ICS-E4020: Week 2 - Correlated pairs

Néstor Castillo García (472081)
nestor.castillogarcia@aalto.fi

April 25, 2015

1 Correlated Pairs

1.1 Description

In this task, m input vectors with n numbers are given. The correlations between all pairs of input vectors is calculated. A red pixel at (i, j) indicates a positive correlation between vectos i and j, and a blue pixel indicates a negative correlation. The correlation coefficient r can be calculated in the following way.

$$SS_{xx} = \sum x^2 - n\bar{x}^2 \tag{1}$$

$$SS_{yy} = \sum y^2 - n\bar{y}^2 \tag{2}$$

$$SS_{xy} = \sum xy - n\bar{x}bary \tag{3}$$

$$r^2 = \frac{ss_{xy}^2}{ss_{xx}ss_{yy}} \tag{4}$$

1.2 Implementation

For practical and efficiency reasons. The previous calculations were simplified by normalizing the input matix such that each the sum of squares in each row was one and its mean zero. Such normalization simplifies the correlation coefficient and the calculation is simplified to the following equation.

$$r^2 = \sum xy \tag{5}$$

1.3 Hardware

The computers had the following specifications: Intel Xeon E3-1230v2, 4 cores, 8 thread, 3,3 GHz, RAM: 16 GB, GPU: Nvidia K2000.

1.4 Performance

The computationally intensive part of the correlation image is the dot product, that in the vector solution was efficiently improved by using AVX operations. The matrix normalization is done in O(n) time.

As expected, performance increased with the number of threads. The multi-threded version was in average 2,6 times faster than the single threaded solution for the non vectorized code (CP2). In contrast, the multithreaded version of the vectorized code (CP3) was 4,9 times faster than the single threaded vectorized version.
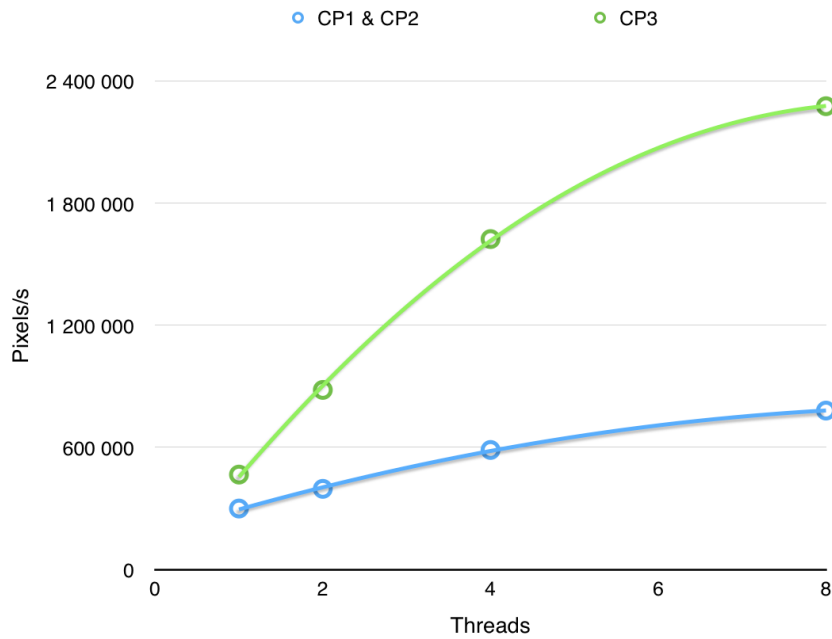

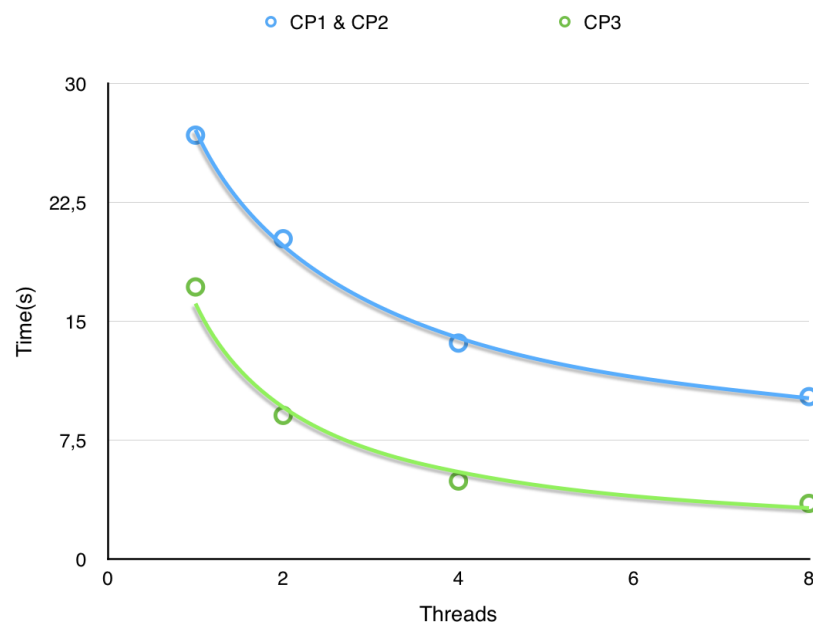Figure 1: Pixels per second in a 4000 x 4000 image

