

# Compresión de archivos linearizados

Vicente Lermenda Candia

## Introducción

El objetivo del proyecto es probar los compresores gzip y bzip2 con distintas imágenes. Además, linealizando estas siguiendo el orden típico de row major, y el "locality friendly" z order (Morton curve). Esto con el fin de comparar los archivos ya comprimidos y verificar así, si existe una relación entre mayor localidad y mejor ratio compresión.

## Dataset

Se utiliza un dataset de kodak [1], y una imagen la cual es un cuadrado que cuenta con 1 color en cada cuadrante de este [2](nos referiremos a el en los resultados como "sq"). Esta última imagen se usa con el fin de probar que ambos métodos de linearización funcionan, además de ser un buen caso de prueba para el objetivo del proyecto.

Cada imagen se recorta para terminar siendo de 256x256 px, para así poder aplicar z-order.

## Descripción del programa

Se utilizan 2 programas escritos en C++ [3], uno que lineariza en row major order, y el otro en z-order.

Se utiliza la librería CImg, la cual permite trabajar de manera sencilla con imágenes, pudiendo así modificarlas y acceder a los valores de sus pixeles.

Se realizan 3 pruebas, cada una de estas obtiene un output distinto para cada imagen de input.

Para la primera prueba se condiera, de cada pixel de la imagen, sus valores  $r$ ,  $g$ ,  $b$ , y estos son almacenados como bytes independientes.

Una vez guardados, estos bytes se concatenan aplicando operaciones de bits.

Luego, en un archivo de texto plano se van guardando estos bytes concatenados, tanto en row major order, como z-order.

Para la segunda prueba, la única diferencia es que se crea un archivo binario en vez de texto plano.

Para la tercera prueba solo consideramos el byte  $r$ , y creamos archivos binarios.

## Ejemplo del programa

A continuación se muestra como ambos algoritmos linearizan la imagen *sq*, primero con orden row major, y después con Z order.

Ambas imágenes son de tamaño  $256 * 256 \times 1px$ .

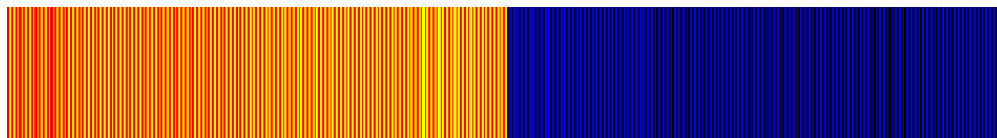


Figura 1: rm.png



Figura 2: zo.png

## Resultados

Para cada una de las pruebas descritas en la sección anterior, se comprime aplicando los programas gzip y bzip2.

Finalmente se aplicará sobre los archivos binarios un programa que nos entrega la entropía de orden  $k$  de un archivo [4]. En este caso  $k \in [0, 5]$ .

## Prueba 1

Image	og size (bytes)	gzip size	bzip2 size	gzip ratio	bzip2 ratio
1	476959	151052	123636	0.316698081	0.2592172493
2	484260	134033	97269	0.276779003	0.2008611077
3	477687	128557	107954	0.2691239242	0.2259931713
4	461716	146422	119160	0.317125679	0.2580807249
5	470120	178441	160930	0.379564792	0.3423168553
6	504611	122684	95019	0.243125893	0.1883014837
7	465120	87814	62280	0.1887985896	0.1339009288
8	471610	158057	137705	0.3351434448	0.2919891436
9	473886	153769	122804	0.3244852137	0.2591424942
10	479136	152936	118760	0.3191912108	0.2478628197
11	459958	113745	98145	0.2472943182	0.213378178
12	499824	109247	87538	0.218570937	0.1751376485
13	459536	159328	131909	0.3467149473	0.28704824
14	471662	164685	144399	0.3491589316	0.3061493188
15	505551	101492	72591	0.2007552156	0.1435878873
16	468956	94466	75258	0.201438941	0.1604798744
17	461311	124401	90441	0.2696684016	0.196052121
18	461659	162416	135416	0.3518094524	0.2933247267
19	471063	174003	143625	0.369383713	0.3048955235
20	517995	89276	63659	0.1723491539	0.1228950086
21	461208	111619	88303	0.242014449	0.1914602522
22	483567	177029	153563	0.366089911	0.3175630264
23	470157	138875	124243	0.2953800539	0.2642585349
24	484764	127382	106585	0.2627711629	0.2198698748
sq	327680	1109	101	0.003384399414	0.0003082275391

