

Hardware Accelerated Scientific Computing with Python

Recep Erol

COSMOS Friday Meeting

April 16, 2021

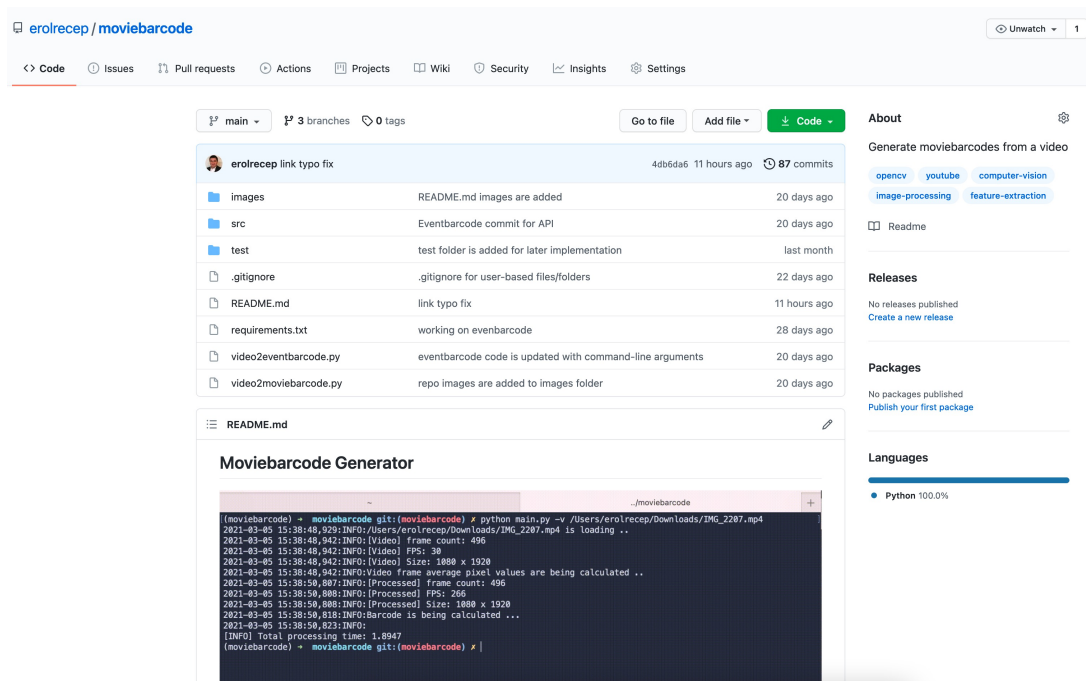
Agenda

- Research Progress
- Tech Stack and Tools
- Future Needs

Research Progress

- YouTube Video Categorization with
 - Moviebarcode
 - Topic Modeling
 - Data Fusion
 - t-SNE
 - Opinion-Mining

Research Progress - Moviebarcode



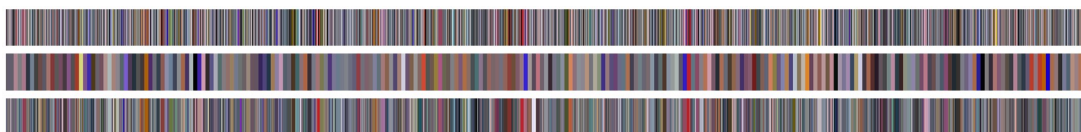
The screenshot shows the GitHub repository page for `erolrecep/moviebarcode`. The repository has 87 commits and 3 branches. The commit history shows recent updates to the README, source code, and test folders. A terminal window displays the following output:

```

(moviebarcode) ~ - ~/moviebarcode
(moviebarcode) ~ - ~/moviebarcode x | python main.py -v /Users/erolrecep/Downloads/IMG_2287.mp4
2021-03-05 15:38:48,929:INFO:Users/erolrecep/Downloads/IMG_2287.mp4 is loading ..
2021-03-05 15:38:48,942:INFO:(Video) Frame count: 496
2021-03-05 15:38:48,942:INFO:(Video) FPS: 30
2021-03-05 15:38:48,942:INFO:(Video) Size: 1080 x 1920
2021-03-05 15:38:48,942:INFO:(Video) frame average pixel values are being calculated ..
2021-03-05 15:38:50,807:INFO:(Processed) frame count: 496
2021-03-05 15:38:50,808:INFO:(Processed) FPS: 266
2021-03-05 15:38:50,808:INFO:(Processed) Size: 1080 x 1920
2021-03-05 15:38:50,818:INFO:Barcode is being calculated ...
2021-03-05 15:38:50,823:INFO:
[INFO] total processing time: 1.8947
(moviebarcode) ~ - ~/moviebarcode x |
  
