

CS 4481A Assignment 3

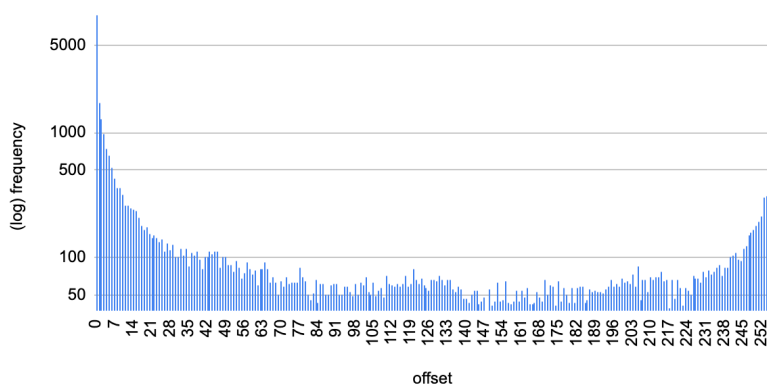
Sunday, November 17, 2024

Benjamin Asokanthan

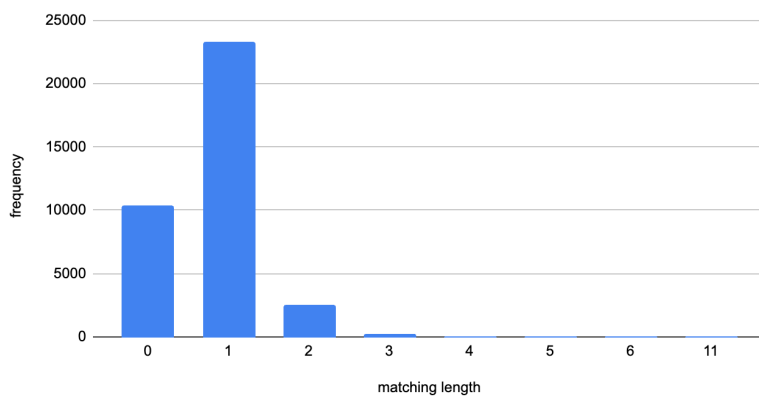
251172292

peppers.raw.pgm

(Logarithmic) Frequency of Offset Values for Peppers (search buffer length = 256)



Frequency of Matching Lengths for Peppers (search buffer length = 256)



