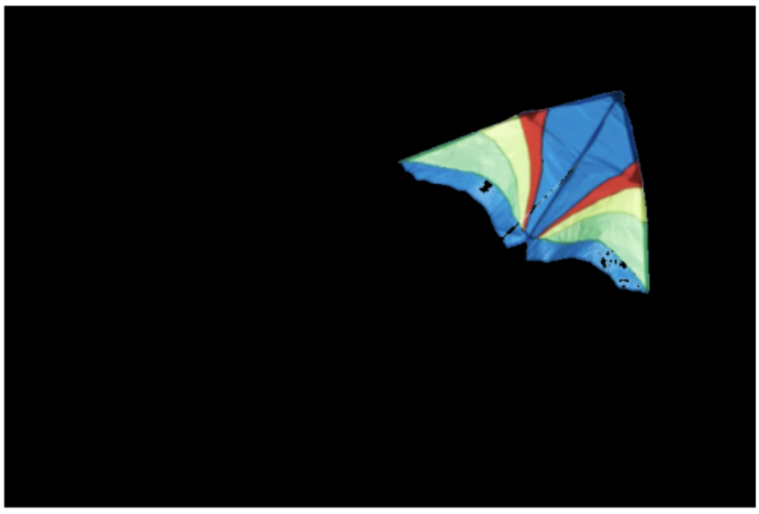
Number of bins / discrete values per channel, threshold

Nbins = 256
Threshold = 2

b. Method 2 (Clustering, discrete):