```

- Single moviebarcode
- Multi-thread moviebarcode
- Multi-process moviebarcode
- GPU-based moviebarcode (in progress)

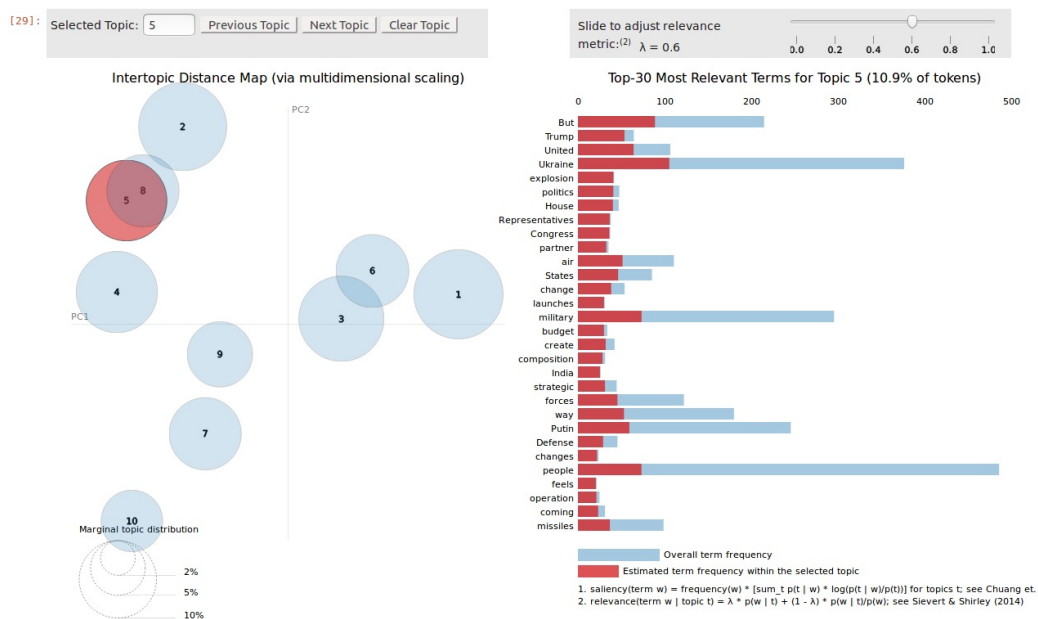
Research Progress - Eventbarcode



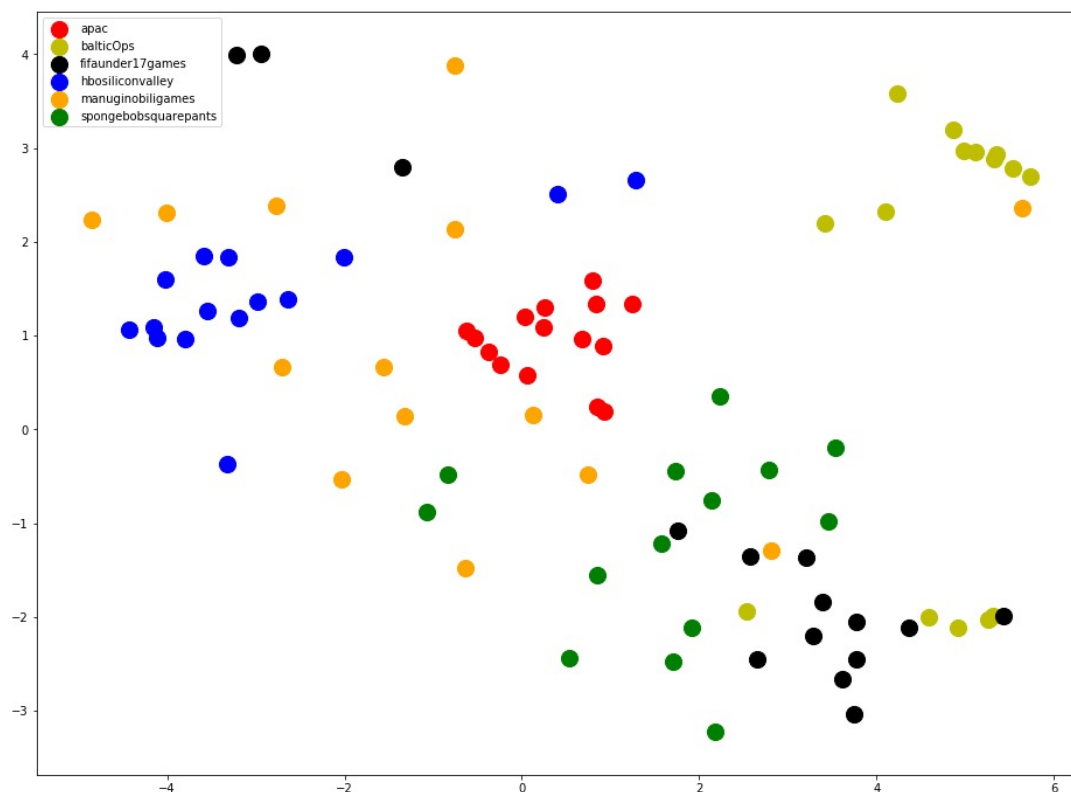
- Types of eventbarcode generations
 - First-frame (0.3 – 0.45 seconds)
 - Dominant Colors (1.2 – 2 seconds)
 - Short-clip (Future Plan)
- Single-thread eventbarcode
- Multi-process eventbarcode (in progress)
- GPU-based eventbarcode (future plan)

Research Progress – Text Processing

- Text2Vec for feature extraction
- Topic modeling
- Language models (such as BERT)
- Opinion scores
 - Emotion
 - Toxicity

