Project # 3. Multimedia Coding

Alexandre Rodrigues (2039952)

January 11, 2022

1 Introduction

This reports is dedicated to explain the usage of the LBG-split algorithm and how to implement it. The Linde-Buzo-Gray algorithm is a lossy coding technique, it uses vector quantization, meaning that a block of input samples (size L) is processed all together. ...

2 Technical Approach

Each vector of size L samples can be defined as

$$x = [x_1, x_2, \ldots, x_L], x \in R^L \tag{1}$$

where $x_i, i = 1, 2, \ldots, L$ are input samples. Using y_i as a codevector,

$$B = \{y_1, y_2, \ldots, y_L\} \tag{2}$$

is the set of reconstruction levels, i.e. the codebook, of size K. The decision cells will be

$$I_i \in R^L, i = 1, 2, \ldots, K, suchthat I_i intersect I_j = 0 \aleph i \neq j and union ki = 1 I_i = R^L \tag{3}$$

3 Results

There were several tests made to fully understand the performance of this method.

3.1 Training Performance

The tests were made using $\epsilon = 0.01$.

Training set	**2,2**	**2,4**	**4,1**	**4,2**
All Audio	1.77×10^5	2.12×10^5	4.75×10^5	2.85×10^5
Music: Say Nada	4.36×10^6	3.70×10^6	1.63×10^7	2.37×10^7
All Music	??	??	??	??
All Music and Audio	??	??	??	??

Table 1: Distortion for each training set and each values of L and R

Training set	**2,2**	**2,4**	**4,1**	**4,2**
All Audio	$0.45s$	$101.49s$	$0.32s$	$84.795s$
Music: Say Nada	$0.73s$	$115.93s$	$0.39s$	$95.22s$
All Music	??	??	??	??
All Music and All Audio	??	??	??	??

Table 2: Time for each training set and each values of L and R

We can see that (4,1) is clearly the fastest training. The training time is mostly dependent on the value K. The distortion is noticeably larger for L = 4.

3.2 Encoding Performance

In summary I got the following results:

Encoded	**2,2**	**2,4**	**4,1**	**4,2**
70mono	$1.20s$	$16.79s$	$0.60s$	$8.705s$
Average Music	$13.40s$	$193.61s$	$6.66s$	$104.92s$
Worst Case	$18.40s$	$232.67s$	$8.22s$	$118.26s$
Best Case	$10.89s$	$149.41s$	$5.52s$	$91.04s$

Table 3: Time for each training set and each values of L and R

There is no noticeable difference when using All Audio or Say Nada as the training set.

Encoded	**2,2**	**2,4**	**4,1**	**4,2**
70mono	5×10^5	2×10^4	4.75×10^5	8×10^4
Average Music	3.72×10^6	5.77×10^5	4.66×10^6	8.45×10^5
Worst Case	1.15×10^7	8.20×10^5	1.36×10^7	2.74×10^6
Best Case	6.2×10^5	3.10×10^4	9.42×10^5	1.07×10^5

Table 4: Distortion for each training set and each values of L and R

Encoding music files using All Audio as training se gives us in average 4 times the distortion when using Say Nada as the training set. For (2,4) we get actually 11 times the distortion.

3.3 Training set including the encoding object...

4 Conclusions

1 - Using music as a training set is clearly superior when we will use the codebook to encode music files. Vice-versa is also valid, although less significantly. Encoding 70mono using All audio as training set produces in average half the distortion.

2 -

3 -

4 -