Offset Average: 73.470406

Offset Standard Deviation: 90.404686

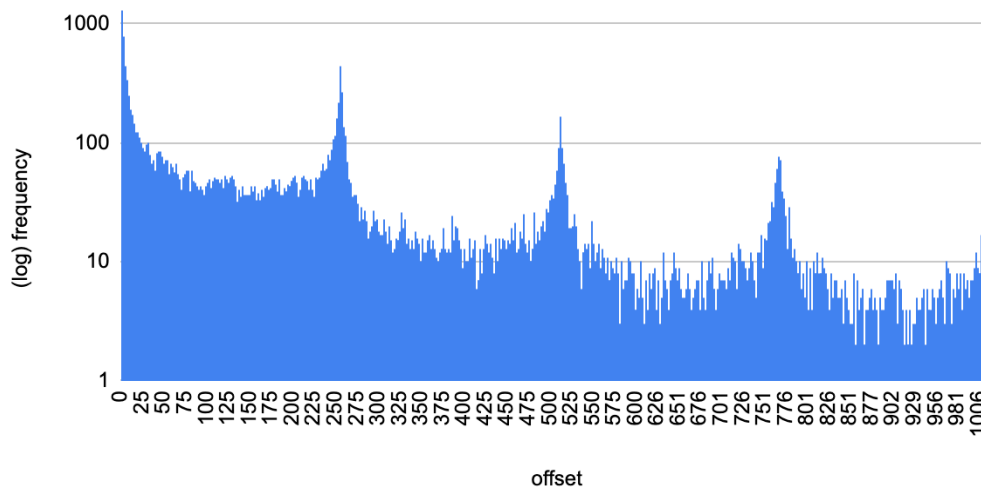
Match Length Average: 0.798216

Match Length Standard Deviation: 0.587050

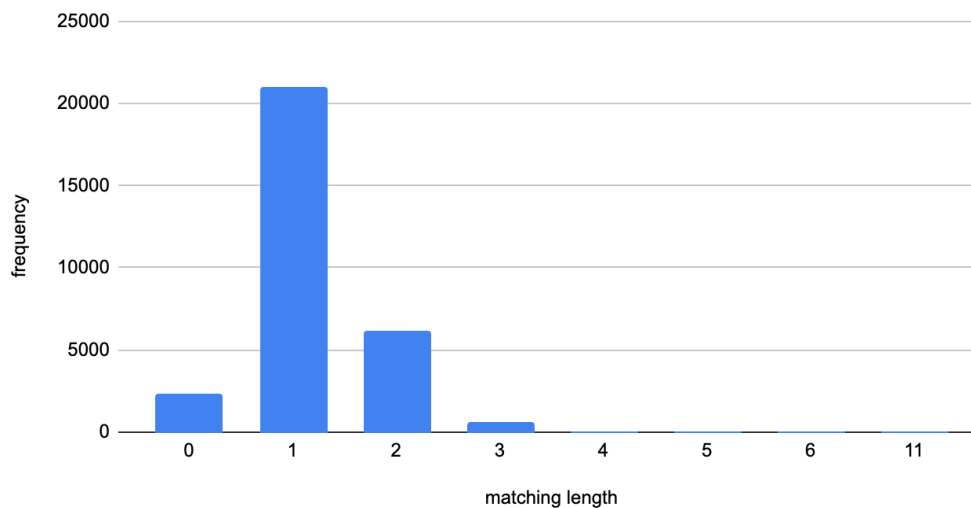
Compression Time: 43543 microseconds

Decompression Time: 5753 microseconds

(Logarithmic) Frequency of Offset Values for Peppers (search buffer length = 1024)



Frequency of Matching Lengths for Peppers (search buffer length = 1024)



Offset Average: 240.576599

Offset Standard Deviation: 258.510498

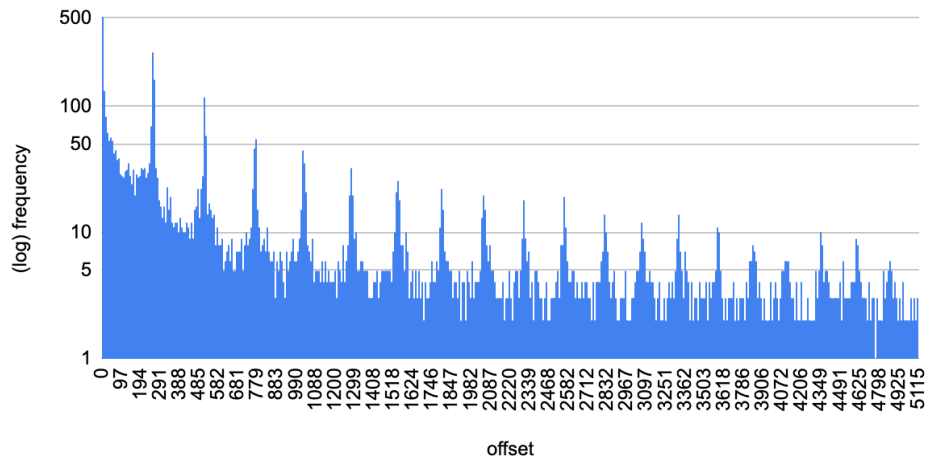
Match Length Average: 1.176658

Match Length Standard Deviation: 0.597689

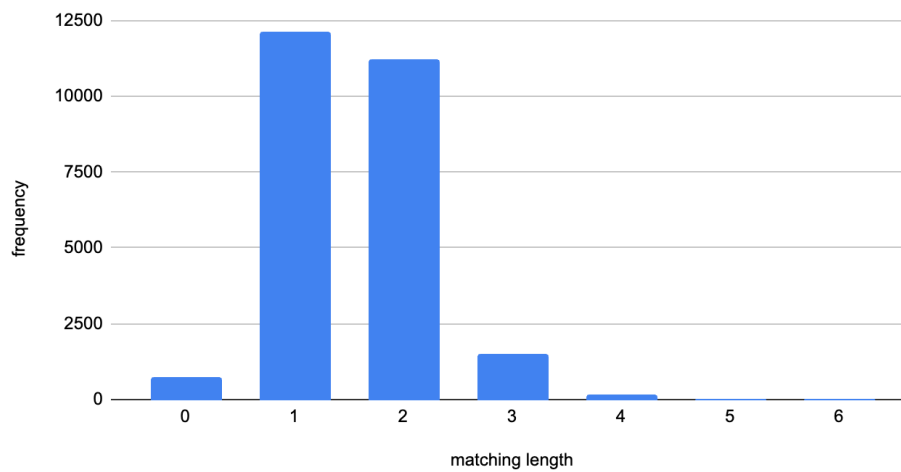
Compression Time: 187726 microseconds

Decompression Time: 6545 microseconds

(Logarithmic) Frequency of Offset Values for Peppers (search buffer length = 5120)



