

# Assignment 2: Blob Analysis

Brian Jacobel

Artificial Intelligence, Fall 2013

## 1 Discussion

### 1.1 Accuracy

A key feature of the data collected in testing this algorithm is the incredible accuracy of the data. Excepting the “angle” geometric measurement (which intentionally did not remain constant over the test sets, and thus cannot be used to examine variance), each of the measurements exhibited a relative standard deviation (coefficient of variation) of around three percent. The absolute worst relative standard deviation came from distinguishing the second principal length with noise and a fuzzy threshold, for which the RSD was 10%. These low error values indicate what we already knew about the blob analysis algorithm: that it is excellent at distinguishing shapes on a flat background. The accuracy was very high in these tests because the tests were designed to play to blob analysis’ strengths - for example, the background was completely flat, no objects overlapped or were out-of-frame, and the objects (for most of their trials) were not aligned to the grid due to the angle variation being tested.

### 1.2 Noise

An interesting result of the test performed with this algorithm was the effect of noise, specifically with respect to the different types of thresholds. Adding one percent noise to the images analyzed with a hard threshold had an almost negligible result. For each of the six geometric measurements, the average change produced in the standard deviation that measurement when noise was introduced was on the order of a tenth of a percent of the mean value. This is not particularly surprising. Noise is by definition random, and blob analysis is designed to find pixels that are connected - that is, non-random. Noise is also uniformly distributed, which means that it should act on all parts of a blob uniformly. If we assume that it is equally likely that noise will cause the algorithm to classify a non-blob pixel as blob and to classify a blob pixel as non-blob, the overall effect should be a wash.

In the two tests with fuzzy thresholds, introducing noise showed to have a similarly negligible impact. This is also not surprising, especially given that fuzzy thresholds are less sensitive to minor changes in local values (such as the changes that noise produces). Rather than a small change triggering a definite classification from blob to not-blob, a fuzzy threshold can adapt and probabilistically classify, which makes the negligible effect of noise on all six geometric measurements even less surprising than in the hard threshold tests.

### 1.3 Thresholds

For most of the measurements, hard and fuzzy thresholds produced similar values with similar accuracy. The exception was the area of the blobs, which the algorithm reported as about six percent smaller at the mean when using fuzzy thresholds. Though extrapolating from boundary data points is probably not productive, it is notable that the difference between maximums was greater than between minimums when switching between hard and fuzzy thresholds. Finally, the standard deviations when using a fuzzy threshold were about half of the standard deviations when using a hard threshold, giving us confidence that the fuzzy values are in fact the correct ones. One possible reason for the differences in the area measurement between the two thresholds is that when the object being analyzed was rotated and many grid squares through which its border crossed became partially blob and partially non-blob, the fuzzy threshold was better at recognizing this reality whereas the hard threshold in this situation "rounded up" and created many new pixels out of these fractional squares that were actually not entirely filled with blob.

## 2 Data

Name	Count	Mean	StDev	Min	Max
Area	1800	226.14	3.1702	212.00	242.00
xCenter	1800	0.00	0.0730	-0.30	0.37
yCenter	1800	0.00	0.0730	-0.30	0.37
Angle	1800	-0.00	6.8837	-10.68	10.68
Length1	1800	7.47	0.2733	7.02	7.96
Length2	1800	4.52	0.4428	3.78	5.18

Figure 1: 0.50 edge, no noise, 48x40 image, 64 black, 192 white, hard threshold  $\geq 128.0$

Name	Count	Mean	StDev	Min	Max
Area	1800	226.25	3.0282	212.00	240.00
xCenter	1800	-0.00	0.0701	-0.32	0.32
yCenter	1800	-0.00	0.0706	-0.37	0.35
Angle	1800	-0.01	6.8675	-10.68	10.68
Length1	1800	7.47	0.2721	7.01	7.96
Length2	1800	4.52	0.4429	3.78	5.18

Figure 2: 0.50 edge, 1.00% noise, 48x40 image, 64 black, 192 white, hard threshold  $\geq 128.0$

Name	Count	Mean	StDev	Min	Max
Area	1800	212.14	1.3606	204.92	218.42
xCenter	1800	0.00	0.0434	-0.20	0.27
yCenter	1800	0.00	0.0434	-0.20	0.27
Angle	1800	-0.00	6.8650	-10.26	10.26
Length1	1800	7.48	0.2737	7.06	7.93
Length2	1800	4.49	0.4497	3.78	5.13

Figure 3: 0.50 edge, no noise, 48x40 image, 64 black, 192 white, fuzzy threshold 96.0, 160.0

Name	Count	Mean	StDev	Min	Max
Area	1800	212.20	1.2047	207.11	216.89
xCenter	1800	-0.00	0.0423	-0.21	0.22
yCenter	1800	0.00	0.0417	-0.21	0.21
Angle	1800	-0.00	6.8563	-10.20	10.19
Length1	1800	7.48	0.2732	7.04	7.92
Length2	1800	4.49	0.4501	3.80	5.14

Figure 4: 0.50 edge, 1.00% noise, 48x40 image, 64 black, 192 white, fuzzy threshold 96.0, 160.0