Figure 2: Processing time in a 4000 x 4000 image

		OMP_NUM_THREADS=1	OMP_NUM_THREADS=2	OMP_NUM_THREADS=4	OMP_NUM_THREADS=8
1	1	0	0	0	0
1	10	0	0	0	0
1	100	0	0	0	0
1	1000	0	0	0	0
1	2000	0	0	0	0
1	4000	0	0	0	0
10	1	0	0	0	0
10	10	0	0	0	0
10	100	0	0	0	0
10	1000	0	0	0	0
10	2000	0	0	0	0
10	4000	0,001	0,001	0	0
100	1	0	0	0	0
100	10	0	0	0	0
100	100	0,001	0,001	0,001	0,001
100	1000	0,011	0,008	0,006	0,003
100	2000	0,02	0,016	0,009	0,006
100	4000	0,019	0,02	0,018	0,012
1000	1	0,002	0,002	0,002	0,003
1000	10	0,003	0,004	0,005	0,004
1000	100	0,039	0,029	0,03	0,02
1000	1000	0,413	0,311	0,189	0,119
1000	2000	0,842	0,635	0,378	0,222
1000	4000	1,689	1,268	0,829	0,55
2000	1	0,013	0,018	0,009	0,01
2000	10	0,014	0,015	0,012	0,017
2000	100	0,153	0,117	0,072	0,052
2000	1000	1,682	1,272	0,8	0,458
2000	2000	3,356	2,526	1,599	1,062
2000	4000	6,714	5,062	3,307	2,37
4000	1	0,061	0,047	0,04	0,04
4000	10	0,067	0,053	0,039	0,027
4000	100	0,612	0,456	0,309	0,155
4000	1000	6,745	5,107	3,242	2,135
4000	2000	13,387	10,11	6,628	4,834
4000	4000	26,734	20,203	13,625	10,246

Figure 3: CP2 benchmark result

		CP3			
		OMP_NUM_THREADS=1	OMP_NUM_THREADS=2	OMP_NUM_THREADS=4	OMP_NUM_THREADS=8
1	1	0	0	0	0
1	10	0	0	0	0
1	100	0	0	0	0
1	1000	0	0	0	0
1	2000	0	0	0	0
1	4000	0	0	0	0
10	1	0	0	0	0
10	10	0	0	0	0
10	100	0	0	0	0
10	1000	0	0	0	0
10	2000	0,001	0	0	0
10	4000	0,001	0,001	0	0
100	1	0	0	0	0
100	10	0	0	0	0
100	100	0,001	0,001	0	0
100	1000	0,006	0,003	0,003	0,001
100	2000	0,01	0,006	0,003	0,003
100	4000	0,011	0,013	0,007	0,006
1000	1	0,012	0,007	0,01	0,007
1000	10	0,013	0,009	0,01	0,003
1000	100	0,028	0,014	0,011	0,006
1000	1000	0,185	0,096	0,073	0,043
1000	2000	0,424	0,234	0,155	0,106
1000	4000	1,014	0,566	0,432	0,216
2000	1	0,048	0,026	0,015	0,015
2000	10	0,051	0,028	0,018	0,011
2000	100	0,113	0,057	0,042	0,024
2000	1000	0,853	0,48	0,294	0,184
2000	2000	1,835	0,997	0,57	0,427
2000	4000	4,223	2,272	1,229	0,859
4000	1	0,208	0,108	0,064	0,04
4000	10	0,217	0,117	0,076	0,042
4000	100	0,447	0,225	0,12	0,092
4000	1000	3,758	2,068	1,116	0,742
4000	2000	7,659	4,041	2,259	1,713
4000	4000	17,163	9,066	4,926	3,513

Figure 4: CP3 benchmark result

I