Frequency of Matching Lengths for Peppers (search buffer length = 5120)



Offset Average: 1034.755737

Offset Standard Deviation: 1297.178223

Match Length Average: 1.543941

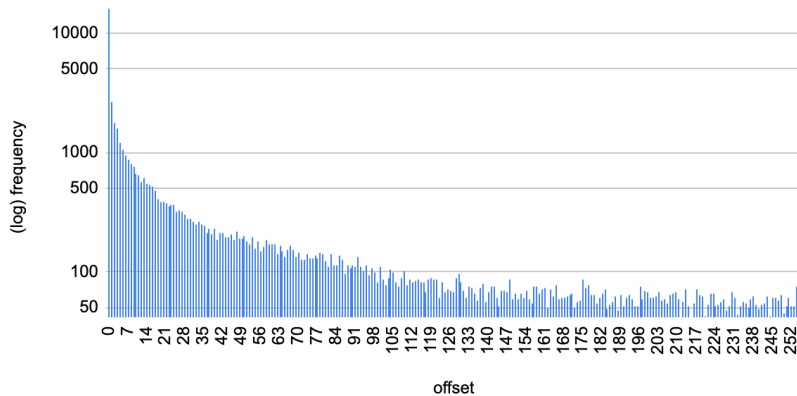
Match Length Standard Deviation: 0.683500

Compression Time: 546022 microseconds

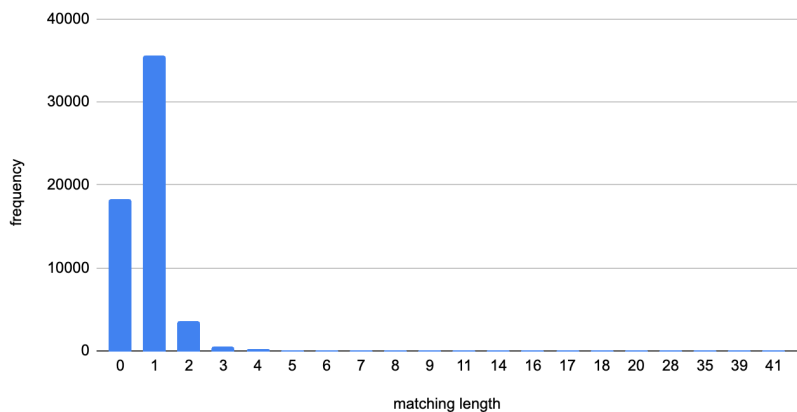
Decompression Time: 5487 microseconds

goldhill.raw.pgm

(Logarithmic) Frequency of Offset Values for Goldhill (search buffer length = 256)



Frequency of Matching Lengths for Goldhill (search buffer length = 256)



