Name:

Hanliang Jiang

Netid:

hj33

CS 441 - HW3: PDFs and Outliers

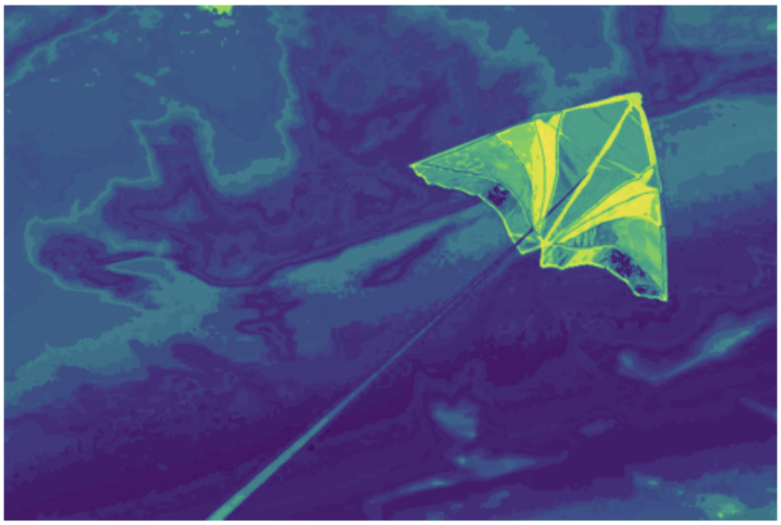
Complete the sections below. You do not need to fill out the checklist.

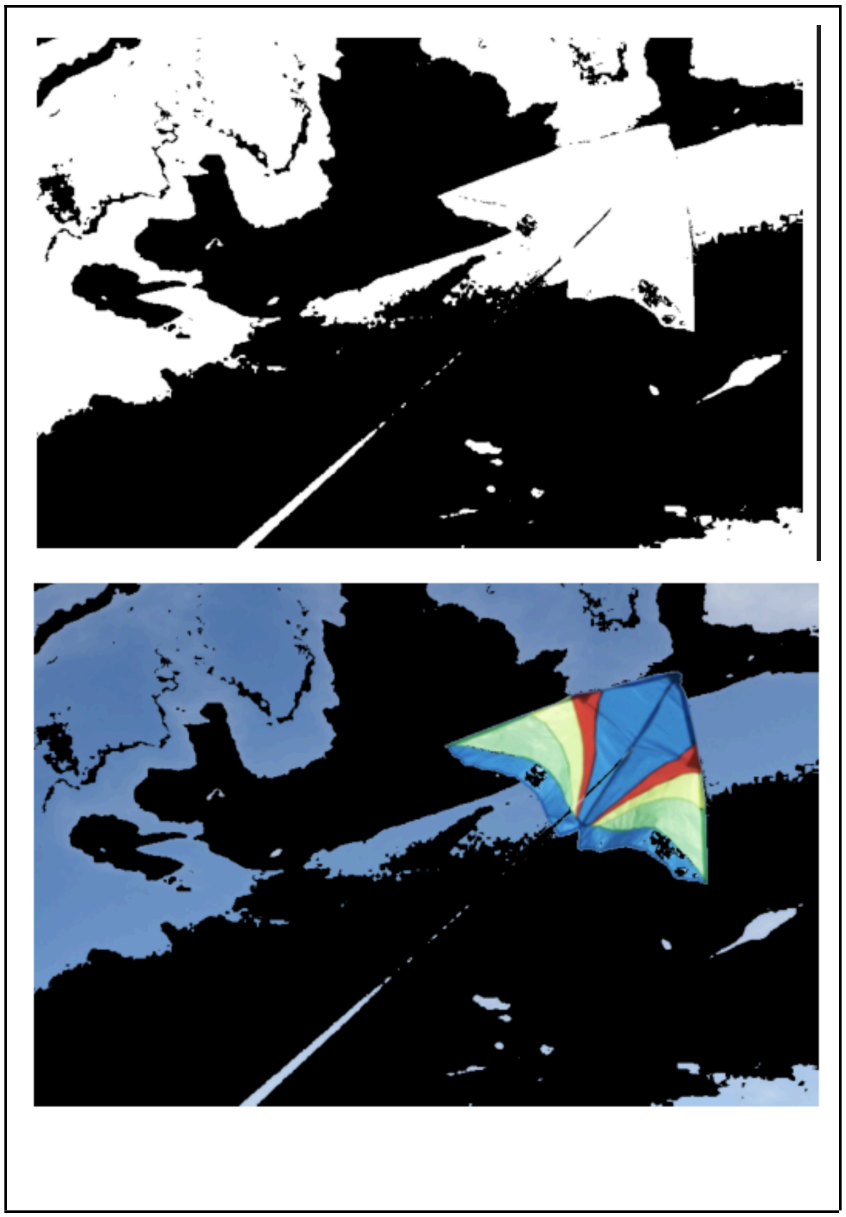
Total Points Available	**[ ] / 160**
1. Estimating PDFs	
a. Segmentation with per-channnel PDFs	[ ] / 15
b. Segmentation with clustered value PDFs	[ ] / 15
c. Segmentation with GMMs	[ ] / 20
2. Robust Estimation	
a. Assume no noise	[ ] / 10
b. Robust estimation with percentiles	[ ] / 15
c. Robust estimation 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.

