Audio and Music Processing Lab

Essentia and Freesound

12/04/2021

Nicolás Schmidt

In this lab my main objective was to have a reconstruction of the audio target as similar as possible with respect to rhythm. However, I did not want to have too many sonic resources to reconstruct the target audio too well. One very exciting part about mosaicing, from my personal perspective, is being able to identify the sources used to re-create the target sound. In contrast to the concrete music, where the sound source is not supposed to be identifiable, in this exercise the interesting is to re-version the target sound with some other that could or could not share sonic features.

In this assignment I was especially interested to work on one particular suggestion written in the task description. The one that refers to trying to reconstruct the sound synchronously. This means, to be able to create a sound that has a similar rhythm and tempo. In addition to this, the fact that the audio of Notebook number 3 has very attractive rhythmic features made me go for this challenge.

Considering this problem, the first thing I did was to change the order of the analysis of target and sources, so that the target would be analyzed first and the sources later. From this first analysis, and complemented with the use of a method to estimate the beat of the target, I used the value of the beat as a quantized grid. One of the parameters of the *compute_frame_size_from_beat* function that I implemented is the resolution, which refers to how many subdivisions of the measure will be. After this I was able to estimate the *frame_size* as a fixed number aligned with the beat of the target audio.

Next, I proceeded to implement 3 additional feature extraction algorithms from Essentia. Considering the previously 2 implemented algorithms in Jupyter Notebook number 2, my final work considers 5 different features: Loudness, MFCC, MelBands, OnsetDetection and HPCP.

I thought that the Loudness feature, implemented previously in the Notebook, was a feature that has to be preserved in order to obtain a nice sonic analysis of the target and the source files. Loudness is a measure of volume that makes perfect sense to be involved in the feature comparison to perform good mosaicing. Considering this, I kept this feature as part of the analysis.

On the other hand, the MFCC features are known to be a very good timbre descriptor. Since our mosaicing task has to sound like the target one, having the timbre in consideration would be a nice feature to preserve.

The first feature added by me was the MelBands standard algorithm. This one was used to have a complement to the MFCC timbre features, but in terms of the energy on the Mel frequency bands.

The second Essentia algorithm implemented by me to process the source audio was the onset estimator. I picked this one in order to allow a better frame selection according to the global and local beat. This allowed me to have a correction factor considering the bpm variations, and the fact that I was using a fixed frame window.

Finally, I incorporated a last timbre estimation based on the Harmonic Pitch Class Profile computed over the pitch peaks in the spectrogram.

With all these features I make sure that the K-nearest neighbors algorithm, used as a feature selection part of the analysis, considers the three: pitch, rhythmic and loudness features. I added all these features to the K nearest neighbors algorithm.

Along with this, I added 3 queries to the initial dataset. These three sounds were kick, snare and hihat. The intention of this was to give the system a set of sounds from which to reconstruct percussive components of the target. Using this, my dataset increased to up to 120 different samples.

Finally, I changed a little bit the *chose_frame_from_source_collection* function adding a certain amount of randomness. I did this in two different ways; the selection of K now is a random integer variable between 1 and 10; also, the selection of the sample within the K nearest neighbors is also random. Using this a variability factor was obtained in the generation of the mosaic resulting in always-different pieces to reconstruct the target sound. However, considering that the next sample is always within the 10 most similar in terms of the used Essentia features, this guarantees that the selected frame will always be similar to the target one.

Personally, I found this lab very useful to learn how to use the freesound API. At the same time, it gave us the necessary tools in order to use Essentia as a tool to analyze sounds, both at frame level (time and spectrogram), as well as at global level. I really enjoyed doing it.