HW2 Report

themis12@kaist.ac.kr / 010 2830 3651
20184448 Jo MinKi

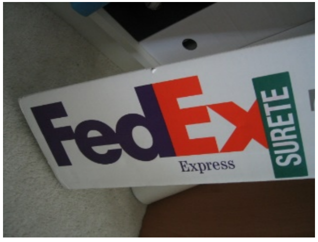
1 Introduction

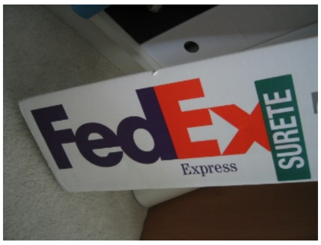
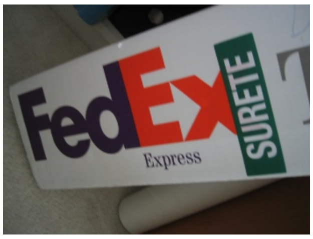
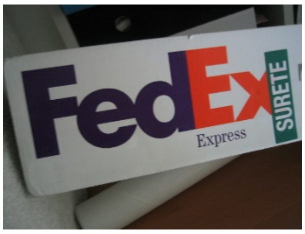
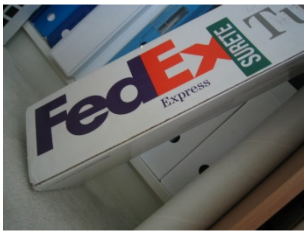
This work conduct the image searching task with Quantization.

2 Result

Correct case

Query

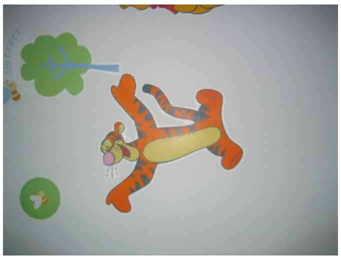


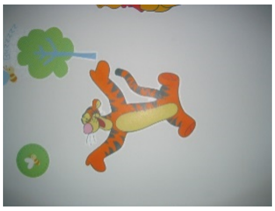
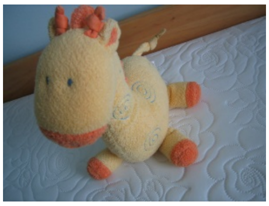
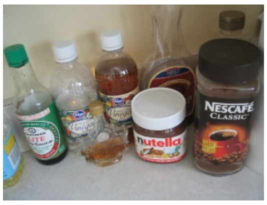
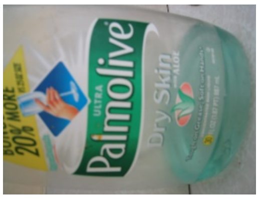





dist = [3.6398244, 90.92494, 107.60222, 110.9077]

Error cases

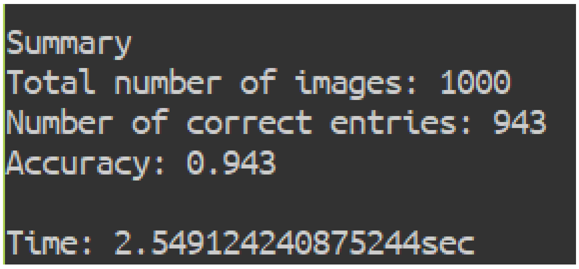
Query



dist = [4.801503, 88.44673, 96.51866, 97.753426]

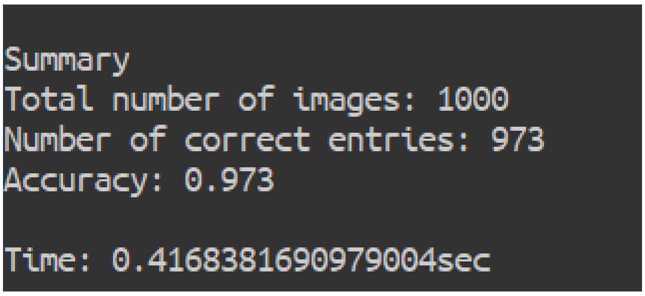
VGG16_fc7 layer / Exhaustive search



```
Summary
Total number of images: 1000
Number of correct entries: 943
Accuracy: 0.943

Time: 2.549124240875244sec
```

VGG16_fc7 layer / index search



```
Summary
Total number of images: 1000
Number of correct entries: 973
Accuracy: 0.973

Time: 0.4168381690979004sec
```

The exhaustive search from the homework 1 used the 4096 features for calculation of the L1 or L2 distance between the data. The data itself has a big size, and the calculation process also required some times and most importantly, it calculate the all distances for every data which makes too much unnecessary job. In order to make it efficient, the data quantized samler binary sequence, and it conduct the product quantization and search the entries by finding the close data with calculation of the distance between them.

Through the process, the row data which has 4096 of 32bit floating point was converted to the 32 of 8bit unsigned int, which require only 2416 bit. In other word, this process compress the data 2^9times. This brings a huge advantage to the runtime memory and running time. In addition, comparing the distance between two data could be more simple. If the process optimized well, then it can calculate the distance by just a bit level XOR operation.

Accordingly, the search performance got dramatic improvement compare to the exhaustive search. The running time got 6 times faster, but in high probability, it could be much better if the system gets bigger dataset. In this time the system only handled 1000 data so the size of the data is not enough to observe the improvement. Meanwhile, the accuracy also improved. In the case of the naïve L2 distance, the measure could be distorted by some critical value. However, the quantization divides the space and made it binary, so if there are some critical value in the vector, it gives smaller influence on the distance which makes the model more stable.

3 Discussion

Exhaustive search

For the last assignment, any compression or coding method and fancy data structure has been used. It was fine for this assignment which require searching only 1000 images. However, when the number of the image gets bigger such like a search engine, this method will not work.

One of the solution is quantization. You can quantize the weight and find the cluster based on the distance. If you choose the exhaustive way for quantization, then it also takes n^2 times, but it must be much faster when you conduct a search. (If you group each portion with 256 images, it will only take 256 time. Furthermore, there are many ways that conduct the quantization in $\mathrm{n}\log n$ time.

However, it seems the improvement of the accuracy seems too much dramatic. It's because the calculation process at the homework 1 has some imperfectness, and some influence by small number of data. It requires bigger dataset to observe precisely.