Offset Average: 45.372269

Offset Standard Deviation: 65.948509

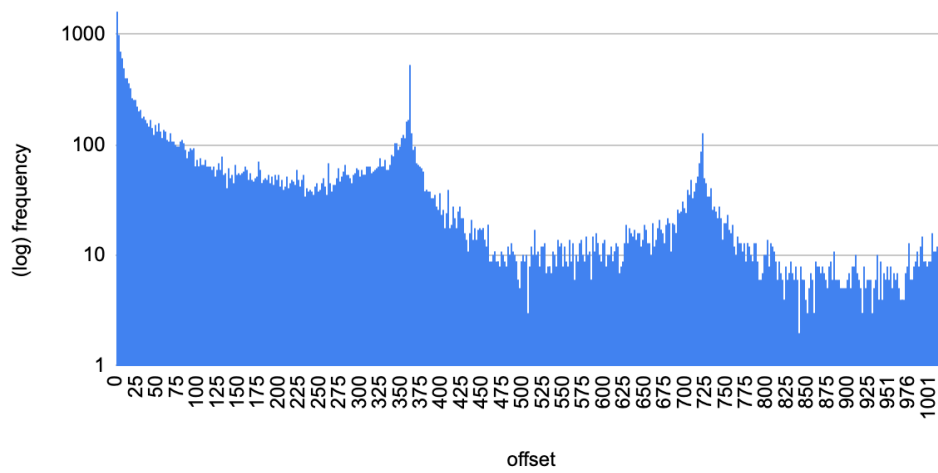
Match Length Average: 0.781198

Match Length Standard Deviation: 0.724496

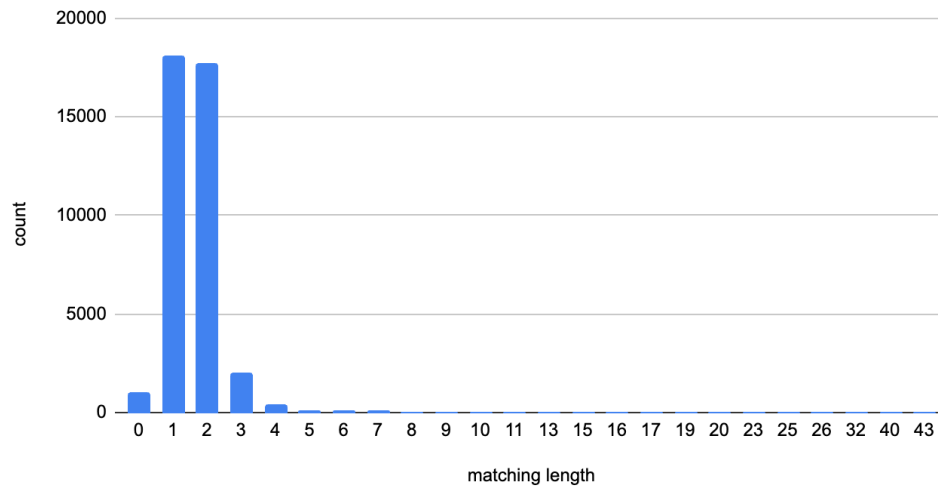
Compression Time: 69161 microseconds

Decompression Time: 8314 microseconds

(Logarithmic) Frequency of Offset Values for Goldhill (search buffer length = 1024)



Frequency of Matching Lengths for Goldhill (search buffer length = 1024)