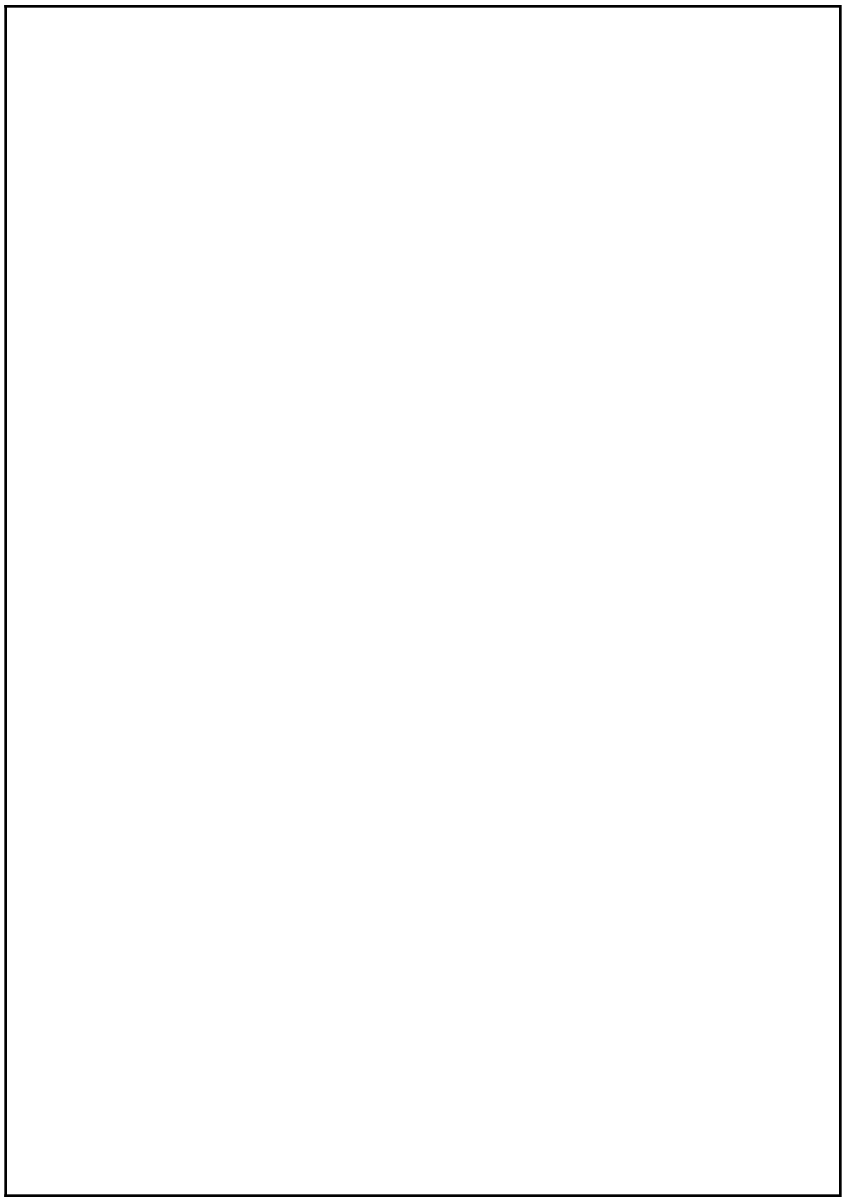


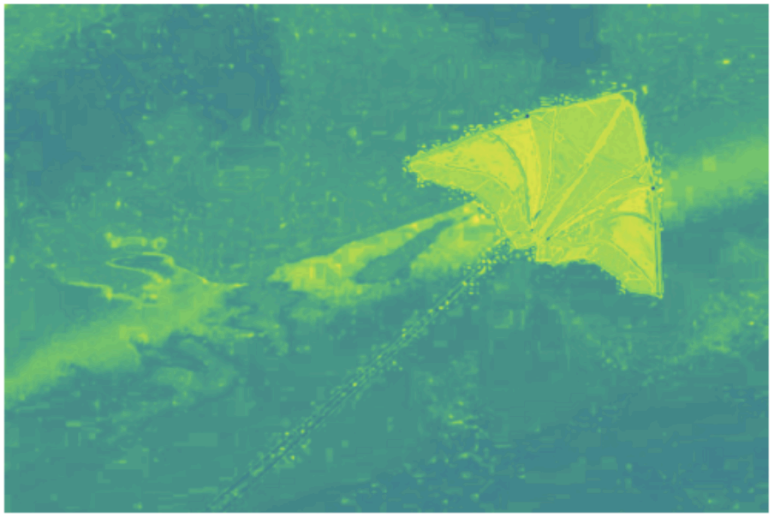


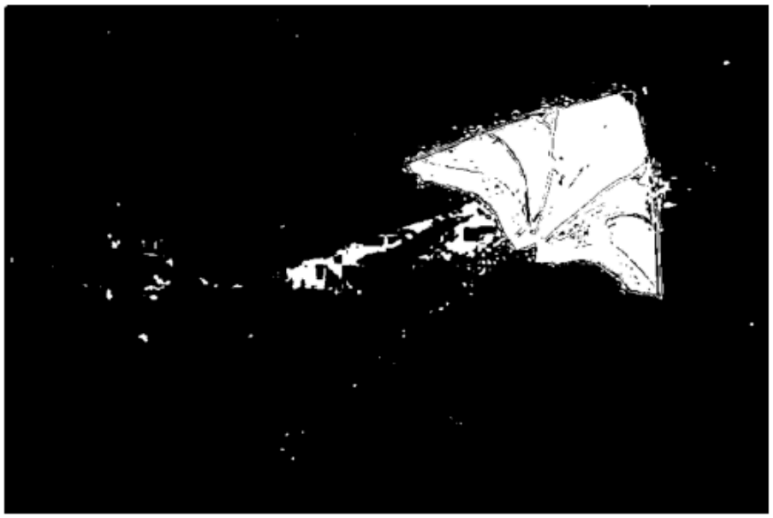
Number of clusters, threshold

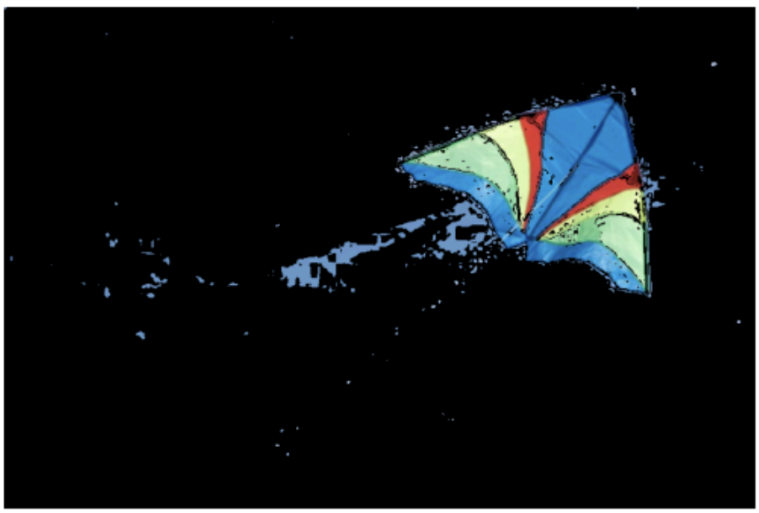
Clusters= 64
Threshold = 1

c. Method 3 (Gaussian Mixture Model):









Number of components, variance model, threshold

Components: 10 Covariance: full Threshold: 1.25

2. Robust Estimation

Round to nearest whole number.

	a. No noise	b. Percentiles	c. EM
Min	64694	75494	64694
Mean	123750	113879	111984
Std	61954	15876	17966
Max	611,494	159,901	169,008

First five indices of invalid data (based on EM solution, you add last 3)

18	28	49	127	128

3. Stretch Goals

a. Impact of school on salary

Report mean salary overall and for each school

	Average Salary
Overall	
School 0 (UIUC)	
School 1 (MIT)	
School 2 (Cornell)	

Describe your approach to estimate this.

b. Impact of years of experience on salary

How much are salaries expected to increase with one year of experience?

$ 1107.17

Describe your approach to estimate this.

I used the EM algorithm from the previous part to filter out samples which were likely to be invalid. To account for school, I added the school as a numerical feature (-1,0,1,2). Then I trained a linear regressor and took the coefficient for the experience feature, since this represents the change in salary per change in years of experience.

c. Mutual information of sex and age, discrete approach

Mutual information (base natural log)

0.092593

d. Mutual information of sex and age, GMM approach

Mutual information (base natural log)

0.02427

Acknowledgments / Attribution

Code for EM is taken from lecture slides and adapted to the problem.