Cuadro 1: Prueba 1 row major order

Image	og size (bytes)	gzip size	bzip2 size	gzip ratio	bzip2 ratio
1	476959	155980	128307	0.327030206	0.2690105439
2	484260	141376	104082	0.291942345	0.2149299963
3	477687	128033	112196	0.2680269716	0.2348734632
4	461716	151256	123416	0.3275953183	0.2672985125
5	470120	179210	164149	0.3812005445	0.3491640432
6	504611	131009	99510	0.2596237498	0.1972014086
7	465120	90720	69712	0.1950464396	0.149879601
8	471610	161036	140348	0.3414601047	0.2975933504
9	473886	150089	123813	0.316719633	0.2612716983
10	479136	144443	118626	0.3014655547	0.2475831497
11	459958	123639	102189	0.2688049778	0.2221702851
12	499824	124393	101502	0.2488736035	0.2030754826
13	459536	161431	133609	0.3512913025	0.2907476237
14	471662	170518	149920	0.3615258384	0.317854735
15	505551	96980	74702	0.1918303	0.1477635293
16	468956	109356	83761	0.2331903206	0.1786116395
17	461311	126416	97714	0.2740363876	0.2118180577
18	461659	163743	139490	0.3546838684	0.302149422
19	471063	170963	147877	0.3629302238	0.313921917
20	517995	90521	65229	0.1747526521	0.1259259259
21	461208	119617	92144	0.2593558655	0.1997883818
22	483567	175524	154166	0.3629776225	0.3188100098
23	470157	135570	125566	0.2883504872	0.2670724886
24	484764	125421	108377	0.2587258955	0.223566519
sq	327680	539	100	0.001644897461	0.0003051757813

Cuadro 2: Prueba 1 Z order

## Prueba 2

Image	og size (bytes)	gzip size	bzip2 size	gzip ratio	bzip2 ratio
1	262144	132953	101995	0.5071754456	0.3890800476
2	262144	116016	83199	0.442565918	0.3173789978
3	262144	115362	88452	0.440071106	0.3374176025
4	262144	129582	97687	0.4943161011	0.3726463318
5	262144	167553	143456	0.6391639709	0.5472412109
6	262144	109000	81415	0.415802002	0.3105735779
7	262144	77367	54170	0.2951316833	0.2066421509
8	262144	143349	115479	0.5468330383	0.4405174255
9	262144	140298	105910	0.535194397	0.4040145874
10	262144	136455	97756	0.5205345154	0.3729095459
11	262144	102406	82694	0.3906478882	0.3154525757
12	262144	94377	72311	0.3600196838	0.275844574
13	262144	143495	112049	0.5473899841	0.4274330139
14	262144	149291	125015	0.5694999695	0.4768943787
15	262144	88480	62782	0.3375244141	0.2394943237
16	262144	84114	63123	0.3208694458	0.2407951355
17	262144	110088	75911	0.4199523926	0.2895774841
18	262144	147376	116473	0.5621948242	0.4443092346
19	262144	159286	119621	0.6076278687	0.4563179016
20	262144	78762	57905	0.300453186	0.2208900452
21	262144	98115	76063	0.3742790222	0.2901573181
22	262144	158514	129582	0.6046829224	0.4943161011
23	262144	126782	103154	0.4836349487	0.3935012817
24	262144	114729	93887	0.4376564026	0.3581504822
sq	262144	1358	80	0.005180358887	0.0003051757813

Cuadro 3: Prueba 2 row major order

Image	og size (bytes)	gzip size	bzip2 size	gzip ratio	bzip2 ratio
1	262144	136726	106791	0.5215682983	0.4073753357
2	262144	120903	89152	0.4612083435	0.3400878906
3	262144	112989	91410	0.4310188293	0.3487014771
4	262144	132683	101546	0.5061454773	0.3873672485
5	262144	166558	147261	0.6353683472	0.561756134
6	262144	115382	85775	0.4401473999	0.327205658
7	262144	79478	61366	0.3031845093	0.2340927124
8	262144	145065	118151	0.5533790588	0.4507102966
9	262144	134779	107479	0.5141410828	0.4099998474
10	262144	127830	97676	0.4876327515	0.3726043701
11	262144	109312	86271	0.4169921875	0.3290977478
12	262144	106019	84375	0.4044303894	0.3218650818
13	262144	144188	113463	0.5500335693	0.4328269958
14	262144	154029	130043	0.5875740051	0.4960746765
15	262144	85266	64521	0.3252639771	0.2461280823
16	262144	97132	70716	0.3705291748	0.2697601318
17	262144	110320	82659	0.4208374023	0.3153190613
18	262144	147850	121466	0.5640029907	0.4633560181
19	262144	155161	124908	0.5918922424	0.4764862061
20	262144	80147	59524	0.3057365417	0.22706604
21	262144	104671	79565	0.3992881775	0.3035163879
22	262144	157113	130490	0.5993385315	0.4977798462
23	262144	120801	105125	0.4608192444	0.40102005
24	262144	112041	95687	0.4274024963	0.3650169373
sq	262144	333	67	0.001270294189	0.0002555847168

