CSCE 350 Programming Assignment

For this programming assignment, we were given sets of data based on action groups for facial expressions. The different action groups were as follows: AU01 (representing surprise), AU12, (representing happiness), and AU17 (representing sadness). Each action group was recognized from images, producing lists of decimal values. We were given 47 different subjects, with each subject containing a query for each action group. These queries were then compared for similarity against each entry in the template file. The similarity of the lists of data was calculated using the cosine similarity, where similarity = sum(x\*y)/(sqrt(sum(x\*x)) \* sqrt(sum(y\*y))). This produces a range of values, from [0, 1] (assuming positive input data), where 1 is the most similar and 0 the least.

In order to solve this problem, we began by reading the files into vectors. Each of the three query file was stored into single vector of doubles, while the template was stored in a vector containing a vector of doubles. Then the outer vector of the template was looped through and the cosine similarity was computed for each query and entry in the template. To improve the efficiency of the algorithm, calculations were performed on each query together, to prevent reading and calculating the magnitude of the same template data on three separate times. After these calculations were completed for one index of the template, the 1-cosine similarity was stored in a multimap using the similarity as a key and the index as the value. A multimap was necessary as some indices have the same similarity value. This improves efficiency by sorting the data into a red-black binary tree on insertion. The function then simply outputs the first ten elements of the map.

This methodology can be seen in the pseudocode for the algorithm, which is as follows:

Algorithm findCosineSim(vector<vector<double> > template, vector<double> query01, vector<double> query12, vector<double> query17)

// Input: a template vector and three query vectors, 0-indexed

// Output: the ten vectors with the greatest cosine similarities for each query vector

multimap<double, int> result01

multimap<double, int> result12

multimap<double, int> result17

for i = 0 to 137

dotProduct01 = 0, dotProduct12 = 0, dotProduct17 = 0

xSquared01 = 0, xSquared12 = 0, xSquared17 = 0

ySquared = 0

for j = 0 to 5631

dotProduct01 += template[i][j] \* query01[j]

dotProduct12 += template[i][j] \* query12[j]

dotProduct17 += template[i][j] \* query17[j]

xSquared01 += query01[j] \* query01[j]

xSquared12 += query12[j] \* query12[j]

xSquared17 += query17[j] \* query17[j]

ySquared += template[i][j] \* template[i][j]

vectorYlength = sqrt(ySquared)

// We use (1 – similarity value) because the values are stored in the // map from smallest to largest, but we need to output the ten largest // values, so we must invert the order to find the nearest pairs  
cosSim01 = 1 - (dotProduct01 / (sqrt(xSquared01) \* vectorYlength))

cosSim12 = 1 - (dotProduct12 / (sqrt(xSquared12) \* vectorYlength))

cosSim17 = 1 - (dotProduct17 / (sqrt(xSquared17) \* vectorYlength))

result01.insert(pair(cosSim01, i+1))

result12.insert(pair(cosSim12, i+1))

result17.insert(pair(cosSim17, i+1))

trim result01 to its first 10 entries

trim result12 to its first 10 entries

trim result17 to its first 10 entries

output result01, result12, and result17

As can be seen in the appendix, the program took 7.56 seconds to run to completion on the ash machine in the linux lab. When optimizing the program, it was easy to reduce the run time from the original 40 seconds to around 10 seconds by changing the way the files were read in. Originally, the template was read in while calculations were performed, but this was changed to an overhead cost before calculations began. However, it was much harder to further optimize the program, which was done by performing calculations for the 3 query files together. Further optimization could be done in the future by taking advantage of multithreading, which would be especially helpful when reading in the files.

Appendix:

../../data/001\_AU01\_query.dat: 8 3 38 33 19 119 116 126 41 91

../../data/001\_AU12\_query.dat: 41 3 38 91 75 135 119 87 33 106

../../data/001\_AU17\_query.dat: 6 131 115 71 118 36 2 55 56 86

../../data/002\_AU01\_query.dat: 4 5 9 7 72 114 128 49 60 41

../../data/002\_AU12\_query.dat: 3 135 128 87 70 75 8 41 38 13

../../data/002\_AU17\_query.dat: 6 9 30 7 103 104 134 128 55 56

../../data/003\_AU01\_query.dat: 7 10 85 117 114 73 93 4 51 28

../../data/003\_AU12\_query.dat: 7 10 85 117 114 73 93 4 51 28

../../data/003\_AU17\_query.dat: 131 115 55 56 130 37 9 7 2 15

../../data/004\_AU01\_query.dat: 69 85 49 93 114 51 35 124 117 61

../../data/004\_AU12\_query.dat: 105 128 126 70 111 50 59 13 43 91

../../data/004\_AU17\_query.dat: 125 103 104 39 60 105 61 86 131 115

../../data/005\_AU01\_query.dat: 88 7 8 10 114 61 85 93 42 65

../../data/005\_AU12\_query.dat: 19 116 95 119 63 57 41 54 64 79

../../data/005\_AU17\_query.dat: 94 62 57 90 55 56 44 130 27 92

../../data/006\_AU01\_query.dat: 114 17 49 10 77 69 85 52 53 18

../../data/006\_AU12\_query.dat: 128 91 87 41 50 38 135 11 79 63

../../data/006\_AU17\_query.dat: 16 17 62 63 94 37 115 131 9 127

../../data/007\_AU01\_query.dat: 77 114 26 39 69 73 120 35 132 47

../../data/007\_AU12\_query.dat: 79 18 54 63 19 46 17 116 127 95

../../data/007\_AU17\_query.dat: 18 62 127 63 37 115 17 94 130 90

../../data/008\_AU01\_query.dat: 107 69 21 35 77 117 73 82 47 17

../../data/008\_AU12\_query.dat: 63 116 41 106 79 15 87 20 54 95

../../data/008\_AU17\_query.dat: 22 78 79 108 84 50 63 18 116 109

../../data/009\_AU01\_query.dat: 49 97 10 12 39 42 36 124 107 121

../../data/009\_AU12\_query.dat: 79 24 64 43 22 41 33 83 20 87

../../data/009\_AU17\_query.dat: 25 42 27 36 131 81 115 52 53 50

../../data/010\_AU01\_query.dat: 77 39 21 85 117 125 12 35 69 124

../../data/010\_AU12\_query.dat: 27 91 41 50 87 135 123 43 59 16

../../data/010\_AU17\_query.dat: 115 131 55 56 86 48 52 53 136 137

../../data/011\_AU01\_query.dat: 93 7 8 38 52 53 10 117 85 124

../../data/011\_AU12\_query.dat: 93 7 8 38 52 53 10 117 85 124

../../data/011\_AU17\_query.dat: 136 137 125 117 71 74 12 39 5 133

../../data/012\_AU01\_query.dat: 69 125 85 124 112 121 122 52 53 12

../../data/012\_AU12\_query.dat: 79 43 135 50 83 11 38 25 41 54

../../data/012\_AU17\_query.dat: 69 125 85 124 112 121 122 52 53 12

../../data/013\_AU01\_query.dat: 85 69 49 51 10 58 114 39 117 124

../../data/013\_AU12\_query.dat: 95 116 90 63 57 54 38 111 87 15

../../data/013\_AU17\_query.dat: 50 136 137 86 125 71 12 52 53 10

../../data/014\_AU01\_query.dat: 117 69 29 124 85 93 12 125 77 49

../../data/014\_AU12\_query.dat: 135 87 41 116 50 16 106 91 79 22

../../data/014\_AU17\_query.dat: 115 130 131 116 57 127 54 90 9 63

../../data/015\_AU01\_query.dat: 65 12 10 49 72 115 13 97 61 114

../../data/015\_AU12\_query.dat: 87 106 63 22 116 91 41 50 16 57

../../data/015\_AU17\_query.dat: 98 110 112 134 48 118 133 113 128 66

../../data/016\_AU01\_query.dat: 115 63 130 131 45 57 90 22 127 61

../../data/016\_AU12\_query.dat: 95 54 50 123 59 70 135 44 11 111

../../data/016\_AU17\_query.dat: 136 137 125 126 117 64 12 52 53 39

