Magneto Summer2020 Summary

Benjamin Ott

DigitizationTODO.py

The need for a program that told us what images had been digitized by previous versions became apparent, as it wouldn't be possible to re-digitize all of the images when the program was stopped. This was easiest to run in python because of its efficient computing capability. It asks the user for two .txt files *Figure: D1*; These files contain the path of one item, in our case these were the paths of the origional image and the path of the digitized image.

Original Image Path

/home/ben/magneto/Images/AGC-D-19010530-19040829/AGC-D-19031103-19031105.tif

Digitized Image Path

/home/ben/magneto/Digitized_Data/Algorithm2/AGC-D-19031103-19031105.tif.csv

---

The path was needed so that the digitize function would be able to find the image that hasn't been digitized yet (see TISForAutomation for more details about digitization function).

The path was then split into the name, the path and the file type; This program was designed to take any file type (.tiff, .tif, .tif.csv, .tif.png) as it separates by both "." and "/" to obtain

---

File Path: */home/ben/magneto/Images/AGC-D-19010530-19040829/*

File Name: *AGC-D-19031103-19031105*

File Type: *.tif*

This aloud us to ony compare the file names of images and digitized images, without needing to worry about the path or the file type. This caused an decrease of the computation time as there wasn't any splitting done when the comparison was happening. This caused another problem, as we were only comparing the names, in some cases there was digitized images that don't have original images; code was created to identify when this happens and print it to screen for the user to see. We recorded these in a document **DigitizedNonImages.txt** to hopefully identify the original images one day.

as the comparisons happen in the function, They are recorded and outputed to a .csv file that gets saved locally. Example of the output *Figure: D2*. In order of the columns, it saves: **Image Path, Image Name, If the image has been digitized, Digitized Path, Digitized Name**. Along with this .csv, there is also information that is printed to the user about the data *Figure: D3*; This includes information to ensure that all of the files are accounted for in the final document, along with the number of digitized images and the number of images that still need to be digitized.

Testing new part

jfdlshfjdshlfdskl

```
C:\Users\rocke\Documents\PycharmProjects\venv\Scripts\python.exe C:/Users/rocke/Documents/PycharmProjects/Magnetograms/digitizationTODO.py
What is the image file?
Images.txt
What is the comparison file?
DigitizedBatch5.txt
What is the desired output file name? (must be .csv)
DigitizedB5.csv
```

Figure 1: D1: The user input

/home/ben/magneto/Images/AGC-D-18980111-19010529/	AGC-D-18990911-18990913.tif	TRUE	/home/ben/Magneto2020/DigitizedImages/	AGC-D-18990911-18990913.tif.csv
/home/ben/magneto/Images/AGC-D-18980111-19010529/	AGC-D-18990913-18990915.tif	TRUE	/home/ben/Magneto2020/DigitizedImages/	AGC-D-18990913-18990915.tif.csv
/home/ben/magneto/Images/AGC-D-18980111-19010529/	AGC-D-18990915-18990917.tif	TRUE	/home/ben/Magneto2020/DigitizedImages/	AGC-D-18990915-18990917.tif.csv
/home/ben/magneto/Images/AGC-D-18980111-19010529/	AGC-D-18990917-18990919.tif	TRUE	/home/ben/Magneto2020/DigitizedImages/	AGC-D-18990917-18990919.tif.csv
/home/ben/magneto/Images/AGC-D-18980111-19010529/	AGC-D-18990919-18990921.tif	TRUE	/home/ben/Magneto2020/DigitizedImages/	AGC-D-18990919-18990921.tif.csv
/home/ben/magneto/Images/AGC-D-18980111-19010529/	AGC-D-18990921-18990923.tif	FALSE		
/home/ben/magneto/Images/AGC-D-18980111-19010529/	AGC-D-18990923-18990925.tif	TRUE	/home/ben/Magneto2020/DigitizedImages/	AGC-D-18990923-18990925.tif.csv

Figure 2: D2: .csv Output File Example

```
The amount digitized are: 3139
The amount that still needs to be done is: 23271
There are  12856 duplicate images in  Images.txt
There are  0 duplicate comparison files in  DigitizedBatch5.txt

VERY IMPORTANT
The ones that are not matched: [['AGC-D-19010729-19010730.tif.csv' 'in first array']
 ['AGC-D-19010908-19010910-BlackTraceOnly.tif.csv' 'in first array']
 ['AGC-D-19011008-19011010.tif.csv' 'in first array']
 ['AGC-D-19020402-19020404-allthecrazyredtrace.tif.csv' 'in first array']
 ['AGC-D-19020418-19020420.tif.csv' 'in first array']
 ['AGC-D-19020526-19020528-startTrace1.tif.csv' 'in first array']
 ['AGC-D-19020923-19020925.tif.csv' 'in first array']] First: comparison No duplicates, Second: matches with images
Length of notAccountedFor is 7

Check for the Total Comparison files:  3146 includes the files that aren't accounted for
Total Comparison files:  3146
Total Image files:  39266 subtract 12856 = 26410
Check for the total in end Array:  26410
Length of the final array made into csv:  26410
```

Figure 3: D3: Extra Information for User