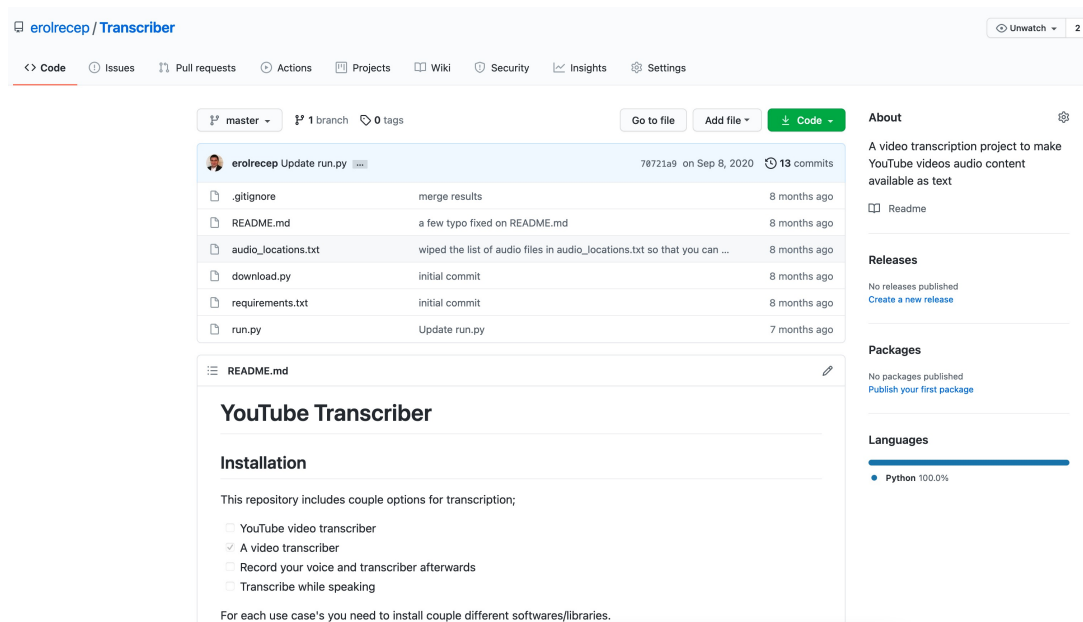


Research Progress – Unsupervised Visualization



- t-SNE
- UMAP (Future Plan)

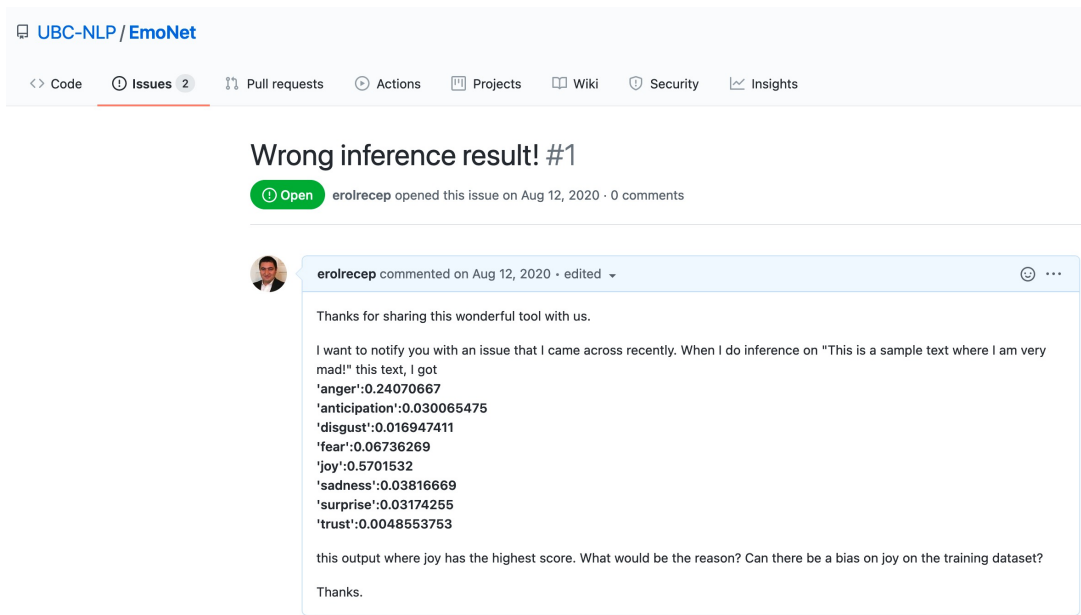
Research Progress – Speech Recognition



The screenshot shows the GitHub interface for the repository 'erolrecep/Transcriber'. At the top, there are navigation tabs for Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. Below this, the repository name and a 'Code' button are visible. A commit history table lists several commits, with the most recent one being 'erolrecep Update run.py' on Sep 8, 2020. The README file is open, showing the title 'YouTube Transcriber' and an 'Installation' section with a list of options: 'YouTube video transcriber', 'A video transcriber' (checked), 'Record your voice and transcriber afterwards', and 'Transcribe while speaking'. The right sidebar contains sections for 'About', 'Releases', 'Packages', and 'Languages', with 'Languages' showing 'Python 100.0%'.