Cuadro 4: Prueba 2 Z order

### Prueba 3

Image	og size (bytes)	gzip size	bzip2 size	gzip ratio	bzip2 ratio
1	262144	76192	49789	0.2906494141	0.1899299622
2	262144	64964	41981	0.2478179932	0.1601448059
3	262144	65608	40671	0.2502746582	0.1551475525
4	262144	67272	43821	0.2566223145	0.1671638489
5	262144	82351	50585	0.3141441345	0.1929664612
6	262144	50256	31623	0.1917114258	0.1206321716
7	262144	49876	28681	0.1902618408	0.1094093323
8	262144	76875	51015	0.2932548523	0.194606781
9	262144	78552	48954	0.2996520996	0.1867446899
10	262144	77478	46435	0.2955551147	0.1771354675
11	262144	58930	36916	0.2248001099	0.1408233643
12	262144	53875	33583	0.2055168152	0.1281089783
13	262144	76623	50471	0.2922935486	0.1925315857
14	262144	78037	48730	0.2976875305	0.1858901978
15	262144	57173	33812	0.2180976868	0.1289825439
16	262144	51853	32230	0.1978034973	0.1229476929
17	262144	64268	38443	0.2451629639	0.146648407
18	262144	73237	45553	0.2793769836	0.1737709045
19	262144	82310	50312	0.3139877319	0.1919250488
20	262144	19132	10927	0.07298278809	0.04168319702
21	262144	52441	34288	0.2000465393	0.1307983398
22	262144	74912	48545	0.2857666016	0.1851844788
23	262144	61351	36080	0.2340354919	0.1376342773
24	262144	61169	38940	0.233341217	0.1485443115
sq	262144	300	53	0.00114440918	0.0002021789551

Cuadro 5: Prueba 3 Row major order

Image	og size (bytes)	gzip size	bzip2 size	gzip ratio	bzip2 ratio
1	262144	78401	53267	0.2990760803	0.2031974792
2	262144	68939	46281	0.2629814148	0.1765480042
3	262144	64827	43367	0.2472953796	0.1654319763
4	262144	69242	46318	0.2641372681	0.1766891479
5	262144	82090	53991	0.3131484985	0.2059593201
6	262144	54054	35109	0.206199646	0.1339302063
7	262144	51088	33315	0.1948852539	0.1270866394
8	262144	77745	52476	0.2965736389	0.2001800537
9	262144	74166	50188	0.2829208374	0.1914520264
10	262144	70456	46418	0.2687683105	0.1770706177
11	262144	61247	39703	0.2336387634	0.1514549255
12	262144	61204	40741	0.2334747314	0.1554145813
13	262144	76363	51542	0.2913017273	0.1966171265
14	262144	80927	53129	0.3087120056	0.202671051
15	262144	52985	34678	0.2021217346	0.1322860718
16	262144	58409	37398	0.2228126526	0.1426620483
17	262144	64960	42502	0.2478027344	0.1621322632
18	262144	73279	48514	0.2795372009	0.1850662231
19	262144	82338	54995	0.3140945435	0.2097892761
20	262144	19055	11825	0.0726890564	0.04510879517
21	262144	55128	36019	0.2102966309	0.1374015808
22	262144	73682	49101	0.2810745239	0.1873054504
23	262144	58143	36860	0.2217979431	0.1406097412
24	262144	59316	39912	0.226272583	0.1522521973
sq	262144	302	51	0.001152038574	0.0001945495605

