Assignment 2

A solution for digitizing numerical hand-written data

In this project the question is whether there a way to automatically convert the hand-written numbers into digital data for computer use. 253+ million Federal tax documents are filed in FY 2019 according to IRS. A considerable portion of these documents are hand-written. A professional typist types at least 50 to 80 words per minute (wpm). It can be even higher for some typists. Each page in a typical tax document, there are around 50 numeric fields to fill which means almost 300 digits per page. Each filing could have several pages. As a result, it takes a long time even for a group of professional typists to enter such a volume of data especially when they are hand-written. A professional typist may have up to 92% accuracy in typing (Source Ratatype.com). The financial data are critical and an error in entering them might have serious consequences.

In this project I present a solution for automatically digitizing hand-written tax forms and prepare them for computer use. Optical Character Recognition (OCR) classifiers are what we are using. This technology has also been used for these applications among others:

* + zip code recognition on mail for postal mail sorting.
  + Processing bank check amounts.
  + Numerical input from tablets with electronic pen.

I have shown that my method has up to 95.72% accuracy that beats a professional typist. The results are presented in attached documents.