../../data/017\_AU01\_query.dat: 42 49 85 124 120 21 69 61 38 114

../../data/017\_AU12\_query.dat: 20 116 57 95 130 127 63 37 54 115

../../data/017\_AU17\_query.dat: 133 134 118 42 113 125 112 92 91 86

../../data/018\_AU01\_query.dat: 85 38 10 114 69 42 124 61 50 17

../../data/018\_AU12\_query.dat: 46 128 126 87 44 41 11 70 16 39

../../data/018\_AU17\_query.dat: 52 53 38 10 61 85 69 93 86 71

../../data/019\_AU01\_query.dat: 54 136 137 71 85 124 69 125 10 39

../../data/019\_AU12\_query.dat: 95 57 63 46 116 20 53 22 135 44

../../data/019\_AU17\_query.dat: 54 136 137 71 85 124 69 125 10 39

../../data/020\_AU01\_query.dat: 131 12 63 9 115 134 125 136 137 86

../../data/020\_AU12\_query.dat: 63 116 57 95 130 44 106 22 115 37

../../data/020\_AU17\_query.dat: 131 12 63 9 115 134 125 136 137 86

../../data/021\_AU01\_query.dat: 85 38 10 52 69 107 82 54 114 124

../../data/021\_AU12\_query.dat: 91 46 83 11 105 44 75 64 87 16

../../data/021\_AU17\_query.dat: 12 42 71 125 103 104 89 61 131 85

../../data/022\_AU01\_query.dat: 54 12 125 85 114 10 52 42 55 56

../../data/022\_AU12\_query.dat: 116 60 22 44 57 95 130 115 90 87

../../data/022\_AU17\_query.dat: 94 19 18 115 116 60 92 130 51 40

../../data/023\_AU01\_query.dat: 45 68 64 10 13 12 88 66 85 58

../../data/023\_AU12\_query.dat: 119 91 87 75 79 25 95 44 105 135

../../data/023\_AU17\_query.dat: 67 68 134 113 51 133 44 98 61 58

../../data/024\_AU01\_query.dat: 85 117 107 10 124 42 93 38 71 52

../../data/024\_AU12\_query.dat: 66 44 68 60 22 106 57 69 87 116

../../data/024\_AU17\_query.dat: 69 58 59 68 60 130 116 57 127 131

../../data/025\_AU01\_query.dat: 45 64 42 12 105 91 10 93 52 103

../../data/025\_AU12\_query.dat: 111 135 11 53 46 125 126 128 16 87

../../data/025\_AU17\_query.dat: 125 136 137 72 134 109 111 55 56 112

../../data/026\_AU01\_query.dat: 21 117 93 114 42 72 85 7 8 29

../../data/026\_AU12\_query.dat: 83 91 67 135 119 62 11 87 105 111

../../data/026\_AU17\_query.dat: 133 42 112 113 136 137 125 103 104 117

../../data/027\_AU01\_query.dat: 21 29 42 72 114 85 38 17 132 117

../../data/027\_AU12\_query.dat: 46 25 67 36 11 79 62 119 87 44

../../data/027\_AU17\_query.dat: 22 95 79 66 116 20 115 24 30 125

../../data/028\_AU01\_query.dat: 124 85 55 56 102 61 54 114 17 72

../../data/028\_AU12\_query.dat: 20 25 22 66 36 16 53 67 41 135

../../data/028\_AU17\_query.dat: 53 125 44 12 55 56 108 89 57 115

../../data/029\_AU01\_query.dat: 61 38 124 54 85 80 114 72 55 56

../../data/029\_AU12\_query.dat: 78 91 62 105 46 135 82 16 36 111

../../data/029\_AU17\_query.dat: 86 24 81 16 53 131 40 82 74 91

../../data/030\_AU01\_query.dat: 72 61 124 93 114 38 117 52 10 42

../../data/030\_AU12\_query.dat: 44 91 106 41 16 53 67 135 119 66

../../data/030\_AU17\_query.dat: 108 125 134 39 136 137 74 118 12 115

../../data/031\_AU01\_query.dat: 13 120 42 114 124 21 72 93 10 64

../../data/031\_AU12\_query.dat: 130 116 66 37 131 115 44 60 22 57

../../data/031\_AU17\_query.dat: 125 131 89 115 84 74 63 12 10 64

../../data/032\_AU01\_query.dat: 88 72 42 10 117 76 134 114 54 52

../../data/032\_AU12\_query.dat: 90 62 44 78 86 16 67 106 125 105

