**Hashing.ipynb(Jupiter notebook)**

Import hashlib

import scipy.misc

from imageio import imread

import matplotlib.pyplot as plt

%matplotlib inline

import time

import numpy as np

from hashlib import md5

def file\_hash(filepath):

with open(filepath, 'rb') as f:

return md5(f.read()).hexdigest()

import os

os.getcwd()

os.chdir(r'C:/4')

os.getcwd()

files\_list = os.listdir()

print(len(files\_list))

import hashlib, os

duplicates = []

hash\_keys = dict()

for index, filename in enumerate(os.listdir('.')):

if os.path.isfile(filename):

with open(filename, 'rb')as f:

filehash = hashlib.md5(f.read()).hexdigest()

if filehash not in hash\_keys:

hash\_keys[filehash] = index

else:

duplicates.append((index,hash\_keys[filehash]))

duplicates

for file\_indexes in duplicates[:40]:

try:

plt.subplot(121),plt.imshow(imread(files\_list[file\_indexes[1]]))

plt.title(file\_indexes[1]), plt.xticks ([]), plt.yticks([])

plt.subplot(122),plt.imshow(imread(files\_list[file\_indexes[0]]))

plt.title(str(file\_indexes[0]) + ' duplicate'), plt.xticks ([]), plt.xticks([]), plt.yticks([])

plt.show()

except OSError as e:

continue