- SoX
- DeepSpeech
- Wav2letter++ (future plan)
- Nemo (future plan)

Research Progress – Opinion Mining



UBC-NLP/EmoNet

<> Code Issues 2 Pull requests Actions Projects Wiki Security Insights

Wrong inference result! #1

Open erolrecep opened this issue on Aug 12, 2020 · 0 comments

erolrecep commented on Aug 12, 2020 · edited

Thanks for sharing this wonderful tool with us.

I want to notify you with an issue that I came across recently. When I do inference on "This is a sample text where I am very mad!" this text, I got

```
'anger':0.24070667
'anticipation':0.030065475
'disgust':0.016947411
'fear':0.06736269
'joy':0.5701532
'sadness':0.03816669
'surprise':0.03174255
'trust':0.0048553753
```

this output where joy has the highest score. What would be the reason? Can there be a bias on joy on the training dataset?

Thanks.

- EmoNet
 - Training needed
 - Labeled data
- Audio-based opinion-mining (future plan)
- Visual content-based opinion mining (future plan)

Research Progress – Video Characterization



- Plotly based treemap visualization
- Building web tool for YouTube Video Characterization

Data

APAC

BalticOps

CanadaElections2019-Influential

CanadaElections2019-Suspicious

Dimensions

Movie Barcode

Topic Modeling

Toxicity Features

Emotion Features

Network Clusters

Video Categorization

Organic/Inorganic Analysis

Topic Diversity & Periodicity

Visualization

2D Cluster Results

3D Cluster Results

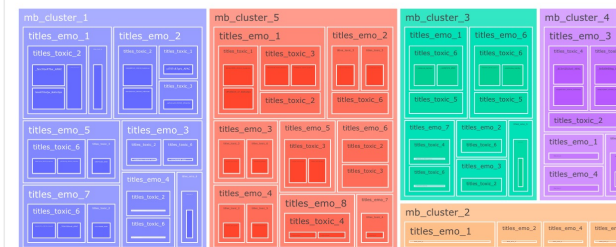
T-SNE

Hierarchical Clust. Dendo.

Network Communities

Treemap

Topic Modeling Results



Tech Stack and Tools - Current

- Python & C++
- NumPy, Scikit-Learn, TensorFlow, and Keras
- OpenCV
- CUDA, CUDA-C
- SoX – Sound Exchange
- CPU Programming
- Parallel Computing
- Distributed Computing
- Shared Memory Programming (OpenMP & MPI)
- GPU Programming
- Computer Vision Programming

Tech Stack and Tools - Future

- CPython
- Numba (or PyCUDA)
- Faiss
- CUDA – cuDNN
- JAX

Tech Stack and Tools - Future



Build for Scientific Computing

Just in Time (JIT) Compiler

LLVM Compiler Support

CUDA and cuDNN Support

C and Fortran Codebase

NumPy code -> Hardware Code -> Machine Code

```
from numba import jit
import random

@jit(nopython=True)
def monte_carlo_pi(nsamples):
    acc = 0
    for i in range(nsamples):
        x = random.random()
        y = random.random()
        if (x ** 2 + y ** 2) < 1.0:
            acc += 1
    return 4.0 * acc / nsamples
```

Tech Stack and Tools - Future

- AutoGrad (Deprecated)
- JIT
- XLA
- JAX
- Hardware Acceleration on GPU and TPU



Tech Stack and Tools - Future

- Python optimized AVX instruction Sets
- Built-in CPU acceleration for NumPy, Scikit-Learn, and visualization libraries

Intel® Distribution for Python*

Develop fast, performant Python code with this set of essential computational packages including NumPy, SciPy, scikit-learn*, and more.

Future Needs

- New coders are needed for computer vision & GPU programming
- Tool development engineers are needed:
 - Computer vision programming (OpenCV & Python)
 - Machine learning algorithm programmer (TensorFlow & CUDA)

Hardware Accelerated Scientific Computing with Python

Recep Erol

COSMOS Friday Meeting

April 16, 2021