

CSI2120 Project Part 2 Report

Zhiyu Lin 300255509

Yitao Cui 300345709

Functions

1.

```
(define (read-helper p)
; helper function used to read all numbers from the
; query histogram and store them in a list
; returns a list that stores the histogram data
```

2.

```
(define (readQuery queryFilename)
; read query histogram file and store it in a list,
; ex: (0.0029 0.0157 ...), already normalized.
; returns a list that stores the normalized histogram
data.
```

3.

```
(define (readDataset-helper p)
; helper function used to read all numbers from the
; dataset histogram, and normalize each histogram data.
; returns a list that stores the histogram data
```

4.

```
(define (readDatasetHistogram datasetFilename)
; read dataset histogram file and store it in a list.
; returns a list that stores the normalized histogram
data.
```

5.

```
(define (get-txt-files directory)
; read dataset files names
; returns a list that stores dataset files name
; ex: (1001.jpg.txt 1002.jpg.txt 1003.jpg.txt ... )
```

6.

```
(define (process-files file-list directory)
```

```
; processes each file in a list by constructing sublists
that contain each dataset image's filename and its
histogram.
; returns a list that stores dataset files names and
histograms. ex: ((1001.jpg.txt (histogram)) ... )
```

7.

```
(define (compare-histograms query-hist dataset-hist)
; helper function used to calculate the similarity score
between a query histogram and a dataset histogram.
; query-hist (list), dataset-hist (list) -> Output:
similarity score (number)
```

8.

```
(define (process-and-compare-files file-list directory
query-histogram)
; helper function used to process and compare each
histogram in the dataset with the query histogram.
; returns a list of similarity scores paired with
filenames.
; file-list (list), directory (string), query-histogram
(list) -> Output: (list)
```

9.

```
(define (compare-query-with-dataset directory
queryFilename)
; compares the query histogram to all histograms in the
dataset.
; returns a list of similarity scores and corresponding
filenames.
; directory (string), queryFilename (string) -> Output:
(list)
```

10.

```
(define (insert-sorted result item)
; inserts an item into a sorted list based on the item's
first element. Sort from biggest to smallest.
; returns the sorted list
```

11.

```
(define (sort-results results)
; sort the similarity results list (from biggest to
smallest)
; returns the sorted list
```

12.

```
(define (insert-sorted result item)
  ; inserts an item into a sorted list based on the item's
  first element. Sort from biggest to smallest.
  ; returns the sorted list
```

13.

```
(define (take-first-5 lst)
  ; helper function for display the final results.
  ; returns the first 5 elements from a list.
```

15.

```
(define (display-results results count)
  ; helper function for display the final results.
  ; Output: displays the Top 1 to Top 5 file names and
  similarity.
```

16.

```
(define (similaritySearch queryHistogramFilename
  imageDatasetDirectory)
  ; main function, get the query histogram file name and
  the dataset directory.
  ; Output: displays the final results.
```

Results

q00

```
Welcome to DrRacket, version 8.11.1 [cs].
Language: racket, with debugging; memory limit: 256 MB.
> (similaritySearch "q00.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 1144.jpg.txt | Similarity: 1.0
Top 2 = 3806.jpg.txt | Similarity: 0.7083179468242244
Top 3 = 3714.jpg.txt | Similarity: 0.6694112352289511
Top 4 = 3756.jpg.txt | Similarity: 0.6630239106351552
Top 5 = 2462.jpg.txt | Similarity: 0.6501396325701626
>
```

q01

```
> (similaritySearch "q01.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 3588.jpg.txt | Similarity: 0.5716338164697193
Top 2 = 2536.jpg.txt | Similarity: 0.5625519294682424
Top 3 = 3553.jpg.txt | Similarity: 0.5379315915805022
Top 4 = 1875.jpg.txt | Similarity: 0.529692115952733
Top 5 = 2592.jpg.txt | Similarity: 0.5064334841211227
```

q02

```
> (similaritySearch "q02.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 1091.jpg.txt | Similarity: 0.6553729689807973
Top 2 = 3005.jpg.txt | Similarity: 0.6436484490398816
Top 3 = 850.jpg.txt | Similarity: 0.6304295144017725
Top 4 = 3728.jpg.txt | Similarity: 0.6253519663958637
Top 5 = 1090.jpg.txt | Similarity: 0.6246941931314622
```

q03

```
> (similaritySearch "q03.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 1255.jpg.txt | Similarity: 0.5500369276218612
Top 2 = 3532.jpg.txt | Similarity: 0.5434476550960117
Top 3 = 4507.jpg.txt | Similarity: 0.5397779726735596
Top 4 = 692.jpg.txt | Similarity: 0.5384162666174296
Top 5 = 4737.jpg.txt | Similarity: 0.532842503692762
```

q04

```
> (similaritySearch "q04.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 1799.jpg.txt | Similarity: 0.8469580871491874
Top 2 = 848.jpg.txt | Similarity: 0.4948936022895128
Top 3 = 4261.jpg.txt | Similarity: 0.49057768648449057
Top 4 = 1800.jpg.txt | Similarity: 0.4729966765140327
Top 5 = 1937.jpg.txt | Similarity: 0.4696558807237816
```