Offset Average: 213.600327

Offset Standard Deviation: 242.650040

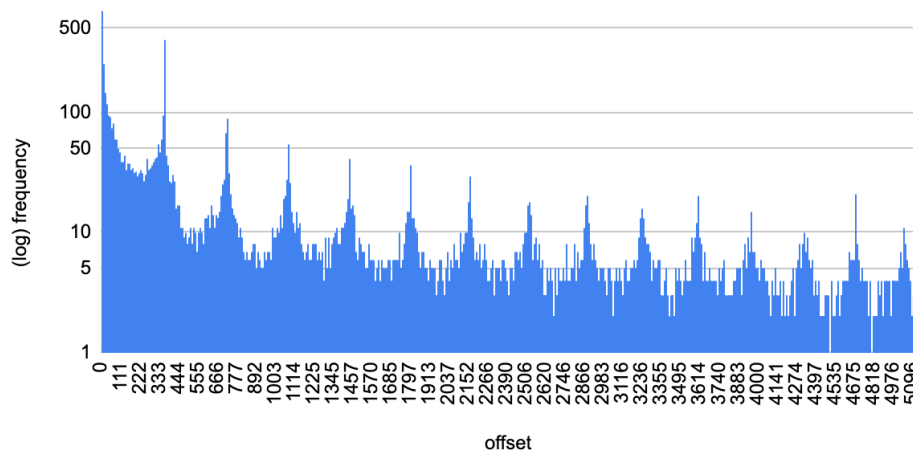
Match Length Average: 1.226565

Match Length Standard Deviation: 0.826137

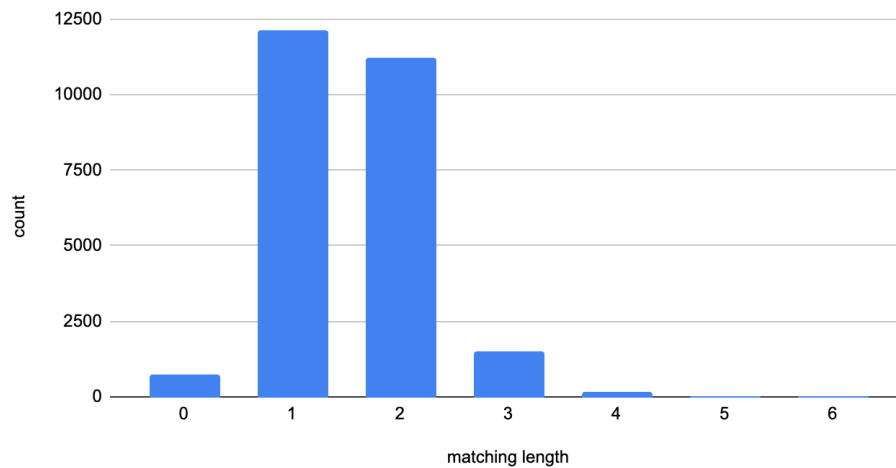
Compression Time: 201215 microseconds

Decompression Time: 10338 microseconds

(Logarithmic) Frequency of Offset Values for Goldhill (search buffer length = 5120)



Frequency of Matching Lengths for Goldhill (search buffer length = 5120)



Offset Average: 1042.253296

Offset Standard Deviation: 1312.707764

Match Length Average: 1.613891

Match Length Standard Deviation: 0.940284

Compression Time: 829058 microseconds

Decompression Time: 8983 microseconds

Performance Summary

Below is a table summarizing the performance results of my LZ77 Encoding and Decoding program for both test images at 256, 1024, and 5120 searching_buffer_size values. All decoded images were identical to the original images (MAE = 0.000000).

Image	<i>peppers.raw.pgm</i>			<i>goldhill.raw.pgm</i>		
<i>searching_buffer_size</i>	256	1024	5120	256	1024	5120
<i>Offset Average</i>	73.470406	240.576599	1034.755737	45.372269	213.600327	1042.253296
<i>Offset Standard Deviation</i>	90.404686	258.510498	1297.178223	65.948509	242.650040	1312.707764
<i>Match Length Average</i>	0.798216	1.176658	1.543941	0.781198	1.226565	1.613891
<i>Match Length Standard Deviation</i>	0.587050	0.597689	0.683500	0.724496	0.826137	0.940284
<i>Compression Time (microseconds)</i>	43543	187726	546022	69161	201215	829058
<i>Decompression Time (microseconds)</i>	5753	6545	5487	8314	10338	8983

Comments

The Shape of the Offset Histograms

The most frequent offset for both images encoded with LZ77 (at all 3 search buffer lengths) is 0. I believe this occurs since both images have many unique pixel values within the search buffer, even at the maximum search buffer length of 5120.

Additionally, even for repeated sections, LZ77 will always have to encode the first occurrence of the sequence of symbols with a 0 offset and 0 matching length.

I believe that the offset value histograms have more impulses/peaks as the search buffer length increases. I think the observed peaks in offset frequency occur because, as the search buffer increases, there are common semi-repeated pixel sequences separated by random pixel values.

Additionally, the LZ77 encoding algorithm flattens a 2D image into a 1D sequence of pixels to encode it. Repeated pixels in a column being separated by an entire row of pixels in the symbol sequence might explain the visual impulses in the offset histograms.

The Shape of the Matching Length Histograms

For both images encoded with LZ77 at all 3 search buffer lengths, the most frequent matching length is 1. As the search buffer increases, the match length histogram becomes less and less left skewed. This is seen in the matching length average increasing between the 3 search buffer lengths. The spread of the graph remains relatively consistent between the graphs. However, there are no instances of tokens with matching lengths of 11 in the LZ77 encoding of peppers using a search buffer length of 5120, even though there is in searching buffer length of 256 and 1024. I am not sure why this occurs, but the same phenomenon is shown even more drastically in the results using the goldhill image.

Final Recommendation

I believe that for both images, using LZ77 with a searching_buffer_size of 1024 displays the best results while maintaining a lower compression and decompression time cost. As searching buffer size increases, there is a dramatic compression performance cost as shown in the performance summary table. At this searching_buffer_size, the average matching length is above 1. Additionally, it was observed that there are tokens with higher matching lengths with a searching buffer size of 1024 for both images.

Practically, this means that there would be less tokens in the encoding, meaning that higher compression could be achieved with higher matching lengths when LZ77 is used in conjunction with a codeword compression algorithm. Searching_buffer_size at 1024 also does not see the huge performance hit seen at 5120, which is 2.6/4.1 slower at compression.