Cuadro 6: Prueba 3 Z order



## Entropía archivos binarios

Image	orden 0	orden 1	orden 2	orden 3	orden 4	orden 5
1	7.188806	5.438693	2.420917	1.418368	0.464160	0.251924
2	6.939660	4.314920	2.318815	1.584496	0.893043	0.535578
3	7.394490	5.104286	1.862456	1.062980	0.518734	0.369840
4	7.088059	5.340648	2.344089	1.408990	0.609057	0.266393
5	7.625689	6.143482	2.430864	0.727777	0.215834	0.132593
6	6.452739	4.448452	2.004497	1.172581	0.596007	0.516116
7	6.627994	3.664690	1.451188	1.076682	0.820157	0.699894
8	7.306041	5.397482	2.716380	1.278084	0.389883	0.188958
9	7.344970	5.093158	2.563459	1.399264	0.495611	0.247833
10	7.437764	5.027316	2.642952	1.409195	0.483475	0.227003
11	6.819584	4.301292	1.843178	0.946529	0.537587	0.427836
12	7.017421	4.721420	2.248342	1.239880	0.722926	0.512826
13	6.910892	5.760564	2.779643	1.479095	0.323894	0.111146
14	7.389702	5.773725	2.484375	1.090785	0.341259	0.183495
15	7.232955	3.766305	1.626102	1.137312	0.788149	0.656533
16	6.865355	4.509855	1.673725	1.118164	0.754593	0.612122
17	7.089272	4.270147	2.490357	1.565461	0.818320	0.510742
18	6.673363	5.614590	2.700331	1.237081	0.561466	0.335098
19	7.496556	5.779821	2.643339	1.272843	0.263658	0.076596
20	4.421442	3.223655	1.987428	1.194059	1.030573	0.920382
21	6.552065	4.674768	2.172645	1.417956	0.955805	0.658645
22	7.330686	6.042357	2.600653	0.908096	0.313836	0.171611
23	7.422263	5.635144	2.065096	0.939117	0.530301	0.339851
24	6.271956	4.243792	1.741049	0.829757	0.417640	0.333375
sq	0.918296	0.906710	0.863700	0.024354	0.024308	0.024301

Cuadro 7: Entropía orden Row Major

Image	orden 0	orden 1	orden 2	orden 3	orden 4	orden 5
1	7.188806	5.483077	2.498026	1.454061	0.424003	0.204518
2	6.939660	4.374284	2.448955	1.709138	0.871825	0.471495
3	7.394490	5.125255	1.955061	1.136625	0.515550	0.338695
4	7.088059	5.377328	2.416957	1.475497	0.573877	0.219833
5	7.625689	6.188331	2.464421	0.760315	0.195861	0.105970
6	6.452739	4.527854	2.029463	1.220229	0.588686	0.511394
7	6.627994	3.765664	1.592033	1.202009	0.794862	0.656853
8	7.306041	5.412529	2.744970	1.292557	0.353390	0.152705
9	7.344970	5.099040	2.548935	1.399336	0.508988	0.253053
10	7.437764	5.015518	2.625702	1.404517	0.504502	0.235750
11	6.819584	4.341387	1.927824	1.018039	0.536440	0.403976
12	7.017421	4.870315	2.488948	1.469976	0.644766	0.371099
13	6.910892	5.776876	2.803341	1.493298	0.307952	0.095629
14	7.389702	5.841179	2.566859	1.160497	0.279762	0.114952
15	7.232955	3.781690	1.679851	1.196489	0.807833	0.650552
16	6.865355	4.587818	1.877590	1.314147	0.813621	0.587457
17	7.089272	4.369084	2.611578	1.677394	0.759531	0.420665
18	6.673363	5.646113	2.789177	1.295983	0.525368	0.284362
19	7.496556	5.837145	2.665593	1.280388	0.224246	0.050417
20	4.421442	3.233471	2.012877	1.238137	1.052846	0.932157
21	6.552065	4.682622	2.239852	1.500906	0.976217	0.622282
22	7.330686	6.050552	2.599316	0.917091	0.318876	0.168090
23	7.422263	5.647870	2.089270	0.954411	0.522521	0.332038
24	6.271956	4.274565	1.796791	0.896772	0.457395	0.343614
sq	0.918296	0.906710	0.863700	0.000482	0.000269	0.000249

Cuadro 8: Entropía Z order

## Conclusión

No se ve la dominancia que se espera al aplicar los compresores sobre los archivos linealizados con Z order.

La entropía de orden 0 es la misma para ambas linealizaciones tal como lo esperado, más aún, para ordenes mayores, en especial el orden 5, se apreciaba una tendencia donde las imágenes linealizadas con Z order tienen entropía menor.

Dado que la entropía de las imágenes linealizadas respalda la hipótesis planteada, algo indeterminado ocurre en el proceso de compresión con gzip y bzip2.

## Referencias

- [1] Kodak dataset  
<https://www.site.uottawa.ca/~edubois/demosaicking/>
- [2] Cuadrado de 4 colores.  
<https://i.stack.imgur.com/Zsy0G.png?s=256>
- [3] Códigos utilizados para los programas de linearización  
<https://github.com/vlermandac/proyecto>
- [4] Programa utilizado para calcular la entropía  
<http://pizzachili.dcc.uchile.cl/experiments.html>