q05

```
> (similaritySearch "q05.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 2223.jpg.txt | Similarity: 0.615496907311669
Top 2 = 3091.jpg.txt | Similarity: 0.6013086225997045
Top 3 = 2550.jpg.txt | Similarity: 0.5317404449778435
Top 4 = 2883.jpg.txt | Similarity: 0.5302171805760709
Top 5 = 2901.jpg.txt | Similarity: 0.5243664604874442
```

q06

```
> (similaritySearch "q06.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 2278.jpg.txt | Similarity: 0.776605197562777
Top 2 = 2224.jpg.txt | Similarity: 0.7175094627031022
Top 3 = 1108.jpg.txt | Similarity: 0.7135282034711964
Top 4 = 1917.jpg.txt | Similarity: 0.6985898264401772
Top 5 = 3806.jpg.txt | Similarity: 0.6956471565731168
```

q07

```
> (similaritySearch "q07.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 2377.jpg.txt | Similarity: 1.0
Top 2 = 2379.jpg.txt | Similarity: 0.7677252584933543
Top 3 = 2079.jpg.txt | Similarity: 0.6407288589364853
Top 4 = 1653.jpg.txt | Similarity: 0.6352185653618916
Top 5 = 1896.jpg.txt | Similarity: 0.6232920974889229
```

q08

```
> (similaritySearch "q08.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 2587.jpg.txt | Similarity: 0.7240583456425408
Top 2 = 1837.jpg.txt | Similarity: 0.7017748338257017
Top 3 = 2586.jpg.txt | Similarity: 0.6723539050960118
Top 4 = 1515.jpg.txt | Similarity: 0.6700978581979322
Top 5 = 1217.jpg.txt | Similarity: 0.6689554098966027
```

q09

```
> (similaritySearch "q09.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 2839.jpg.txt | Similarity: 0.7908338718611522
Top 2 = 3414.jpg.txt | Similarity: 0.5104724427621862
Top 3 = 2617.jpg.txt | Similarity: 0.45269225443131506
Top 4 = 4363.jpg.txt | Similarity: 0.4259543482274744
Top 5 = 4816.jpg.txt | Similarity: 0.3865283419497788
```

q10

```
> (similaritySearch "q10.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 2976.jpg.txt | Similarity: 0.7456033050221571
Top 2 = 2971.jpg.txt | Similarity: 0.661252538774003
Top 3 = 3108.jpg.txt | Similarity: 0.645362121491876
Top 4 = 2194.jpg.txt | Similarity: 0.642315592688331
Top 5 = 4118.jpg.txt | Similarity: 0.6353455040620388
```

q11

```
> (similaritySearch "q11.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 3017.jpg.txt | Similarity: 0.612184961225997
Top 2 = 3660.jpg.txt | Similarity: 0.5770229412850811
Top 3 = 2411.jpg.txt | Similarity: 0.5712299206056127
Top 4 = 1010.jpg.txt | Similarity: 0.5571051052437221
Top 5 = 2067.jpg.txt | Similarity: 0.5549125276957163
```

q12

```
> (similaritySearch "q12.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 3135.jpg.txt | Similarity: 1.0
Top 2 = 3133.jpg.txt | Similarity: 0.6428233474889217
Top 3 = 2447.jpg.txt | Similarity: 0.5746457256277693
Top 4 = 4429.jpg.txt | Similarity: 0.5446362629246676
Top 5 = 2964.jpg.txt | Similarity: 0.5137901587887739
```

q13

```
> (similaritySearch "q13.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 3225.jpg.txt | Similarity: 1.0
Top 2 = 2083.jpg.txt | Similarity: 0.49636493722304287
Top 3 = 2082.jpg.txt | Similarity: 0.4960302806499264
Top 4 = 2078.jpg.txt | Similarity: 0.4904911373707539
Top 5 = 2607.jpg.txt | Similarity: 0.4739371768833089
```

q14

```
> (similaritySearch "q14.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 2193.jpg.txt | Similarity: 0.44794936299852284
Top 2 = 3108.jpg.txt | Similarity: 0.38240283419497784
Top 3 = 3318.jpg.txt | Similarity: 0.3815604228212703
Top 4 = 4028.jpg.txt | Similarity: 0.3730612998522895
Top 5 = 3085.jpg.txt | Similarity: 0.37240352658788767
```

q15

```
> (similaritySearch "q15.ppm_normalized.hist.txt" "imageDataset2_15_20")
5 most similar images to the query image:
Top 1 = 4756.jpg.txt | Similarity: 0.7382524002954206
Top 2 = 2462.jpg.txt | Similarity: 0.7244795513293941
Top 3 = 3756.jpg.txt | Similarity: 0.6980186022895124
Top 4 = 908.jpg.txt | Similarity: 0.6945854874446085
Top 5 = 3675.jpg.txt | Similarity: 0.6932814807976365
```