../../data/032\_AU17\_query.dat: 51 65 127 115 53 44 66 108 93 75

../../data/033\_AU01\_query.dat: 99 97 98 53 94 125 11 12 112 128

../../data/033\_AU12\_query.dat: 57 116 46 66 119 111 60 37 125 15

../../data/033\_AU17\_query.dat: 65 131 125 130 58 59 89 60 115 40

../../data/034\_AU01\_query.dat: 99 94 52 125 12 53 42 45 124 44

../../data/034\_AU12\_query.dat: 99 94 53 128 41 11 135 78 105 73

../../data/034\_AU17\_query.dat: 99 125 53 133 108 51 134 52 89 39

../../data/035\_AU01\_query.dat: 74 111 138 136 137 73 128 110 126 121

../../data/035\_AU12\_query.dat: 111 123 126 138 74 11 136 137 46 16

../../data/035\_AU17\_query.dat: 110 74 72 134 111 121 122 136 137 83

../../data/036\_AU01\_query.dat: 12 74 63 96 125 10 11 77 134 42

../../data/036\_AU12\_query.dat: 11 94 12 62 53 86 67 46 111 78

../../data/036\_AU17\_query.dat: 12 74 63 96 125 10 11 77 134 42

../../data/037\_AU01\_query.dat: 72 42 117 124 10 88 61 38 74 23

../../data/037\_AU12\_query.dat: 44 90 94 22 116 66 41 60 111 46

../../data/037\_AU17\_query.dat: 109 89 74 12 125 53 90 136 137 84

../../data/038\_AU01\_query.dat: 96 105 74 88 72 110 121 122 117 103

../../data/038\_AU12\_query.dat: 73 98 135 11 126 74 46 128 90 123

../../data/038\_AU17\_query.dat: 111 74 127 89 98 73 53 135 136 137

../../data/039\_AU01\_query.dat: 88 21 52 17 10 80 64 38 72 96

../../data/039\_AU12\_query.dat: 133 74 134 77 121 122 136 137 73 11

../../data/039\_AU17\_query.dat: 134 133 121 122 74 77 51 42 118 125

../../data/040\_AU01\_query.dat: 72 42 88 96 124 110 10 76 120 118

../../data/040\_AU12\_query.dat: 66 130 60 98 22 44 57 93 127 41

../../data/040\_AU17\_query.dat: 131 130 40 66 44 127 89 12 30 92

../../data/041\_AU01\_query.dat: 42 120 91 21 124 50 132 72 125 117

../../data/041\_AU12\_query.dat: 67 98 90 94 78 15 119 44 114 135

../../data/041\_AU17\_query.dat: 90 125 94 89 120 51 74 136 137 116

../../data/042\_AU01\_query.dat: 116 115 72 52 105 42 113 74 96 125

../../data/042\_AU12\_query.dat: 46 53 114 90 135 108 44 125 11 94

../../data/042\_AU17\_query.dat: 116 115 72 52 105 42 113 74 96 125

../../data/043\_AU01\_query.dat: 88 72 42 10 120 110 52 55 56 117

../../data/043\_AU12\_query.dat: 128 11 53 114 73 138 41 16 74 44

../../data/043\_AU17\_query.dat: 74 12 89 92 136 137 98 94 42 134

../../data/044\_AU01\_query.dat: 55 56 74 111 73 136 137 112 46 57

../../data/044\_AU12\_query.dat: 129 11 53 16 138 114 73 44 41 6

../../data/044\_AU17\_query.dat: 119 53 112 118 130 57 66 40 111 44

../../data/045\_AU01\_query.dat: 72 88 38 42 80 21 123 52 120 134

../../data/045\_AU12\_query.dat: 119 118 93 66 40 60 44 22 98 90

../../data/045\_AU17\_query.dat: 118 40 119 58 59 92 9 12 93 89

../../data/046\_AU01\_query.dat: 116 74 89 128 72 136 137 115 96 51

../../data/046\_AU12\_query.dat: 114 41 73 90 46 16 78 53 98 57

../../data/046\_AU17\_query.dat: 116 77 115 74 51 128 136 137 72 101

../../data/047\_AU01\_query.dat: 74 128 55 56 39 137 89 12 136 111

../../data/047\_AU12\_query.dat: 131 114 11 129 73 103 16 67 115 74

../../data/047\_AU17\_query.dat: 74 128 55 56 39 137 89 12 136 111

Run time: 7.56 s