a. Histogram:

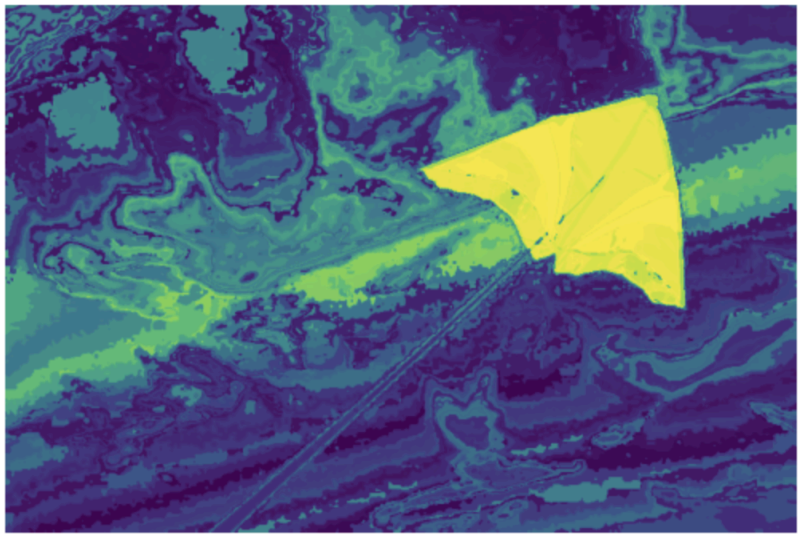


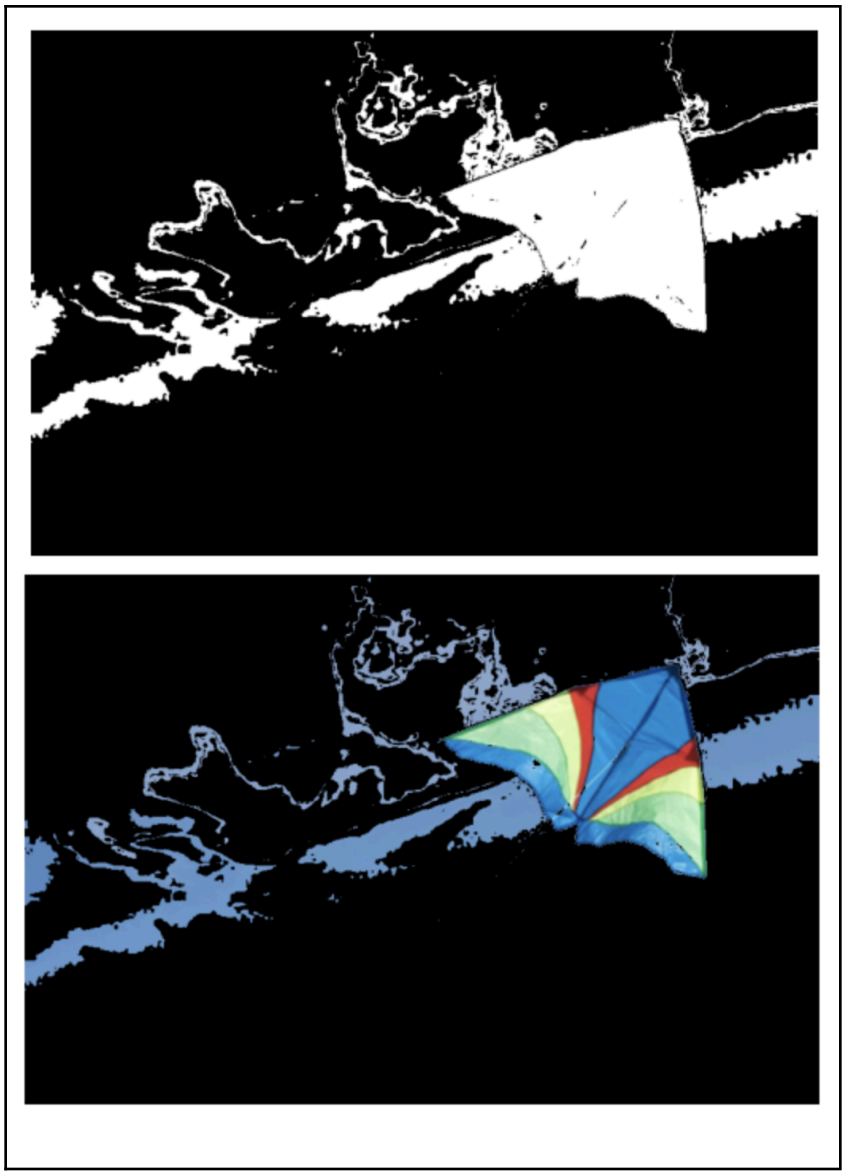


Number of bins / discrete values per channel, threshold

Nbins = 40 Threshold = -2

b. Clustering:

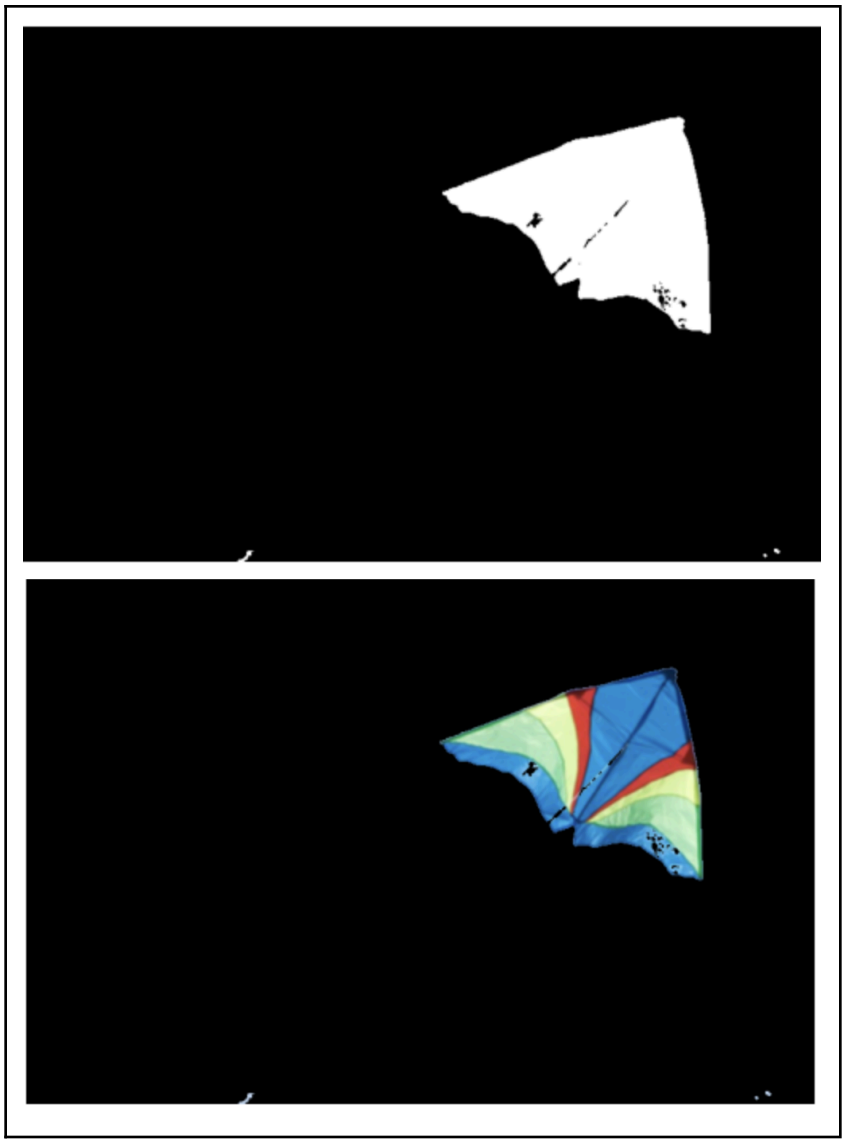




Number of clusters, threshold

Clusters= 200 Threshold = -1

c. Gaussian mixture Model:



Number of components, variance model, threshold

Components: 5 Covariance: ‘diag’ Threshold: -1

2. Robust Estimation

Round to nearest whole number.

	a. No noise	b. Percentiles	c. EM
Min	64,694	75,494	64,694

Mean	123,750	113,879	83,457
Std	61,954	15,876	35,916
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	123,750
School 0 (UIUC)	120,046
School 1 (MIT)	125,266
School 2 (Cornell)	127,559

Describe your approach to estimate this.

I assume all three are gaussian distribution, then for each Mr.school-unknown's salary, I calculate the possibility given school is 0 or 1 or 2, and assign the number(0,1,2) with the highest probability.

b. Impact of years of experience on salary

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

4,153

Describe your approach to estimate this.

I calculate the average gap between two years to get the answer.

c. Mutual information of sex and age, discrete approach

Mutual information (base natural log)

0.09259264627746928

d. Mutual information of sex and age, GMM approach

Mutual information (base natural log)

Acknowledgments / Attribution

None